Transient Effects of Linear Dynamical Systems

von Elmar Plischke

Dissertation

zur Erlangung des Grades eines Doktors der Naturwissenschaften
– Dr. rer. nat. –

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im Juli 2005

Datum des Promotionskolloquiums: 19. August 2005

Gutachter: Prof. Dr. Diederich Hinrichsen (Universität Bremen)
PD Dr. Fabian Wirth (NUI Maynooth, Irland)

Contents

Introduction

From a practical point of view the concept of stability may be deceiving: Asymptotic stability allows for arbitrary growth before a decay occurs. These transient effects which are only of temporary nature have no influence on the asymptotic stability of a dynamical system. However, these effects might dominate the system's performance. Hence we are in need of information which describes the short-time behaviour of a dynamical system.

Motivation

Everyone has noticed that devices need some time to get ready for use, like an old radio warming up or a computer booting. On a larger scale, plants also need an amount of time to reach their working point. But in this initial phase, the plant is particularly vulnerable. The faster one wants to reach the working point, the more stress the plant has to endure: there may be overshots which carry some parts of the plant to the limit of their capacity. One may think of chemicals, which advance towards toxic or explosive concentrations, before reaching the desired concentration of the reagents, or an autopilot steering the wrong way before getting on track.

One could believe that such a distinctive behaviour in the initial phase can only occur for complex dynamical systems. However, this behaviour can already be observed for linear differential equations which provide simple models for dynamical processes.

To avoid catastrophes like those indicated above, one wishes to eliminate the bad influences in the starting phase, or at least, to keep them small. Furthermore, methods are needed that allow to predict if the system under consideration shows these transient effects, and if so, to obtain information on the duration and intensity of these excursions.

This work is mainly concerned with questions dealing with the last two issues, namely finding bounds on the exponential growth of linear systems. Although there are many results on exponential bounds, there is still no systematic treatment in the literature.

The mathematical model of a plant will in general not yield the accurate description of the behaviour of the real plant. Hence we are in need of results which are robust under small perturbations of the mathematical model. Fortunately, these results follow directly from our systematic treatment of the exponential bounds.

In addition to the general theory, we study two major classes of linear systems, namely positive systems and delay-differential systems, which are used frequently in economics and biology.

Moreover, we study the influence of state feedback on the transient behaviour. We obtain necessary and sufficient conditions to obtain a closed-loop system without transient excursions.

Finite-dimensional linear systems are mostly used as an approximation of more complex dynamical systems. These are obtained by linearization or discretization. Let us now discuss two possible ways in which the transients of linear systems may influence the dynamics.

From Transience to Turbulence

Most linear dynamical systems are obtained by linearizing a nonlinear model of a real process around an equilibrium point. Now, Liapunov's theorem implies that the nonlinear system is asymptotically stable if the linearization is asymptotically stable.

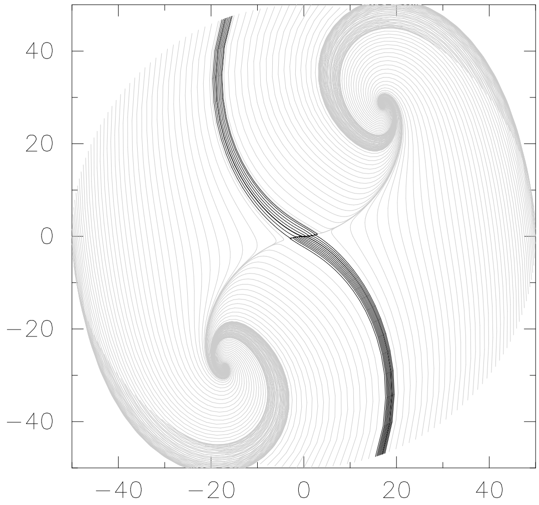

Figure 1: Toy model for turbulence.

But if the asymptotically stable linear system has solutions which move far afield before eventually decaying, these solutions of the linear system may leave the domain for which the linear system is a valid approximation of the nonlinear system. Hence small perturbations from the equilibrium point may incite nonlinearities. Models of this kind have been suggested in Baggett and Trefethen [8] to explain why turbulence of certain flows occurs at Reynolds numbers much smaller than predicted from a spectral analysis. For example, let us investigate the following nonlinear time-invariant ordinary differential equation

$$\dot{x} = Ax + B(x) = \begin{pmatrix} -5 & 36 \\ 0 & -20 \end{pmatrix} x + \|x\| \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x, \qquad x \in \mathbb{R}^2, \tag{1}$$

where A is an asymptotically stable, but nonnormal matrix and the nonlinearity $B(x)$ is conservative (energy-preserving), thus B only adds a rotation of the state-space to the linear system $\dot{x} = Ax$. Figure 1 shows many trajectories of system (1) starting on a circle of radius 50.

The trajectories which converge to the origin are colored in black, which gives a rough approximation of the domain of attraction for the origin. One observes that this domain of attraction is flat, the nonnormality of the linear system $\dot{x} = Ax$ quickly drives the state into regions where the nonlinearity has strong effects on the state. Note that in this example nonlinearity and nonnormality form opposite forces which create a sort of conveyor belt driving the states away from the stable origin. The picture drastically changes when replacing $B(x)$ by $-B(x)$.

From Transience to Permanence

Another interesting observation can be made when approximating infinite dimensional systems by finite dimensional approximants. Now assume that there exists a sequence of finite dimensional matrices which approximate an infinite dimensional linear operator, and that this sequence consists only of stable matrices. Then the infinite dimensional system need not be stable. But this can be detected by studying the transient behaviour of the approximants. Let us consider the matrix exponential of Jordan-blocks J_n of growing size n associated with the eigenvalue $\lambda_0 = -1/2$. These blocks approximate a "multiply-and-shift" operation J_∞ on the sequence space $\ell^2(\mathbb{C})$ given by

$$J_\infty : \ell^2(\mathbb{C}) \to \ell^2(\mathbb{C}),\ (x_k) \mapsto (x_{k+1} - \tfrac{1}{2}x_k).$$

But the spectrum of J_∞ consists of a whole unit ball centered around λ_0, because it is a Toeplitz operator with symbol $s \mapsto -\frac{1}{2} + s$, see Böttcher and Silbermann [21], and

$$\partial\sigma(J_\infty) = -\tfrac{1}{2} + \mathbb{S},\ \text{where } \mathbb{S} = \{s \in \mathbb{C} \mid |s| = 1\}.$$

Hence the asymptotic growth rate of the semigroup generated by J_∞ is given by $\alpha(J_\infty) = \sup\{\operatorname{Re}\lambda \mid \lambda \in \sigma(J_\infty)\} = 1/2$. Figure 2 shows the growth of $\|\exp(J_n t)\|$ for $n = 2, 4, 8, 16, 32$. Although all J_n, $n \in \mathbb{N}$, are stable, the limit $1/2$ of the transient growth rates given by $\mu_n = \frac{d}{dt^+} \left\| e^{J_n t} \right\| \big|_{t=0}$ coincides with the asymptotic growth rate of J_∞.

The Curse of Nonnormality

Both of these examples suffer from the same defect: the linear matrix A is highly nonnormal, i.e., there exists no orthogonal basis of eigenvectors. Moreover, if there are eigenvectors which roughly point into the same direction then there are vectors of small size for which the coordinates will blow up if they are represented in an eigenvector basis, see Figure 3 where $w = 7/2 v_1 - 13/4 v_2$. Clearly, if the angle spanned by v_1 and v_2 becomes more acute, this results in larger coordinates of w in the $\{v_1, v_2\}$-basis. Grossmann [49, 50] calls this behaviour the "blind spot" of such a basis.

Henrici [54] has identified the nonnormality as a cause for the failure of many numerical algorithms. He introduced the *departure from normality* as a measure of nonnormality.

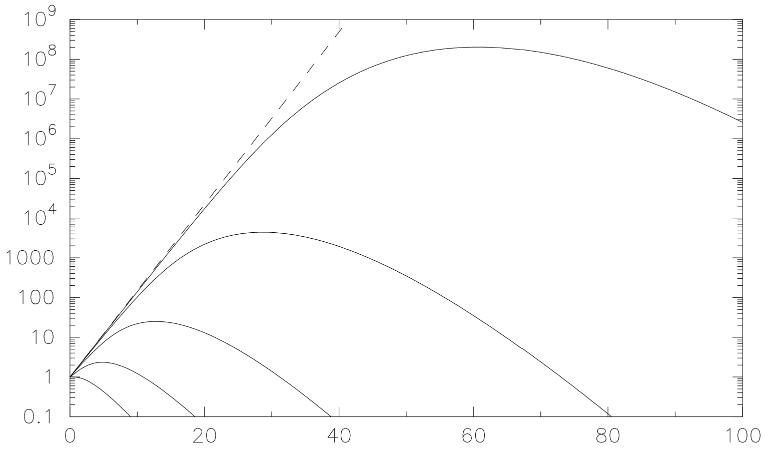

Figure 2: Exponential growth of Jordan blocks.

However, there are matrices for which the departure from normality is small, but transient effects are present, e.g. $\operatorname{dep}(J_n) = 1$ for a Jordan block of dimension n.

Outline of the Thesis

The next chapter is devoted to some mathematical preliminaries which fix the notations used throughout this thesis and cover some topics from linear algebra and functional analysis that are used in the later chapters.

In Chapter 2 we collect some facts for generators of strongly continuous semigroups and semigroups of contractions. We introduce an indicator for contractions which forms the basis of much of the later work. This indicator is related to a convex Liapunov function. We also consider some duality issues here and discuss the stability of differential inclusions.

Chapter 3 deals with a suitable concept of stability and discusses several types of estimates for the norm of the matrix exponential. Unfortunately, bounds based on spectral properties have several drawbacks. Hence we are investigating alternatives, we consider exponential bounds derived from quadratic Liapunov function and from resolvent estimates.

In Chapter 4 several small results are presented, including a discussion of transient norms and quadratic Liapunov functions of minimal condition number for 2×2 matrices.

In Chapter 5, results for positive systems are derived. In a sense these systems exhibit the "worst" transient behaviour as no cancellation of terms in the matrix exponential can occur. We show that Liapunov functions for positive systems are of simple structure, and so we can derive simple exponential bounds for positive systems.

Moreover, we can compare the transient behaviour of an arbitrary system with the behaviour of a positive one which allows us to apply the simple bounds to a large class of matrices.

We close the "analysis" part of the thesis by considering differential delay equations in Chapter 6. In order to obtain comparable results for the transient estimates a special class

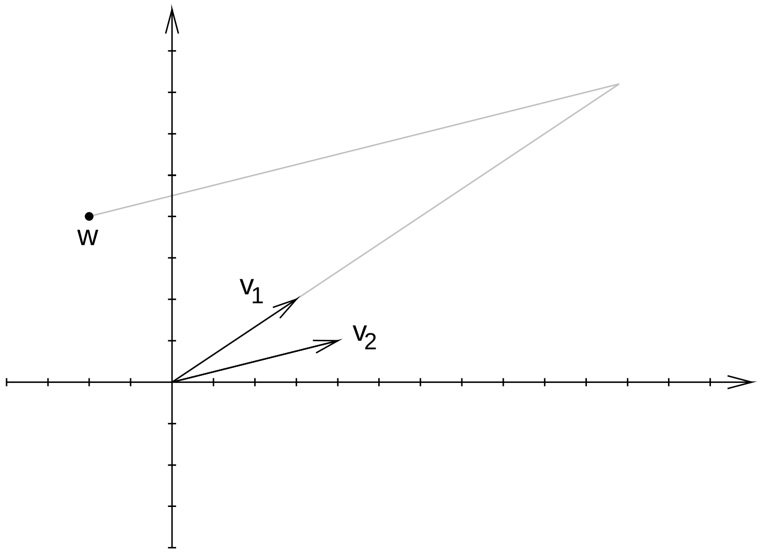

Figure 3: Nonnormality blows up coordinates.

of Liapunov functionals is introduced and studied. We also discuss numerical issues related with the computation of these Liapunov functionals.

On the "synthesis" side, Chapter 7 is devoted to results under which a closed-loop system satisfies given exponential bounds when state feedback matrices are introduced. We discuss this topic for general norms, as well as for the special case of quadratic norms.

Acknowledgments

I thank my advisor Prof. Dr. Diederich Hinrichsen and PD Dr. Fabian Wirth for their valuable advice and criticism. I also thank Prof. Dr. Anthony Pritchard and Prof. Dr. Vladimir Kharitonov for fruitful discussions.

I express my gratitude to the Hamilton Institute, NUI Maynooth, Ireland, for sponsoring a visitor's stay. I gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG) via the Graduiertenkolleg "Komplexe dynamische Systeme" and the grant from the project "Spektrale Wertemengen und Übergangsverhalten unsicherer dynamischer Systeme".

Chapter 1

Preliminaries

In this chapter we fix some terminology and notations used throughout later chapters. Let us denote the field of real numbers by $\mathbb{R}$ and the field of complex numbers by $\mathbb{C}$. The natural numbers are given by $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, and the ring of integers is called $\mathbb{Z}$.

1.1 Matrix Analysis

This section fixes the notation for some standard notions from linear algebra. The n-dimensional vector space over the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ is denoted by $\mathbb{K}^n$. If S is a subset of a vector space $\mathbb{K}^n$ the *span* or *linear hull* of S is the set

$$\operatorname{span} S := \left\{ \sum_{i=1}^{k} \alpha_i x_i \;\middle|\; \alpha_i \in \mathbb{K},\ x_i \in S,\ k = 1, 2, \dots \right\}.$$

The set $\operatorname{span} S$ is always a linear subspace of $\mathbb{K}^n$.

The space of linear operators from $\mathbb{K}^n$ into $\mathbb{K}^m$ is denoted by $\mathbb{K}^{m \times n}$. Its elements are called *matrices*. If $A = (a_{ij}) \in \mathbb{K}^{m \times n}$ then $A^\top = (a_{ji}) \in \mathbb{K}^{n \times m}$ is its *transpose* and $A^* = (\bar{a}_{ji}) \in \mathbb{K}^{n \times m}$ is its *Hermitian adjoint*. The *kernel* of $A \in \mathbb{K}^{n \times m}$ is given by $\ker A = \{x \in \mathbb{K}^n \,|\, Ax = 0\} \subset \mathbb{K}^n$ and the *image* of A by $\operatorname{im} A = \{Ax \,|\, x \in \mathbb{K}^n\} \subset \mathbb{K}^m$. For square matrices with $n = m$ we say that A is *symmetric* if $A = A^\top$, and it is *Hermitian* if $A = A^*$. It is called *normal* if $AA^* = A^*A$. The square matrix $A \in \mathbb{K}^{n \times n}$ is invertible if $\ker A = \{0\}$. Then there exists an inverse matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$. Here I_n is the identity matrix of dimension n. The set of all invertible square matrices $A \in \mathbb{K}^{n \times n}$ is the *general linear group* $\mathrm{Gl}_n(\mathbb{K})$.

Now let us consider the inverse of a partitioned matrix, see [70, Section 0.7.3].

Lemma 1.1. *Let $M = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ be an invertible matrix. Then the inverse M^{-1} can be* *obtained via one of the following formulas if the used inverses exist.*

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BQCA^{-1} & -A^{-1}BQ \\ -QCA^{-1} & Q \end{pmatrix}, \qquad Q = (D - CA^{-1}B)^{-1},$$

$$M^{-1} = \begin{pmatrix} R & -RBD^{-1} \\ -D^{-1}CR & D^{-1} + D^{-1}CRBD^{-1} \end{pmatrix}, \qquad R = (A - BD^{-1}C)^{-1}.$$

These two descriptions yield the Sherman-Morrison-Woodbury formula

$$A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} = (A - BD^{-1}C)^{-1}. \tag{1.1}$$

For the determinant of M we have

$$\det(M) = \det(A)\det(D - CA^{-1}B) = \det(A - BD^{-1}C)\det(D).$$

The matrices $D - CA^{-1}B$ and $A - BD^{-1}C$ are called Schur complements *of M.*

1.2 Properties of Norms

In this section we recall some facts for vector norms on the vector space $\mathbb{K}^n$. References for this material can be found in the books by Horn and Johnston [70] and by Bhatia [18], and the article by Bauer, Stoer and Witzgall [12]. Let us first study vector norms.

Definition 1.2. A norm $\|\cdot\|$ on $\mathbb{K}^n$ is called

1. *absolute*, if $\|x\| = \|\,|x|\,\|$ for all $x \in \mathbb{K}^n$, where $|x| = (|x_i|)_{i=1,\dots,n}$,
2. *monotone*, if $\|x\| \leq \|y\|$ for all $x, y \in \mathbb{K}^n$ with $|x_i| \leq |y_i|, i = 1, \dots n$,
3. *symmetric*, if $\|x\| = \|Px\|$ for all $x \in \mathbb{K}^n$ and all perturbation matrices $P \in \{0,1\}^{n\times n}, P^2 = I_n$.

Proposition 1.3. *The p-norms on $\mathbb{K}^n$, given by*

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, \; 1 \leq p < \infty, \; \textit{and} \; \|x\|_\infty = \max_i |x_i|,$$

satisfy properties (1)-(3) of Definition 1.2.

Proposition 1.4 ([70, Theorem 5.5.10])**.** *A norm on $\mathbb{K}^n$ is absolute if and only if it is monotone.*

If $v(\cdot)$ is a norm on $\mathbb{K}^n$, the dual of $v(\cdot)$ is defined by

$$v^*(y) = \sup_{v(x)=1} |\langle x, y\rangle_2| = \sup_{v(x)=1} \operatorname{Re}\langle x, y\rangle_2.$$

Here $\langle x, y\rangle_2 = y^*x$ is the inner product of x and y. It is easy to see that $v^*(\cdot)$ is a norm. In fact, $v^*(\cdot)$ is a norm even if $v(\cdot)$ is a function which does not satisfy the triangle inequality, but meets all other requirements of a norm. If $\mathbb{B} = \{x \in \mathbb{K}^n \mid \|x\| \leq 1\}$ is the closed unit ball of $\|\cdot\|$ then we denote the unit ball of the dual norm by $\mathbb{B}^*$.

Proposition 1.5 ([12, Theorem 1]). *If $v(\cdot)$ is a monotone norm on $\mathbb{K}^n$, then so is $v^*(\cdot)$.*

Proposition 1.6. *The dual norm of $\|\cdot\|_p$ is given by $\|\cdot\|_q$ where $\frac{1}{p} + \frac{1}{q} = 1$. Especially the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are dual norms on $\mathbb{K}^n$.*

Proposition 1.7 ([70, Theorem 5.5.14]). *The bidual ν^{**} of a norm ν on $\mathbb{K}^n$ equals ν, $\nu^{**}(x) = \nu(x)$ for all $x \in \mathbb{K}^n$.*

A norm $\|\cdot\|$ on $\mathbb{K}^n$ is called *smooth* if it is Gâteaux-differentiable in every $x \neq 0$. A norm $\|\cdot\|$ is smooth if and only if for every x_0 with $\|x_0\| = 1$ there exists a uniquely determined $y_0 \in \mathbb{K}^n$ with $\|y_0\|^* = \langle y_0, x_0 \rangle_2 = 1$, see Werner [147, Satz III.5.3]. The dual norm of a smooth norm is in general not smooth.

Let us now turn to norms on $\mathbb{K}^{n\times n}$. A norm $\|\cdot\|$ on $\mathbb{K}^{n\times n}$ is called a *matrix norm*, if it is sub-multiplicative, that is, it satisfies $\|AB\| \leq \|A\| \, \|B\|$ for all $A, B \in \mathbb{K}^{n\times n}$. If $\|\cdot\|$ is a norm on $\mathbb{K}^n$ then $A \mapsto \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$ is a norm on $\mathbb{K}^{n\times n}$ called the *operator norm* which we also denote with $\|\cdot\|$. Each operator norm on $\mathbb{K}^{n\times n}$ is a matrix norm.

Example 1.8. A norm on $\mathbb{K}^{n\times n}$ which is not a matrix norm is given by the *numerical radius* of $A \in \mathbb{K}^{n\times n}$,

$$r_{\text{num}}(A) = \sup_{x\neq 0} \left| \frac{\langle x, Ax \rangle_2}{\langle x, x \rangle_2} \right|.$$

However, for every norm ν on $\mathbb{K}^{n\times n}$ there exists a positive real constant $\alpha > 0$ such that $\alpha\nu$ is a matrix norm. Here the norm $\nu(A) = 4r_{\text{num}}(A)$ is a matrix norm on $\mathbb{K}^{n\times n}$, see [70, p. 331]. ■

The following lemma shows some properties of operator norms induced by monotone vector norms.

Lemma 1.9. *If $\|\cdot\|$ is a monotone vector norm on $\mathbb{K}^n$ then the associated operator norm satisfies the following properties, see [12].*

(i) If $A = (a_{ij})$ and $B = (b_{ij})$ are nonnegative matrices in $\mathbb{R}^{n\times n}$ that satisfy $a_{ij} \geq b_{ij} \geq 0$ then $\|A\| \geq \|B\|$.

(ii) For $A = (a_{ij}) \in \mathbb{K}^{n\times n}$ we have $\|A\| \leq \| \, |A| \, \|$ where $|A| = (|a_{ij}|) \in \mathbb{R}^{n\times n}$.

(iii) The vector norm $\|\cdot\|$ is monotone if and only if the induced operator norm satisfies $\|\Lambda\| = \max_{i=1,\dots,n} |\lambda_i|$ for all diagonal matrices $\Lambda = \operatorname{diag}(\lambda_i)$.

Hence, the operator norm induced from a monotone norm is not monotone by itself, but satisfies the monotonicity condition only on the positive orthant.

1.3 Spectral Value Sets and Stability Radii

In this section we present some notions which are used to analyse robustness issues for matrices, see Hinrichsen and Pritchard [67]. Spectral value sets have been introduced in [64, 65] as a tool to cope with the behaviour of highly nonnormal systems. These sets are used to study the robustness of dynamical systems, including the analysis of numerical algorithms.

Definition 1.10. The pair $(\mathbf{\Delta}, \|\cdot\|_{\mathbf{\Delta}})$ is called a *perturbation structure* in $\mathbb{K}^{\ell\times q}$ if $\mathbf{\Delta}\subset\mathbb{K}^{\ell\times q}$ is a closed convex cone and $\|\cdot\|_{\mathbf{\Delta}}$ is a norm on the linear span of $\mathbf{\Delta}$, $\operatorname{span}(\mathbf{\Delta})$. If $\mathbb{C}\mathbf{\Delta}=\mathbf{\Delta}$ then we call $(\mathbf{\Delta}, \|\cdot\|_{\mathbf{\Delta}})$ a *complex* perturbation structure.

Let us consider affine perturbations of the form

$$A \rightsquigarrow A_\Delta := A + B\Delta C, \qquad \Delta\in\mathbf{\Delta}, \tag{1.2}$$

where $B\in\mathbb{K}^{n\times\ell}$ and $C\in\mathbb{K}^{q\times\ell}$ are given *structure matrices*, and $(\mathbf{\Delta}, \|\cdot\|)$ is a perturbation structure in $\mathbb{K}^{\ell\times q}$.

Definition 1.11. Let A be a matrix in $\mathbb{K}^{n\times n}$, and $B\in\mathbb{K}^{n\times\ell}$ and $C\in\mathbb{K}^{q\times n}$ be structure matrices. For a given perturbation structure $(\mathbf{\Delta}, \|\cdot\|)$ in $\mathbb{K}^{\ell\times q}$ we define the following notions. The *spectral value set* of A corresponding to the perturbation level $\varepsilon\geq 0$ is given by

$$\sigma_\varepsilon(A,B,C\,|\,\mathbf{\Delta}) = \sigma(A)\cup\bigcup_{\Delta\in\mathbf{\Delta},\|\Delta\|<\delta}\sigma(A+B\Delta C).$$

The *structured pseudospectral abscissa* of A corresponding to the level $\varepsilon\geq 0$ is given by

$$\alpha_\varepsilon(A,B,C\,|\,\mathbf{\Delta}) = \sup\left\{\operatorname{Re}s\,|\,s\in\sigma_\varepsilon(A,B,C\,|\,\mathbf{\Delta})\right\}.$$

The *structured stability radius* of $A\in\mathbb{K}^{n\times n}$ is defined by

$$r(A,B,C\,|\,\mathbf{\Delta}) = \inf\left\{\varepsilon\geq 0\,|\,\alpha_\varepsilon(A,B,C\,|\,\mathbf{\Delta})\geq 0\right\}.$$

For *full block perturbations* $\mathbf{\Delta}=\mathbb{K}^{\ell\times q}$ where $\|\cdot\|$ is an operator norm on $\mathbb{K}^{\ell\times q}$, we drop the dependence on $\mathbf{\Delta}$. In the unstructured case $B=C=I_n$ we write $\sigma_\varepsilon(A\,|\,\mathbf{\Delta})$, $\alpha_\varepsilon(A\,|\,\mathbf{\Delta})$ and $r(A\,|\,\mathbf{\Delta})$.

The spectral value set is the union of all the spectra of the perturbed matrices A_Δ where $\Delta\in\mathbf{\Delta}$ and $\|\Delta\|\leq\varepsilon$. The stability radius measures the robustness of the stability of the matrix A under perturbations of the form (1.2).
Instead of characterizing the spectral value sets in terms of spectra of perturbed matrices, we have the following description in terms of the resolvent of A, $R(s,A)=(sI_n-A)^{-1}$.

Theorem 1.12 ([67, Theorem 5.2.16])**.** *Given $A\in\mathbb{C}^{n\times n}$, let $B\in\mathbb{C}^{n\times\ell}$ and $C\in\mathbb{C}^{q\times n}$ be given structure matrices. If $\mathbf{\Delta}=\mathbb{C}^{\ell\times q}$ and $(\mathbf{\Delta}, \|\cdot\|)$ is a full block perturbation structure the spectral value set of A for the level ε is given by*

$$\sigma_\varepsilon(A,B,C) = \sigma(A)\cup\left\{s\in\mathbb{C}\setminus\sigma(A)\,\middle|\,\left\|C(sI-A)^{-1}B\right\|>\varepsilon^{-1}\right\}.$$

For the unstructured case, if $B = I_n = C$ and if $\|\cdot\|$ is the spectral norm $\|\cdot\|_2$, we denote $\sigma_\varepsilon(A, B, C) = \sigma_\varepsilon(A)$ and obtain

$$\sigma_\varepsilon(A) = \sigma(A) \cup \{s \in \varrho(A) \,|\, \sigma_{\min}(sI_n - A) < \varepsilon\}\,.$$

Here $\sigma_{\min}(A) = \sigma_n(A)$ denotes the smallest singular value of A. An unstructured spectral value set of level ε is also called an *ε-pseudospectrum* of A.

1.4 Linear Operators

Let X and Y be Banach or Hilbert spaces over $\mathbb{K}$ equipped with the norm $\|\cdot\|$ where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. Let $A : D \to Y$ be a linear map defined on a linear subspace $D \subset X$ and taking its values in a Banach space Y. A is called a *linear operator* with *domain* $D(A) := D \subset X$ and *range* $A[X] := \{Ax \,|\, x \in D(A)\} \subset Y$. The symbol $\mathcal{L}(X, Y)$ stands for the Banach space of all *bounded linear operators* from X into Y (endowed with the operator norm). We write $\mathcal{L}(X)$ instead of $\mathcal{L}(X, X)$. The *identity operator* on X is denoted by I_X or just by I. An operator A is said to be *closed* if the *graph* of A defined by $\{(x, Ax) \in X \times Y \,|\, x \in D(A)\}$ is a closed subset of $X \times Y$, and it is called a *dense* operator if $A[X]$ is dense in Y. A is *densely defined* if $D(A)$ is dense in X.

Definition 1.13. Let A be a closed linear operator on X.

1. The *resolvent set* of A is given by

$$\varrho(A) = \left\{s \in \mathbb{C} \,\middle|\, (sI_X - A)^{-1} \text{ exists in } \mathcal{L}(X)\right\}.$$

2. The operator function $R(s, A) : \varrho(A) \to \mathcal{L}(X)$, $s \mapsto (sI_X - A)^{-1}$ is called the *resolvent* of A.

3. The complement in $\mathbb{C}$ of the resolvent set is called the *spectrum* of A, $\sigma(A) := \mathbb{C} \backslash \varrho(A)$. We define the following subsets of the spectrum,

$$\begin{aligned}
\sigma_P(A) &= \{s \in \mathbb{C} \,|\, sI_X - A \text{ is not injective}\}\,,\\
\sigma_C(A) &= \{s \in \mathbb{C} \,|\, sI_X - A \text{ is injective, not surjective, with dense range}\}\,,\\
\sigma_R(A) &= \{s \in \mathbb{C} \,|\, sI_X - A \text{ is injective and without dense range}\}\,.
\end{aligned}$$

$\sigma_P(A)$ is called the *point* spectrum of A, $\sigma_C(A)$ the *continuous* spectrum of A, and $\sigma_R(A)$ the *residual* spectrum of A. A point $\lambda \in \sigma_P(A)$ is an *eigenvalue* of A and $x \in D(A)$, $x \neq 0$ such that $\lambda x = Ax$ is a corresponding *eigenvector*.

Like in the matrix case, we have the following tool for robustness analysis of closed linear operators, see Hinrichsen, Gallestey and Pritchard [60].

Definition 1.14. Let A be a closed and densely defined linear operator on a Banach space X. The following set associated with the perturbation level $\varepsilon > 0$,

$$\sigma_\varepsilon(A) = \bigcup_{\Delta \in \mathcal{L}(X), \|\Delta\| < \varepsilon} \sigma(A + \Delta),$$

is called the *ε-pseudospectrum* of A. The *ε-pseudospectral abscissa* of A is given by

$$\alpha_\varepsilon(A) = \sup\{\operatorname{Re} s \mid s \in \sigma_\varepsilon(A)\}.$$

The ε-pseudospectrum can also be characterized via the resolvent of A.

Theorem 1.15. *Let A be a closed and densely defined linear operator on a Banach space X. If $\varepsilon \in \left(0, \sup_{s \in \varrho(A)} \|R(s,A)\|^{-1}\right)$ then*

$$\sigma_\varepsilon(A) = \sigma(A) \cup \left\{ s \in \varrho(A) \mid \|R(s,A)\| > \varepsilon^{-1} \right\}.$$

1.4.1 Block-Diagonal Operators

We now study a special class of linear operators for which the spectrum just consists of the point spectrum. We consider the Hilbert space

$$X = \bigoplus_{k \in \mathbb{N}} \mathbb{C}^{n_k} = \left\{ (x^k)_{k \in \mathbb{N}} \;\middle|\; x^k \in \mathbb{C}^{n_k}, \sum_{k \in \mathbb{N}} \|x_k\|^2 < \infty \right\},$$

which is called the *Hilbert direct sum* of $X_k = \mathbb{C}^{n_k}$, $n_k \geq 1$, see also [36, Definition IV.4.17].

We denote the elements of X by $(x^k)_{k \in \mathbb{N}}$ or for short, (x^k). Here each x^k is contained in $\mathbb{C}^{n_k}$. The space X is equipped with the inner product $\left\langle (x^k), (y^k) \right\rangle_2 = \sum_{k \in \mathbb{N}} \langle x_k, y_k \rangle_2$. Given a sequence of square matrices $A_k \in \mathbb{C}^{n_k \times n_k}$ for $k \in \mathbb{N}$, we define the *block-diagonal operator*

$$A = \bigoplus_{k \in \mathbb{N}} A_k : X \to X, \qquad A(x^k)_{k \in \mathbb{N}} = (A_k x^k)_{k \in \mathbb{N}}.$$

The domain of A is given by

$$D(A) = \left\{ x \in X \;\middle|\; \sum_{k \in \mathbb{N}} \left\| A_k x^k \right\|^2 < \infty \right\}.$$

This is a dense subset of X because $D(A)$ contains the following set

$$\{(x^k)_{k \in \mathbb{N}} \in X \mid x^k \neq 0 \text{ for only finitely many } k\},$$

which is a dense subset of X.

Lemma 1.16. $A = \bigoplus_{k \in \mathbb{N}} A_k$ *is a closed and densely defined linear operator on X.*

Proof. We have already seen that A is densely defined. Let us take a sequence (x_j) in $D(A)$ such that $x_j = (x^{k,j})_{k\in\mathbb{N}} \in X$. If we assume that both limits $x_j \to x$ and $y_j = Ax_j \to y$ exist in X then

$$y = \lim_{j\to\infty} Ax_j = \lim_{j\to\infty} \left(\bigoplus_{k\in\mathbb{N}} A_k\right) x^{k,j} = (A_k \lim_{j\to\infty} x^{k,j})_{k\in\mathbb{N}} = (A_k x^k)_{k\in\mathbb{N}} = Ax.$$

Hence $x \in D(A)$ and $Ax = y$ and therefore A is a closed operator in X. □

Let us now have a look at the norm and the spectrum of block-diagonal operators.

Lemma 1.17. *The operator norm of the block-diagonal operator $A = \bigoplus_{k\in\mathbb{N}} A_k : X \to X$ is given by $\|A\| = \sup_{k\in\mathbb{N}} \|(A_k)\|_2$.*

Proof. Given $x = (x^k)$ with $\sum_{k\in\mathbb{N}} \left\|x^k\right\|_2^2 < \infty$. Then

$$\|Ax\|^2 = \sum_{k\in\mathbb{N}} \left\|A_k x^k\right\|_2^2 \leq \sum_{k\in\mathbb{N}} \|A_k\|_2^2 \left\|x^k\right\|_2^2 \leq \sup_{k\in\mathbb{N}} \|A_k\|_2^2 \sum_{k\in\mathbb{N}} \left\|x^k\right\|_2^2 = \left(\sup_{k\in\mathbb{N}} \|A_k\|_2^2\right) \|x\|^2 .$$

On the other hand, there exists a sequence $x^k \in \mathbb{C}^{n_k}$ such that $\left\|x^k\right\|_2 = 1$ and $\|A_k\|_2 = \left\|A_k x^k\right\|_2$. Consider the sequence

$$\tilde{x}_j = ((\tilde{x}^k))_k \subset X \quad \text{where} \quad \tilde{x}^k = \begin{cases} x^j, & k = j, \\ 0, & k \neq j. \end{cases}$$

Then $\|\tilde{x}_j\| = 1$ and $\sup_{x\leq 1} \|Ax\| \geq \sup_j \|A\tilde{x}_j\| = \sup_j \|A_j x^j\|_2 = \sup_j \left\|A_{k_j}\right\|_2$. Therefore $\|A\| = \sup_k \|A_k\|_2$. □

From this lemma we obtain the following implications.

Corollary 1.18. *Let A be a block-diagonal operator in X. The resolvent set of A is given by the complement of*

$$\sigma(A) = \bigcap_{\varepsilon>0} \bigcup_{k\in\mathbb{N}} \sigma_\varepsilon(A_k).$$

On $\varrho(A) = \sigma(A)^{\mathrm{C}}$ the resolvent is given by $R(s, A) = \bigoplus_{k\in\mathbb{N}} R(s, A_k)$. The point spectrum of A is given by $\sigma_P(A) = \bigcup_{k\in\mathbb{N}} \sigma(A_k)$.

Proof. For $s \notin \bigcup_{k\in\mathbb{N}} \sigma(A_k)$ the operator $\bigoplus_{k\in\mathbb{N}} R(s, A_k)$ satisfies $\left(\bigoplus_{k\in\mathbb{N}} R(s, A_k)\right)(sI_X - A) = \bigoplus_{k\in\mathbb{N}}(R(s, A_k)(sI_{X_k} - A_k)) = I_X$ and analogously, $(sI_X - A)\left(\bigoplus_{k\in\mathbb{N}} R(s, A_k)\right) = I_X$. By Lemma 1.17 it is a bounded operator if and only if $\sup_{k\in\mathbb{N}} \|R(s, A_k)\|_2 < \infty$. Now

$$\sup_{k\in\mathbb{N}} \|R(s, A_k)\|_2 = \infty \iff \forall\varepsilon > 0 \,\exists\, k \in \mathbb{N} : \ \|R(s, A_k)\|_2 > \varepsilon^{-1}.$$

By Theorem 1.15 we write the set satisfing this condition as $\bigcap_{\varepsilon>0} \bigcup_{k\in\mathbb{N}} \sigma_\varepsilon(A_k)$. Hence $\bigoplus_{k\in\mathbb{N}} R(s, A_k)$ is undefined for $s \in \sigma_P(A) = \bigcup_{k\in\mathbb{N}} \sigma(A_k)$ and it is unbounded for $s \in \sigma(A) \setminus \sigma_P(A)$. Therefore the resolvent of A is $R(s, A) = \bigoplus_{k\in\mathbb{N}} R(s, A_k)$. □

Clearly $\overline{\bigcup_{k\in\mathbb{N}} \sigma(A_k)} \subset \sigma(A)$. If the ε-pseudospectra of A_k are disjoint for $\varepsilon > 0$ small enough then $\sigma(A) = \sigma_P(A)$.

1.4.2 Self-Adjoint Operators

Let X and Y be Hilbert spaces over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, equipped with the inner products $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$, respectively. The material presented here follows [29, Appendix A.3] and [147]. Let us define the adjoint of an unbounded operator.

Definition 1.19. Let A be a densely defined linear operator on X. Then the adjoint $A^* : D(A^*) \to Z$ is defined as follows. The domain $D(A^*)$ of A^* consists of all $y \in X$ such that there exists a $y^* \in X$ satisfying $\langle y, Ax \rangle_X = \langle y^*, x \rangle_X$ for all $x \in D(A)$. For each such $y \in D(A^*)$ the adjoint operator A^* is defined by $A^* y = y^*$.
We say that a densely defined linear operator A on X is *symmetric* if $\langle x, Ay \rangle_X = \langle Ax, y \rangle_X$ holds for all $x, y \in D(A)$. A symmetric operator is *self-adjoint* if $D(A^*) = D(A)$.

It can be shown that if A is a closed, densely defined linear operator then A^* is also closed and densely defined.
For continuous linear operators we define the following classes of Hilbert space operators.

Definition 1.20. Let $A \in \mathfrak{L}(X, Y)$.

(i) The operator $A^* \in \mathfrak{L}(Y, X)$ which satisfies $\langle Ax, y \rangle_Y = \langle x, A^* y \rangle_X$ for all $x \in X$,$y \in Y$ is called the *adjoint operator* of A.

(ii) A is called *unitary* if A is invertible with $AA^* = I_Y$ and $A^* A = I_X$.

(iii) Let $X = Y$. A is called *self-adjoint* (or Hermitian) if $A = A^*$.

(iv) Let $X = Y$. A is called *normal* if $AA^* = A^* A$.

Clearly, self-adjoint and unitary (in case of $X = Y$) operators are normal.

Lemma 1.21 ([147, Lemma V.5.10]). *For $A \in \mathfrak{L}(X)$ the following facts are equivalent.*

1. *A is normal.*

2. *$\|Ax\| = \|A^* x\|$ for all $x \in X$.*

The norm of a self-adjoint operator can be calculated as follows.

Proposition 1.22. *For a self-adjoint operator $A \in \mathfrak{L}(X)$ the norm is given by*

$$\|A\| = \sup_{\|x\| \leq 1} |\langle x, Ax \rangle_X| .$$

Lemma 1.23. *Let A be an element of $\mathfrak{L}(X)$ where X is a complex Hilbert space. A is self-adjoint if and only if $\langle x, Ax \rangle_X$ is real for all $x \in X$.*

Definition 1.24. A self-adjoint operator A on the Hilbert space X is called *nonnegative*, if $\langle x, Ax \rangle_X \geq 0$ for all $x \in D(A)$, and it is called *positive* if $\langle x, Ax \rangle_X > 0$ for all nonzero $x \in D(A)$. If there exists an $\varepsilon > 0$ such that $\langle x, Ax \rangle_X \geq \varepsilon \|x\|^2$ for all $x \in D(A)$ then A is called *coercive*.

Chapter 2

Contractions and Liapunov Norms

Consider a linear time-invariant dynamical system, $\dot{x} = Ax$ where $A \in \mathbb{C}^{n\times n}$. If all solutions decay in norm with $t \geq 0$ growing, A is said to generate a contraction semigroup. In this chapter we want to address the relationship between stability and contractions, and the dependence of the contraction property on the norm. Moreover, as quadratic Liapunov functions allow an interpretation as norms, we introduce the concept of Liapunov norms which provides a link between stability and contractions.

2.1 One-Parameter Semigroups in Banach Spaces

One-parameter semigroups provide a natural generalization of the flow concept associated with a system of linear ordinary differential equations. In this section we want to recall some of its properties. For an in-depth coverage see Engel and Nagel [38], Curtain and Zwart [29], Pazy [113], and Hille and Phillips [59].

Definition 2.1. Let X be a given (real or complex) Banach space and $(T(t))_{t\in\mathbb{R}_+}$ be a family of operators in $\mathcal{L}(X)$. This family is called a *strongly continuous semigroup* in X if

$$T(t+s) = T(t)T(s), \quad t,s \geq 0, \qquad T(0) = I \quad \text{and}$$
$$\varphi_x : t \mapsto T(t)x \text{ is continuous on } \mathbb{R}_+ \text{ for all } x \in X.$$

The semigroup $(T(t))_{t\in\mathbb{R}_+}$ is called *uniformly continuous* if

$$\lim_{h\to 0} \|T(t+h) - T(t)\| = 0 \quad \text{for all} \quad t \geq 0.$$

In the following T or $(T(t))_{t\in\mathbb{R}_+}$ will denote a strongly continuous semigroup on a real or complex Banach space.

Each uniformly continuous semigroup is also strongly continuous. Strongly continuous semigroups are also called C_0-semigroups. We list some known properties of strongly continuous semigroups in the following proposition, which combines results from [38, Proposition I.5.3], [29, Theorem 2.1.6], and [113, Theorem 1.2.4].

Proposition 2.2. *A strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$ on a Banach space X has the following properties.*

- *For all $x \in X$, $\lim_{t\searrow 0} T(t)x = x$.*
- *For all $x \in X$, $\lim_{t\searrow 0} \frac{1}{t}\int_0^t T(s)xds = x$.*
- *$\|T(t)\|$ is bounded on every finite subinterval of $\mathbb{R}_+$.*

For continuity issues we have the following result.

Lemma 2.3. *[147, Lemma VII.4.3] If $(T(t))_{t\in\mathbb{R}_+}$ is a strongly continuous semigroup on a Banach space X then the map*

$$\mathbb{R}_+ \times X \to X,\ (t,x) \mapsto T(t)x$$

is continuous, and uniformly continuous in T on compact intervals of $\mathbb{R}_+$. In particular, for every $x_0 \in X$ the map $x : t \mapsto T(t)x_0$ is continuous, $x \in C(\mathbb{R}_+, X)$.

Associated with every T is the (infinitesimal) *generator* A given by

$$Ax = \lim_{h\searrow 0} \tfrac{1}{h}(T(h)x - x), \tag{2.1}$$

which is defined for every $x \in X$ for which the limit in the right hand side of (2.1) exists, i.e., the domain of A is given by

$$D(A) = \{x \in X \mid \lim_{h\searrow 0} \tfrac{1}{h}(T(h)x - x) \text{ exists}\}.$$

If T is uniformly continuous then $D(A) = X$ and A is a bounded linear operator. A uniformly continuous semigroup can always be written as the exponential of its generator $T(t)x = e^{At}x, x \in X, t \geq 0$ where the exponential is defined by the familiar power series

$$e^{At} := \sum_{k=0}^{\infty} \tfrac{1}{k!}(tA)^k, \qquad t \geq 0,$$

which is absolutely convergent in $\mathcal{L}(X)$. For all $x \in D(A)$ the orbit map $\varphi_x : t \mapsto T(t)x$ is right differentiable at $t = 0$ and $\dot{\varphi}_x(0)$ and $\frac{d}{dt}T(t)x\big|_{t=0}$ always denote the right derivative at $t = 0$. By the semigroup property, φ_x is differentiable for all $t > 0$, because for $h > 0$, $h < t$

$$\begin{aligned}
&\lim_{h\searrow 0} \tfrac{1}{h}(T(t+h) - T(t))x = T(t) \lim_{h\searrow 0} \tfrac{1}{h}(T(h) - T(0))x = T(t)\dot{\varphi}_x(0)\\
&\lim_{h\searrow 0} \tfrac{1}{h}(T(t) - T(t-h))x = \lim_{h\searrow 0} T(t-h)\tfrac{1}{h}(T(h) - T(0))x = T(t)\dot{\varphi}_x(0).
\end{aligned}$$

Strongly continuous semigroups always admit an exponential bound.

Proposition 2.4 ([38, Proposition I.5.5]). *If $(T(t))_{t\in\mathbb{R}_+}$ is a strongly continuous semigroup then there exist constants $\beta \in \mathbb{R}$ and $M \geq 1$ such that*

$$\|T(t)\| \leq Me^{\beta t}, \qquad t \geq 0. \tag{2.2}$$

Definition 2.5. A strongly continuous semigroup $(T(t))_{t\in\mathbb{R}}$ is called *exponentially stable* if there exist constants $M \geq 1$, $\beta < 0$ such that (2.2) is satisfied. It is called *(M, β)-stable* if the semigroup T satisfies the growth bound (2.2) for prescribed M and β. The semigroup T is called *(marginally) stable* if $t \mapsto \|T(t)\|$ is bounded on $\mathbb{R}_+$. These notions are also applied to the associated generators of the semigroups.

We will study the matrix case in Chapter 3. The set of operators which generate a (M, β)-stable semigroup is described by the following Hille-Yosida Generation Theorem.

Theorem 2.6 ([38, Theorem II.3.8]). *Let A be a linear operator A on a Banach space X and let $M \geq 1$ and $\beta \in \mathbb{R}$. The following statements are equivalent.*

1. *A is the generator of a (M, β)-stable semigroup.*
2. *A is closed and densely defined, and for every real $\alpha > \beta$, α is contained in the resolvent set $\varrho(A)$ of A and the resolvent $R(\alpha, A)$ satisfies*

$$\left\|R(\alpha, A)^k\right\| \leq \frac{M}{(\alpha - \beta)^k}, \qquad k \in \mathbb{N}. \tag{2.3}$$

3. *A is closed and densely defined, and for every $s \in \mathbb{C}$ with $\operatorname{Re} s > \beta$ one has $s \in \varrho(A)$ and*

$$\left\|R(s, A)^k\right\| \leq \frac{M}{(\operatorname{Re} s - \beta)^k}, \qquad k \in \mathbb{N}. \tag{2.4}$$

Note that in order to verify (2.3) the resolvent only needs to be known on the positive half-line (β, ∞). The resolvent of a generator can be used to recover the semigroup.

Theorem 2.7 ([113, Theorem 1.8.3]). *Let A be the generator of a strongly continuous semigroup T. Then for each $x \in X$ and $t > 0$*

$$T(t)x = \lim_{k\to\infty} \left(I - \tfrac{t}{k}A\right)^{-k} x = \lim_{k\to\infty} \left(\tfrac{k}{t}R(\tfrac{k}{t}, A)\right)^k x,$$

and the limit is uniform in t on compact subsets of $\mathbb{R}_+$.

This formula is the main tool for the proof of Theorem 2.6. The term $(I - \frac{1}{h}A)^{-1}$ corresponds to an *implicit Euler step* in numerical analysis.

Example 2.8. Figure 2.1 shows the norm of the matrix exponential for $A = \left(\begin{smallmatrix} -1 & 5 & -20 \\ 0 & -10 & 75 \\ 0 & 0 & -2 \end{smallmatrix}\right)$ and of the resolvent approximations $(I - {}^t\!/\!{}_k A)^{-k}$ of e^{At} for $k = 1, 6$ and 24. ■

The resolvent is obtained from the semigroup by a Laplace transformation.

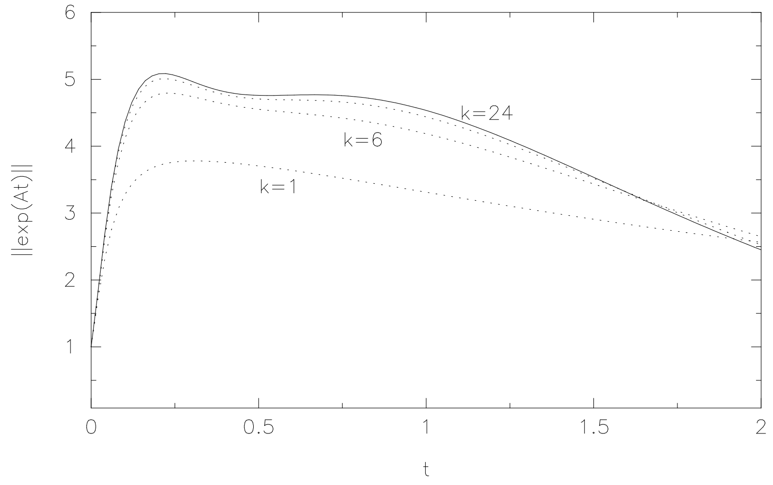

Figure 2.1: Approximation of the matrix exponential by resolvent powers.

Corollary 2.9 ([4])**.** *Let A be a (M,β)-stable generator of the semigroup T. Then for any $s \in \mathbb{C}$ with $\operatorname{Re} s > \beta$ and all $x \in X$ the map $t \mapsto e^{-st}T(t)x$ from $\mathbb{R}_+$ to X is Bochner-integrable and*

$$R(s,A)x = \int_0^\infty e^{-st}T(t)x\,dt. \tag{2.5}$$

In particular, an operator A is the generator of a strongly continuous semigroup T if and only if its resolvent $R(\cdot,A)$ is the Laplace transform of T given by (2.5)*.*

So we have three mathematical objects at our hands which can be mutually reconstructed from one of the other objects: the semigroup itself, its generator, and the resolvent of the generator. For each property of the semigroup, matching properties of the generator and the resolvent can be found. Some connections between these objects are listed in the following proposition.

Proposition 2.10 ([4, Proposition 3.1.9])**.** *Let A be the generator of a strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$ on a Banach space X. Then the following properties hold.*

(i) $\lim_{s\to\infty} sR(s,A)x = x$ *for all* $x \in X$*.*

(ii) $R(s,A)T(t) = T(t)R(s,A)$ *for all* $s \in \varrho(A)$, $t \geq 0$*.*

(iii) $x \in D(A)$ *implies that* $T(t)x \in D(A)$ *and* $AT(t)x = T(t)Ax$*.*

(iv) $\int_0^t T(s)x\,ds \in D(A)$ *and* $A\int_0^t T(s)xds = T(t)x - x$ *for all* $x \in X$ *and* $t \geq 0$*.*

(v) For every $\lambda \in \mathbb{C}$, $(e^{\lambda t}T(t))_{t\geq 0}$ *is a strongly continuous semigroup and* $A - \lambda I$ *is its generator.*

(vi) Let $x \in X$ *and* $\lambda \in \mathbb{C}$*. Then* $x \in D(A)$ *and* $Ax = \lambda x$ *if and only if* $T(t)x = e^{\lambda t}x$ *for all* $t \geq 0$*.*

We can identify the semigroup operation $t \mapsto T(t)x_0$ on X as a solution of an *abstract Cauchy problem* using the generator A,

$$\dot{x}(t) = Ax(t) \qquad \text{with initial value} \quad x(0) = x_0 \in X. \tag{2.6}$$

Proposition 2.11. *[147, Satz VII.4.7] Let A be the generator of the strongly continuous semigroup T on X, and let $x_0 \in D(A)$. Then the function $x : \mathbb{R}_+ \to X, x(t) = T(t)x_0$ is continuously differentiable with values in $D(A)$, and solves* (2.6)*. Moreover, $x(\cdot)$ is the only solution of* (2.6) *with these properties, and $x(t)$ depends continuously on the initial value x_0.*

Hence if $x_0 \in D(A)$ then $x(t, x_0) = T(t)x_0$ is a solution of (2.6). Hence $\frac{d}{dt}(T(t)x_0) = \dot{x}(t, x_0) = Ax(t, x_0) = AT(t)x_0$. Such solutions are called *classical solutions.* If $x_0 \notin D(A)$ then $x(t, x_0) = T(t)x_0$ is not necessarily differentiable anymore. Such a solution is called *mild solution* of (2.6).
We now study a special class of semigroups.

Definition 2.12. A strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$ is said to be a *contraction semigroup*, if $\|T(t)\| \leq 1$ for all $t \geq 0$. It is said to be a *strict*[1] contraction semigroup if $\|T(t)\| < 1$ for $t > 0$, and it is called a *uniform* contraction semigroup if there exists $\beta > 0$ such that $\|T(t)\| \leq e^{-\beta t}, t \geq 0$.

Note that there are strict contractions, which are not uniform contractions as the following finite dimensional example shows.

Example 2.13 ([67, Example 5.5.27 (iii)]). Consider the matrix $A = \left(\begin{smallmatrix} -1 & 2 \\ 0 & -1 \end{smallmatrix}\right)$. The spectral norm of its matrix exponential is given by

$$\left\|e^{At}\right\| = e^{-t}\left(t + \sqrt{1+t^2}\right), \tag{2.7}$$

see Proposition 4.4. The derivative of (2.7) is given by

$$\tfrac{d}{dt}\left\|e^{At}\right\| = e^{-t}\left[\left(t + \sqrt{1+t^2}\right)\left(\sqrt{1+t^2}^{\,-1} - 1\right)\right],$$

which is negative for $t > 0$ because $\sqrt{1+t^2} > 1$. As $\frac{d}{dt}\left\|e^{At}\right\| |_{t=0} = 0$, A generates a strict, but not a uniform contraction semigroup. Interestingly, the first three terms of the Taylor series of $t + \sqrt{1+t^2}$ in $t_0 = 0$ coincide with the series expansion of e^t. ■

Lemma 2.14. *For a strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$ it holds that*

$$\lim_{t\searrow 0} \tfrac{1}{t}\log\|T(t)\| = \sup_{t>0} \tfrac{1}{t}\log\|T(t)\|, \qquad \lim_{t\searrow\infty} \tfrac{1}{t}\log\|T(t)\| = \inf_{t>0} \tfrac{1}{t}\log\|T(t)\|. \tag{2.8}$$

[1] In [67], the notions of “strong” and “strict” contraction semigroups are used instead of “strict” and “uniform” contraction semigroups.

Proof. We show that the supremum and infimum of $t^{-1}f(t)$ with $f(t) = \log\|T(t)\|$ are attained at the boundaries $t = 0$ and $t = \infty$, respectively. To this end, we note that the function $f(t)$ is subadditive, as we have for all $s, t \geq 0$

$$f(t+s) \leq \log(\|T(t)\|\,\|T(s)\|) = f(t) + f(s).$$

As T is a strongly continuous semigroup, there exist $M \geq 1$ and $\beta \in \mathbb{R}$ such that $\|T(t)\| \leq Me^{\beta t}$. Hence $t^{-1}f(t) \leq \beta + t^{-1}\log(M)$. Now [59, Theorem 7.6.1] gives the second equality in (2.8). For the first equality, compare with [59, Theorem 7.11.1]. □

Example 2.15. Note that the function $g : t \mapsto t^{-1}\log\|T(t)\|$ is in general not a monotone decreasing function. To see this, let us consider the following matrix in real Schur form,

$$A = \begin{pmatrix} -1 & -100 & 0 & -150 & 0 & 200 & -1000 \\ 1 & -1 & 1 & -10 & 25 & 11 & -200 \\ & & -1 & 400 & -30 & 0 & 250 \\ & & -1 & -1 & 5 & 5 & 200 \\ & & & & -1 & -2 & 30 \\ & & & & & -1 & -625 \\ & & & & & 1 & -1 \end{pmatrix},$$

where empty entries are filled with zeros. A MAPLE-based computation returns $\left\|e^{2.5A}\right\| = 0.8395$ and $\left\|e^{3A}\right\| = 30.54$. Hence ${}^{\log(0.8395)}/_{2.5} = -0.06998 < 1.13971 = {}^{\log(30.54)}/_{3}$, so that g is not monotonically decreasing. The transient behaviour $t \mapsto \left\|e^{At}\right\|$ is depicted in Figure 7.3. ■

Lemma 2.14 motivates the following definitions.

Definition 2.16. For a strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$, the *initial growth rate* of T is given by $\alpha_0(T) := \lim_{t\searrow 0}\frac{1}{t}\log\|T(t)\|$, the *asymptotic growth rate* of T is given by $\omega_0(T) := \lim_{t\to\infty}\frac{1}{t}\log\|T(t)\|$.

The notation ω_0 is standard, to match this symbolism α_0 is introduced as the initial growth rate. From Lemma 2.14 we immediately get $\alpha_0(A) \geq \omega_0(A)$. Moreover the following characterizations are available for the initial and asymptotic growth rates.

Corollary 2.17. *Let T be a strongly continuous semigroup. Then*

$$\alpha_0(T) = \inf\left\{\beta \in \mathbb{R} \,\middle|\, \textit{for all}\, t \geq 0,\ \|T(t)\| \leq e^{\beta t}\right\}, \tag{2.9}$$

$$\omega_0(T) = \inf\left\{\beta \in \mathbb{R} \,\middle|\, \textit{there exists } M \geq 1 \textit{ such that for all}\, t \geq 0,\ \|T(t)\| \leq Me^{\beta t}\right\}. \tag{2.10}$$

Proof. We only show (2.9) as (2.10) can be found in [38, Proposition IV.2.2]. By Lemma 2.14, we have $\alpha_0(T) = \sup_{t>0}\frac{1}{t}\log\|T(t)\|$. If $\beta \in \mathbb{R}$ is such that $\|T(t)\| \leq e^{\beta t}$ for all $t \in \mathbb{R}_+$ then $\alpha_0(T) \leq \sup_{t>0}\frac{1}{t}\log e^{\beta t} = \beta$. Thus $\alpha_0(T) \leq \inf\{\beta \in \mathbb{R} \,|\, \forall t > 0, \|T(t)\| \leq e^{\beta t}\}$. Let us now consider the semigroup $(S(t))_{t\in\mathbb{R}_+} = (e^{-\alpha_0(T)t}T(t))_{t\in\mathbb{R}_+}$. This is a contraction semigroup, as

$$\begin{aligned} \|S(t)\| &= e^{-\alpha_0(T)t}\|T(t)\| = \exp\left(\log\|T(t)\| - \alpha_0(T)t\right) \\ &= \exp\left(t\left(\tfrac{1}{t}\log\|T(t)\| - \alpha_0(T)\right)\right) \leq e^0 = 1. \end{aligned}$$

Hence $\alpha_0(T) \geq \inf\{\beta \in \mathbb{R} \,|\, \forall t > 0, \|T(t)\| \leq e^{\beta t}\}$ which shows (2.9). □

With respect to our stability investigations we now have the following results.

Corollary 2.18. *A strongly continuous semigroup T is exponentially stable if and only if $\omega_0(T) < 0$. It is a contraction semigroup if $\alpha_0(T) \leq 0$, and it is a uniform contraction semigroup if $\alpha_0(T) < 0$. If T is a uniform contraction semigroup which satisfies $\|T(t)\| \leq e^{\beta t}$, $t \geq 0$, then $(e^{-\beta t}T(t))_{t\in\mathbb{R}_+}$ is a contraction semigroup.*

A semigroup is uniformly continuous if and only if its generator is bounded. For unbounded generators the initial growth rate might be $+\infty$ if $\|T(s)\| > \gamma > 1$ for $s \in (0, \delta)$, and $\delta > 0$ small. An example showing such behaviour will be presented in Example 2.38.
The relation between the growth rates of T and properties of the generator A is studied in the following theorem. We only consider the case of bounded generators.

Theorem 2.19. *Let A be a bounded linear operator on a Banach space X, and $(T(t))_{t\in\mathbb{R}_+}$ be the uniformly continuous semigroup generated by A. Define*

$$\alpha(A) = \sup\{\operatorname{Re}\lambda \mid \lambda \in \sigma(A)\}, \qquad \mu(A) = \lim_{h\searrow 0} \tfrac{1}{h}(\|I + Ah\| - 1). \tag{2.11}$$

Then the following holds,

$$\alpha_0(T) = \mu(A) \geq \alpha(A) = \omega_0(T). \tag{2.12}$$

Proof. We have $e^{At}x = T(t)x$ for all $x \in X$ and all $t \in \mathbb{R}_+$. Then by Lemma 2.14

$$\omega_0(T) = \inf_{t>0} \frac{1}{t} \log \|T(t)\| = \inf_{t>0} \frac{1}{t} \log \left\|e^{At}x\right\|, \qquad x \in X,\ \|x\| = 1.$$

Now we need to know that $\alpha(A) = \inf_{t>0} \frac{1}{t} \left\|e^{At}\right\|$. Gelfand's formula for the spectral radius

$$\rho(A) := \sup\{|\lambda| \mid \lambda \in \sigma(A)\} = \lim_{k\to\infty} \sqrt[k]{\|A^k\|} \tag{2.13}$$

applied to $T(t)$ gives together with Lemma 2.14

$$\rho(T(t)) = \lim_{k\to\infty} \|T(kt)\|^{1/k} = e^{t \lim_{k\to\infty}(kt)^{-1}\log\|T(kt)\|} = e^{\omega_0(T)t}.$$

For bounded generators, the Spectral Mapping Theorem ([38, Lemma I.3.13]) yields

$$\left\{e^{\lambda t} \mid \lambda \in \sigma(A)\right\} = \sigma(T(t)).$$

Therefore there exists an eigenvalue λ_0 of A such that the spectral radius $\rho(T(t))$ of the semigroup T satisfies

$$\rho(T(t)) = e^{\omega_0(T)t} = \left|e^{\lambda_0 t}\right| = e^{\operatorname{Re}\lambda_0 t} = e^{\alpha(A)t}.$$

Hence $\omega_0(T) = \alpha(A)$. For the initial growth rate we have by Lemma 2.14

$$\begin{aligned}\alpha_0(T) &= \sup_{t>0} \tfrac{1}{t} \log \left\|e^{At}\right\| = \lim_{t\searrow 0} \tfrac{1}{t} \log \left\|e^{At}\right\| = \lim_{t\searrow 0} \tfrac{1}{t} \log(\|I + At\|) \\ &= \lim_{t\searrow 0} \tfrac{1}{t}(\|I + At\| - 1) = \mu(A).\end{aligned}$$

Here we approximated e^{At} by the linear part of its Taylor series, $I + At$ and used $\log(1 + r) \approx r$ for $|r|$ small. The inequality between $\alpha_0(T)$ and $\omega_0(T)$ in (2.12) follows from Lemma 2.14. □

However, if A is unbounded then the domain of A is only a subset of X, and only the inequalities $\omega_0(T) \geq \alpha(A)$ and $\alpha_0(T) \leq \mu(A)$ hold. We study a counterexample which illustrates one of these gaps in the following subsection.

2.2 Asymptotic Growth Rates

Let us now turn our attention to the asymptotic growth rate of a strongly continuous semigroup T with generator A. In in section we describe a method of constructing strongly continuous semigroups with $\omega_0(T) > \alpha(A)$. The spectral radii of the semigroup operators $T(t)$ are connected to the asymptotic growth rate of T via

$$\rho(T(t)) = e^{\omega_0(T)t}, \qquad t > 0,$$

which we already used in the proof of Theorem 2.19. This proof is also valid for unbounded generators, see [38, Proposition IV.2.2]. For stability analysis one likes to connect the spectrum of the generator A with the asymptotic growth rate, however one only obtains the following result.

Theorem 2.20 ([38, Theorem IV.3.6]). *The spectrum of a strongly continuous semigroup T and the spectrum of its generator A satisfy*

$$e^{\sigma(A)t} \subset \sigma(T(t)) \qquad \textit{for all}\, t > 0. \tag{2.14}$$

More precisely, the following inclusions hold for all $t \geq 0$,

$$e^{\sigma_P(A)t} \subset \sigma_P(T(t)), \quad e^{\sigma_C(A)t} \subset \sigma_C(T(t)), \quad e^{\sigma_R(A)t} \subset \sigma_R(T(t)).$$

We will demonstrate that there exist semigroups where the generator has only a point spectrum and which yield a proper subset in the spectral inclusion (2.14). An example with $\omega_0(T) \neq \alpha(A)$ is due to Zabczyk [152], see also Trefethen [137]. In the following we present a detailed analysis of this example. For a given sequence of natural numbers $(n_k)_{k\in\mathbb{N}}$ with $n_k \geq 1$ we consider the Hilbert direct sum of the spaces $\mathbb{C}^{n_k}$, denoted by $X = \bigoplus_{k\in\mathbb{N}} \mathbb{C}^{n_k}$. We now investigate semigroups on X.

Proposition 2.21. *If $(T_k(t))_{t\in\mathbb{R}_+}$, $k \in \mathbb{N}$, are strongly continuous semigroups on X_k with generator A_k and for each $t \in \mathbb{R}_+$ the sequence $(\|T_k(t)\|)_{k\in\mathbb{N}}$ is bounded then $T = \bigoplus_{k\in\mathbb{N}} T_k$ is a strongly continuous semigroup on $X = \bigoplus_{k\in\mathbb{N}} X_k$. Its generator is given by $A = \bigoplus_{k\in\mathbb{N}} A_k$ and has the domain $D(A) = \bigoplus_{k\in\mathbb{N}} D(A_k) \subset X$.*

Proof. Let us verify that T is a strongly stable semigroup. We have for $s, t \geq 0$

$$\begin{aligned} T(s+t) &= \bigoplus_{k\in\mathbb{N}} T_k(s+t) = \bigoplus_{k\in\mathbb{N}} T_k(s)T_k(t) = \big(\bigoplus_{k\in\mathbb{N}} T_k(s)\big)\big(\bigoplus_{k\in\mathbb{N}} T_k(t)\big) = T(s)T(t), \\ T(0) &= \bigoplus_{k\in\mathbb{N}} T_k(0) = \bigoplus_{k\in\mathbb{N}} I_{X_k} = I_X. \end{aligned}$$

Thus T is a semigroup. By boundedness, $T(t) \in \mathcal{L}(X)$ for all $t \in \mathbb{R}_+$. Moreover, for all $x = (x^k)_k \in X$

$$\lim_{t \searrow 0} (T(t)x - x) = \lim_{t \searrow 0} (T_k(t)x^k - x^k)_k = 0$$

shows that T is strongly continuous. The domain of its generator A is given by

$$D(A) = \left\{ x \in X \,\middle|\, \lim_{h \searrow 0} \tfrac{1}{h}(T(h)x - x) \text{ exists} \right\} = \left\{ x = (x^k)_k \,\middle|\, x^k \in D(A_k) \right\} = \bigoplus_{k \in \mathbb{N}} D(A_k)$$

and clearly, $A = \bigoplus_{k \in \mathbb{N}} A_k$. □

Note that for stable $A_k \in \mathbb{C}^{n_k \times n_k}$, $(T_k(t))_{t \geq 0} = (e^{A_k t})_{t \geq 0}$ are uniformly continuous semigroups, however $T = \bigoplus_{k \in \mathbb{N}} T_k$ is generally only strongly continuous.
If the following condition is satisfied then the spectral abscissa $\alpha(A) = \{\operatorname{Re} s \,|\, s \in \sigma(A)\}$ of the generator A and the asymptotic growth rate of the semigroup do *not* coincide.

Theorem 2.22. *Let A be a closed and densely defined linear operator on a Hilbert space X that generates the semigroup $(T(t))_{t \in \mathbb{R}_+}$. If the limit of the ε-pseudospectral abscissas satisfies*

$$\lim_{\varepsilon \to 0} \alpha_\varepsilon(A) > \alpha(A) \tag{2.15}$$

then the asymptotic growth rate of the semigroup T satisfies $\alpha(A) < \omega_0(T)$.

Proof. Let us suppose that the asymptotic growth rate of the semigroup $(T(t))_{t \in \mathbb{R}_+}$ generated by A satisfies $\omega_0(T) < \tilde{\alpha} := \lim_{\varepsilon \to 0} \alpha_\varepsilon(A)$. If $\beta \in (\omega_0(T), \tilde{\alpha})$ then for every $\varepsilon > 0$ there exists $\omega \in \mathbb{R}$ such that the resolvent of $\beta + i\omega$ satisfies $\|R(\beta + i\omega, A)\| > \varepsilon^{-1}$. Thus the resolvent $R(\cdot, A)$ is unbounded on $\beta + i\mathbb{R}$, hence (2.4) of Theorem 2.6 does not hold. Therefore $\omega_0(T) \geq \tilde{\alpha} > \alpha(A)$. □

Remark 2.23. For block-diagonal operators $A = \bigoplus_{k \in \mathbb{N}} A_k$ we have $R(s, A) = \bigoplus_{k \in \mathbb{N}} R(s, A_k)$ and $\|R(s, A)\| = \sup_{k \in \mathbb{N}} \|R(s, A_k)\|$. Therefore $\sigma_\varepsilon(A) = \bigcup_{k \in \mathbb{N}} \sigma_\varepsilon(A_k)$, hence $\alpha_\varepsilon(A) = \sup_{k \in \mathbb{N}} \alpha_\varepsilon(A_k)$. If there exists sequence of matrices for which $\alpha(A_k) \equiv \alpha$ is constant and which satisfies, say, $\alpha_{1/k}(A_k) > \alpha + \frac{1}{2}$, then this sequence can be used to construct an operator which satisfies Theorem 2.22.

Our choice of the matrix sequence $(A_k)_{k \in \mathbb{N}}$ such that $A = \bigoplus_{k \in \mathbb{N}^*} A_k$ satisfies Theorem 2.22 will consist of Jordan blocks of growing dimensions.

Remark 2.24. We can regard these Jordan blocks as finite dimensional approximants of an $\ell^2(\mathbb{C})$-Toeplitz-operator which has a continuous spectrum, see [21, 20]. Let us consider the Toeplitz operator $J_\infty(\lambda) : \ell^2(\mathbb{C}) \to \ell^2(\mathbb{C})$, $(x_k) \mapsto (\lambda x_k + x_{k+1})$ with $\lambda \in \mathbb{C}$ which is composed of a multiplication operator and a shift operator on $\ell^2(\mathbb{C})$. Its spectrum consists of all points which are enclosed by $\{a(e^{i\varphi}) \,|\, \varphi \in [0, 2\pi]\}$ with a winding number of 1 where $a(t) = \lambda + t$ is the *symbol* belonging to the Toeplitz operator $J_\infty(\lambda)$. Here the spectrum is a disc of radius 1 centered at λ. A finite dimensional approximation of $J_\infty(\lambda)$ is given by the Jordan block $J_n(\lambda)$ of dimension n.

We will derive properties of Jordan blocks by a direct analysis. Let J_n be the Jordan block of size n for the eigenvalue 0 and set $J_n(\lambda) = \lambda I_n + J_n$. For an estimate of the norm of the resolvent of J_n we use the Neumann series which consists only of finitely many terms since J_n is nilpotent,

$$(sI_n - J_n)^{-1} = s^{-1} \sum_{k=0}^{n-1} \left(\frac{J_n}{s} \right)^k ,$$

which is valid for all $s \neq 0$. Taking norms we have $\left\| J_n^k \right\| = 1$ for $k < n$ and therefore

$$\left\| (sI_n - J_n)^{-1} \right\| \leq |s|^{-1} \sum_{k=0}^{n-1} \frac{\left\| J_n^k \right\|}{|s|^k} = |s|^{-1} \frac{|s|^{-n} - 1}{|s|^{-1} - 1} = \frac{1 - |s|^{-n}}{|s| - 1}, \qquad |s| \neq 0, 1.$$

Now consider the set $\{\lambda \in \mathbb{C} \mid |\lambda| = 1 - \frac{1}{n+1}\}$, $n \in \mathbb{N}^*$. For each of its elements λ we have $\|(\lambda I - J_n)^{-1}\| \leq (n+1) \left[\left(1 - \frac{1}{n+1}\right)^{-n} - 1 \right]$. Note that $(1 - \frac{1}{n+1})^{-n} - 1 \in [1, e-1]$. Hence for every $n \in \mathbb{N}^*$ there exists $\varepsilon > 0$ small enough such that $\sigma_\varepsilon(J_n) \subset \{\lambda \in \mathbb{C} \mid |\lambda| < 1 - \frac{1}{n+1}\}$. Finally we present the construction of an example illustrating the gap between $\alpha(A)$ and $\omega_0(T)$.

Example 2.25. Let us consider the diverging sequence $x_k = 2ik$ and set $A_k = J_k(x_k)$ for $k = 1, 2, \ldots$ We show that the associated block multiplication operator $A = \bigoplus A_k$ is unbounded and has only a discrete spectrum given by $\sigma(A) = \bigcup_{k=1}^\infty \sigma(A_k)$. To this end, note that $(x_k)_{k \in \mathbb{N}^*}$ is an unbounded sequence. Therefore, A is an unbounded operator. Moreover, by the results derived above we see that the spectrum of $\sigma(A) = \bigcap_{\varepsilon > 0} \bigcup_{k \in \mathbb{N}^*} \sigma_\varepsilon(A_k)$, see Corollary 1.18, is just $\bigcup_{k=1}^\infty \sigma(A_k)$. For this, note that the ε-pseudospectra of A_k are disjoint for $\varepsilon > 0$ small enough. In particular, for every $k \in \mathbb{N}^*$ there exists an $\varepsilon > 0$ such that

$$\sigma_\varepsilon(A_k) = \sigma_\varepsilon(x_k I + J_k) \subset \{\lambda \in \mathbb{C} \mid |\lambda - x_k| < 1 - \tfrac{1}{k+1}\}.$$

Hence the spectrum of the operator A consists only of a point spectrum. Now by Proposition 2.21, A generates the strongly continuous semigroup T which is again of block diagonal form

$$T(t) = \bigoplus_{k=1}^{\infty} e^{2i\,kt} e^{J_k t}, \qquad t \in \mathbb{R}_+.$$

Each row of $e^{J_k t}$ contains the first terms of a Taylor expansion of the exponential series e^t. Now consider the sequence $(x_k)_{k \in \mathbb{N}} = ((0_1, 0_2, \ldots 0_{k-1}, \mathbf{1}_k, 0_{k+1}, \ldots))_{k \in \mathbb{N}} \subset X$ where $\mathbf{1}_k = (1 \ldots 1)^\top \in \mathbb{R}^k$ matches an A_k block. Then $T(x_k)_{k \in \mathbb{N}} = (0, \ldots, A_k 1_k$ yields

$$\omega_0(T) \geq \lim_{t \to \infty} \sup_{k \in \mathbb{N}} t^{-1} \log \|T(t) x_k\| / \|x_k\| = \lim_{t \to \infty} \sup_{k \in \mathbb{N}} t^{-1} \log \left\| e^{A_k t} \mathbf{1}_k \right\| = \lim_{t \to \infty} t^{-1} \log e^t = 1.$$

The asymptotic growth rate is therefore at least 1, whereas the spectral abscissa $\alpha(A) = \sup_{\lambda \in \sigma(A)} \operatorname{Re} \lambda = \sup_{k \in \mathbb{N}} \alpha(A_k) = 0$. ■

2.3 Initial Growth Rates

Let X be a Banach space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ with norm $\|\cdot\|$. The *dual space* $X^* = \mathfrak{L}(X, \mathbb{K})$ is the set of all bounded linear functionals which map X into $\mathbb{K}$. Let us denote the value of the functional $y \in X^*$ in $x \in X$ by $\langle y, x\rangle \in \mathbb{K}$. The dual space is a Banach space equipped with the *dual norm*

$$\|y\|^* = \sup \{|\langle y, x\rangle| \mid \|x\| \leq 1\} = \sup \{\operatorname{Re} \langle y, x\rangle \mid \|x\| \leq 1\} . \tag{2.16}$$

Indeed, this is the operator norm for linear functionals $\langle y, \cdot\rangle : (X, \|\cdot\|) \to (\mathbb{K}, |\cdot|)$.

Definition 2.26. We call $(x, y) \in X \times X^*$ a *dual pair (DP)* if $\|x\| \|y\|^* = \langle y, x\rangle \neq 0$ and we speak of a *normed dual pair (NDP)* if a dual pair (x, y) satisfies $\langle y, x\rangle = 1$. If $\|x\| = 1$, $\|y\|^* = 1$ and $\langle y, x\rangle = 1$ then (x, y) is a *unitary dual pair (UDP).* We call $y \in X^*$ a *dual vector* of $x \in X$ if (x, y) is a dual pair.

By the Hahn-Banach Theorem, the set of dual vectors $y \in X^*$ of a given $x \in X$ is never empty. We collect some properties of dual pairs in the following proposition.

Proposition 2.27. *Let X be a Banach space and denote its dual space by X^*. For $x \in X$ and $y, y' \in X^*$ we have*

(i) If (x, y) is a dual pair then $(\alpha x, \beta y)$ is a dual pair for all $\alpha, \beta > 0$. Moreover, if X is a Banach space over $\mathbb{C}$ then $(\zeta^{-1}x, \zeta y)$ is a dual pair for all $\zeta \in \mathbb{C}$, $\zeta \neq 0$.

(ii) In a reflexive Banach space X, if (x, y) is a dual pair then (y, x) is a dual pair in X^*.

(iii) If (x, y) and (x, y') are dual pairs then $(x, \theta y + (1-\theta)y')$ are dual pairs for all $\theta \in (0, 1)$.

(iv) Every dual pair (x, y) with $\|y\|^* = 1$ satisfies the subgradient inequality

$$\textit{for all} \quad z \in X : \qquad \|x + z\| \geq \|x\| + \operatorname{Re} \langle y, z\rangle. \tag{2.17}$$

Proof. Item *(i)* follows directly from Definition 2.26 and properties of the norm. For *(ii)*, recall that a reflexive Banach space X is isomorphic to its bidual X^{**} via $z \mapsto \hat{z}$, $\hat{z} : y \mapsto \langle y, z\rangle$. Hence there exists an isomorphism between dual pairs $(y, \hat{z}) \in X^* \times X^{**}$ and dual pairs $(z, y) \in X \times X^*$. To prove *(iii)* note that by definition of the dual norm (2.16), $|\langle u, x\rangle| \leq \|x\| \|u\|^*$ for all $x \in X$, $u \in X^*$. If (x, y) and (x, y') form dual pairs then let us consider $u = \theta y + (1 - \theta)y' \in X^*$ for $\theta \in [0, 1]$. We have

$$\begin{aligned} \|x\| \|u\|^* &\geq \langle u, x\rangle = \theta \|x\| \|y\|^* + (1 - \theta) \|x\| \|y'\|^* = \|x\| \left(\theta \|y\|^* + (1 - \theta) \|y'\|^*\right) \\ &\geq \|x\| \|\theta y + (1 - \theta)y'\|^* = \|x\| \|u\|^* . \end{aligned} \tag{2.18}$$

Thus equality holds in (2.18), and therefore u is a dual vector of x. For item *(iv)*, consider the unitary dual pair (x, y). If we have a pair $(u, y) \in X \times X^*$ which only satisfies $\|u\| = 1$ and $\|y\|^* = 1$ then $\operatorname{Re} \langle y, u\rangle \leq 1$. Setting $u = \frac{x+z}{\|x+z\|} \in X$ for $z \in X$, $z \neq -x$ gives

$$1 = \langle y, x\rangle \geq \operatorname{Re} \langle y, u\rangle = \operatorname{Re} \left\langle y, \frac{x + z}{\|x + z\|} \right\rangle \qquad \text{for all } z \in X \setminus \{-x\}.$$

By multiplication with $\|x+z\|$ we obtain (2.17). The case $z=-x$ is directly verified, $\|x\| + \operatorname{Re}\langle y, z\rangle = \|x\| - \langle y, x\rangle = 0$. □

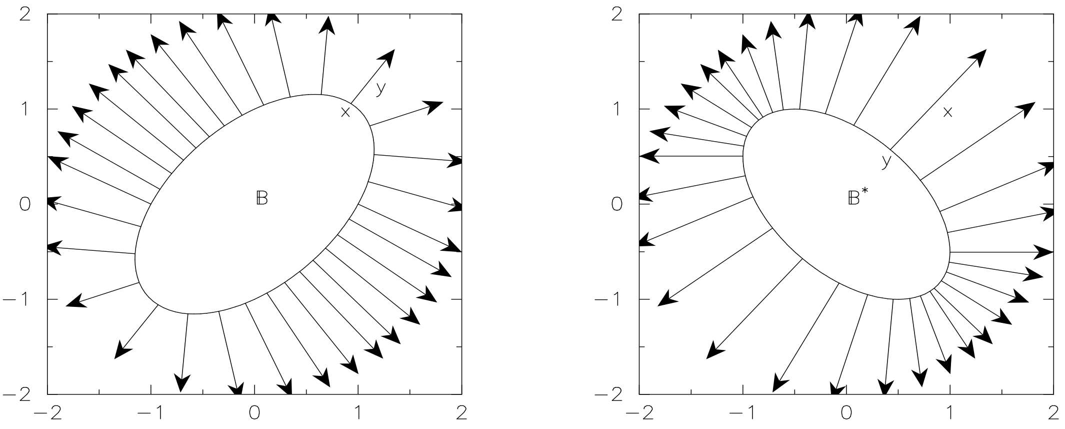

Figure 2.2: Dual pairs and dual norms.

In terms of convex analysis [120], equation (2.17) shows that every dual vector y of x with $\|y\|^* = 1$ is a *subgradient* of the norm $\|\cdot\|$ at the point x. More precisely, Proposition 2.27 *(iv)* implies that (x, y) is a DP if and only if the hyperplane $\{z \in X \mid \langle y, z\rangle = \|x\|\,\|y\|^*\}$ is a supporting hyperplane in x of the ball $\mathbb{B}(r) = \{z \in X \mid \|z\| \le r\}$ with radius $r = \|x\|$. If (x, y) is a UDP then y is an outer normal of $\mathbb{B}$ in x. We demonstrate this property in the following examples.

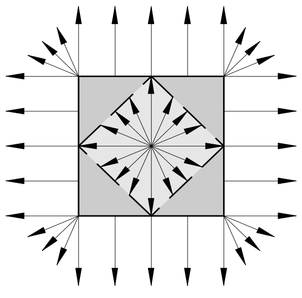

Figure 2.3: Duals of the ∞-norm.

We visualize some dual norms and dual pairs for norms in $\mathbb{R}^2$. Consider a symmetric positive definite matrix $P \in \mathbb{R}^{2\times 2}$. Then $\|x\|_P = \sqrt{x^\top P x}$ defines a norm on $\mathbb{R}^2$. Its dual norm is given by

$$\|y\|_P^* = \max\{\langle y, x\rangle_2 \mid \langle x, Px\rangle_2 = 1\} = \|y\|_{P^{-1}} \qquad (2.19)$$

as $\langle y, x\rangle_2$ is maximal on the unit sphere of $(\mathbb{R}^2, \|\cdot\|_P)$ for $x = \langle y, P^{-1}y\rangle_2^{-1/2} P^{-1} y$ with $\|x\|_P = 1$. Unitary dual pairs are given by (x, Px) with $\|x\|_P = 1$, because these vectors satisfy $1 = \|x\|_P = \langle Px, x\rangle_2 = \langle (Px), P^{-1}(Px)\rangle_2 = \|Px\|_{P^{-1}}^2$. Hence $\|Px\|_{P^{-1}} = 1$. Figure 2.2 shows the unit ball $\mathbb{B}$ and the dual unit ball $\mathbb{B}^*$ of $\|\cdot\|_P$ where $P = \left(\begin{smallmatrix} 1 & -.5 \\ -.5 & 1 \end{smallmatrix}\right)$. We see that if (x, y) is a UDP then y is an outer normal of $\mathbb{B}$ in x. Moreover, the right image shows that $\partial\mathbb{B}^*$ collects all possible dual vectors for all $x \in \partial\mathbb{B}$ so that the pair (y, x) is a unitary dual pair with respect to the dual norm.

Example 2.28. Let us now consider the pair of dual norms $\|x\|_1 = |x_1| + |x_2|$ and $\|x\|_\infty = \max\{|x_1|, |x_2|\}$. For $x \in \{(\pm 1, x_2)$ with $|x_2| < 1$ or $(x_1, \pm 1)$ with $|x_1| < 1\}$ its dual vector is uniquely determined by $y = (\pm 1, 0)$ or $y = (0, \pm 1)$.

However, for $x = (\pm 1, \pm 1)$ the duals are not uniquely determined, see Figure 2.3 for an illustration. The unit box $\mathbb{B}_\infty$ is shaded gray, and in the vertices the dual pairs are not unique. Attaching all these dual vectors to the origin gives the unit ball of $\|\cdot\|_1$. ■

For a given closed linear operator A in X consider $x \in D(A)$ with $\|x\| = 1$. We are interested in the direction in which Ax points with respect to the unit ball of $\|\cdot\|$. This motivates the following definition.

Definition 2.29. Let A be a closed linear operator on X. The *initial growth* of A in $x \in D(A)$ is given by

$$\mu(x, A) = \lim_{h \searrow 0} \tfrac{1}{h} \left(\|x + hAx\| - \|x\|\right). \tag{2.20}$$

For a closed linear operator A, we call $\mu(A) = \sup\{\mu(x, A) \,|\, x \in D(A), \|x\| \leq 1\}$ the *initial growth rate* of A. The closed linear operator A is called *dissipative* if $\mu(A) \leq 0$, it is called strictly dissipative if $\mu(A) < 0$.

We can rewrite $\lim_{h\searrow 0} \frac{1}{h} (\|x + hAx\| - \|x\|)$ as $\lim_{t\to\infty} (\|tx + Ax\| - \|tx\|)$. Then for $s, t \geq 0$,

$$\|(s+t)x + Ax\| - (s+t)\|x\| \leq \|sx + Ax\| + (t - (s+t))\|x\| = \|sx + Ax\| - s\|x\|.$$

Thus the term $\|tx + Ax\| - t\|x\|$ is monotonically decreasing in t. Additionally, $\|tx + Ax\| - \|tx\| \geq -\|Ax\|$ for all $t > 0$, hence the limit in (2.20) exists. We therefore have

$$\mu(x, A) = \inf_{s>0} \|(sI_X + A)x\| - s\|x\|. \tag{2.21}$$

The term "initial growth rate" is slightly misleading, as we have not assumed that A is a generator of a semigroup. However, if A is the generator of a uniformly continuous semigroup the following lemma shows that we regain the initial growth rate used in Theorem 2.19.

Lemma 2.30. *If $A \in \mathcal{L}(A)$ is the generator of a uniformly continuous semigroup then $\mu(A) = \lim_{h\searrow 0} h^{-1}(\|I + Ah\| - 1)$ is the initial growth rate of the semigroup $(e^{At})_{t\in\mathbb{R}_+}$.*

Proof. If A is the generator of a uniformly continuous semigroup, then $D(A) = X$ and

$$\begin{aligned}\sup_{\|x\|=1} \mu(x, A) &= \sup_{\|x\|=1} \lim_{t\to\infty} (\|(tI + A)x\| - t\|x\|) \\ &= \lim_{t\to\infty} \sup_{\|x\|=1} (\|(tI + A)x\| - t\|x\|) = \lim_{t\to\infty} (\|A + It)\| - t) = \mu(A)\end{aligned}$$

as the limit $\lim_{t\to\infty}(\|(tI + A)x\| - t)$ is monotone in t for all $x \in X$, $\|x\| = 1$. Therefore the initial growth rate of a generator A with $D(A) = X$ also satisfies (2.11). □

The following result connects Definition 2.29 with the discussion of dual vectors.

Proposition 2.31. *Given a closed linear operator A in X. Then for all $x \in D(A)$, $x \neq 0$,*

$$\mu(x, A) = \sup\left\{ \operatorname{Re} \frac{\langle y, Ax\rangle}{\langle y, x\rangle} \,\middle|\, y \in X^* \textit{ is a dual vector of } x \right\}. \tag{2.22}$$

Hence the initial growth rate of A satisfies

$$\begin{aligned}\mu(A) &= \sup\left\{ \frac{\operatorname{Re}\langle y, Ax\rangle}{\langle y, x\rangle} \,\middle|\, x \in D(A) \textit{ and } (x, y) \textit{ DP} \right\} \\ &= \sup\left\{\operatorname{Re}\langle y, Ax\rangle \,|\, x \in D(A) \textit{ and } (x, y) \textit{ NDP}\right\}.\end{aligned}$$

Proof. Let us denote the right hand side of (2.22) by $\tilde{\mu}(x, A)$. We first show that $\tilde{\mu}(x, A) \leq \mu(x, A)$. Let y be a dual vector of x with $\|y\|^* = 1$. From (2.17) we have for all $h > 0$ that $\|x\| + h\mathrm{Re}\,\langle y, Ax\rangle = \langle y, x\rangle + h\mathrm{Re}\,\langle y, Ax\rangle = \mathrm{Re}\,\langle y, x + hAx\rangle \leq \|x + hAx\|$, which implies that $\mathrm{Re}\,\langle y, Ax\rangle \leq \frac{1}{h}\left(\|x + hAx\| - \|x\|\right)$ for all $h > 0$, hence

$$\tilde{\mu}(x, A) \leq \lim_{h \searrow 0} \tfrac{1}{h}\left(\|x + hAx\| - \|x\|\right). \tag{2.23}$$

To see the converse inequality $\tilde{\mu}(x, A) \geq \mu(x, A)$ let us fix $x \in D(A)$ with $\|x\| = 1$. Then for all $t > 0$ there exists $y_t \in X^*$ such that $\mathrm{Re}\,\langle y_t, (tI + A)x\rangle = \|(tI + A)x\|$ and $\|y_t\|^* = 1$. With $\mu(x, A) = \inf\{\|sx + Ax\| - s\,\|x\| \mid s > 0\}$, see (2.21), the following inequalities are valid for all $t > 0$

$$\begin{aligned} \mu(x, A) + t \leq \|(tI + A)x\| &= t\,\mathrm{Re}\,\langle y_t, x\rangle + \mathrm{Re}\,\langle y_t, Ax\rangle \\ &\leq \min\{t + \mathrm{Re}\,\langle y_t, Ax\rangle,\, t\,\mathrm{Re}\,\langle y_t, x\rangle + \|Ax\|\}. \end{aligned}$$

Hence $1 + \frac{1}{t}(\mu(x, A) - \|Ax\|) \leq \mathrm{Re}\,\langle y_t, x\rangle \leq 1$ and $\mu(x, A) \leq \mathrm{Re}\,\langle y_t, Ax\rangle$. Now the unit ball of X^* is compact in the weak* topology of X^*, hence there exists a weak* accumulation point of $(y_t)_{t\in\mathbb{R}_+}$ for $t \to \infty$ named y'. This accumulation point satisfies $\mathrm{Re}\,\langle y', x\rangle = 1$. Hence

$$\|y'\|^* \leq 1, \quad \mathrm{Re}\,\langle y', x\rangle = 1, \qquad \mathrm{Re}\,\langle y', Ax\rangle \geq \mu(x, A).$$

But this already implies that $\|y'\|^* = 1$ and $\langle y', x\rangle = 1$. Thus (x, y') is a normed dual pair with $\langle y', x\rangle = 1$ and $\mu(x, A) \leq \mathrm{Re}\,\langle y', Ax\rangle$. This shows $\tilde{\mu}(x, A) \geq \mu(x, A)$. □

Hence a contraction semigroup has a dissipative generator, and a dissipative generator corresponds to a contraction semigroup. Let us now consider a different characterization of dissipativity.

Lemma 2.32 ([147, Theorem VII.4.15])**.** *A closed linear operator A on X is dissipative if and only if for all $x \in D(A)$ and all $\lambda > 0$,*

$$\|(\lambda I - A)x\| \geq \lambda\,\|x\|\,. \tag{2.24}$$

In particular, if $(0, \infty)$ is contained in the resolvent set of A then A is dissipative if and only if for all $x \in X$ and all $\lambda > 0$: $\|\lambda R(\lambda, A)x\| \leq \|x\|$. This characterises dissipativity in terms of the resolvent, which we analyse further by stating a version of the theorem of Hille-Yosida for contractions.

Theorem 2.33 ([147, Theorem VII.4.11])**.** *The closed linear operator A is the generator of a contraction semigroup $(T(t))_{t\in\mathbb{R}_+}$ on X if and only if A is closed and densely defined, every $\lambda > 0$ is contained in the resolvent set of A and satisfies $\|\lambda R(\lambda, A)\| \leq 1$.*

Hence every generator of a contraction semigroup is dissipative. However for the converse implication we need that the positive real half-line is contained in the resolvent set of A which is not enforced by (2.24). The question when a dissipative operator is also the generator of a contraction semigroup is answered by a famous theorem of Lumer and Phillips.

Theorem 2.34 ([4, Theorem 3.4.5])**.** *Suppose that A is a densely defined closed linear operator on a Banach space X. Then A is a generator of a strongly continuous contraction semigroup $(T(t))_{t\in\mathbb{R}_+}$ if and only if A is dissipative and the range $(\lambda I - A)[D(A)] = X$ for some $\lambda > 0$.*

Pazy [113, Corollary 1.4.4] notes the following corollary.

Corollary 2.35. *Suppose that A is a densely defined closed linear operator on a Banach space X. If both A and its adjoint A^* are dissipative then A generates a contraction semigroup.*

If X is a Hilbert space then it can be identified with its dual $X \simeq X^*$. In particular, $\|x\|^2 = \langle x, x\rangle$ for all $x \in X$ so that each $x \in X$ has a uniquely determined dual, namely, x itself.

Lemma 2.36. *Suppose that X is a Hilbert space. The initial growth rate of a bounded linear operator $A \in \mathfrak{L}(X)$ is given by*

$$\mu(A) = \tfrac{1}{2}\alpha(A + A^*).$$

Proof. By Proposition 2.31 we have

$$\mu(A) = \sup_{\|x\|=1} \operatorname{Re}\langle x, Ax\rangle = \tfrac{1}{2} \sup_{\|x\|=1} \langle x, (A + A^*)x\rangle = \tfrac{1}{2}\alpha(A + A^*).$$

For a proof of $\alpha(A + A^*) = \sup_{\|x\|=1} \langle x, (A + A^*)x\rangle$ (the Rayleigh principle) in Hilbert spaces, see [151, Theorem XI.8.2]. □

The dissipativity of A then only depends on properties of the self-adjoint linear operator $A + A^*$, see Definition 1.24.

Corollary 2.37. *Let $A \in \mathfrak{L}(X)$ be a linear operator on a Hilbert space X. Then the initial growth rate μ of A with respect to the norm $\|\cdot\|_X$ satisfies*

$$\begin{aligned} \mu(A) \le 0 &\iff -(A + A^*) \text{ is positive,} \\ \mu(A) < 0 &\iff -(A + A^*) \text{ is coercive.} \end{aligned}$$

For strongly continuous semigroups the initial growth rate may be $+\infty$ as the following example shows.

Example 2.38. Consider the Hilbert space $X = \bigoplus_{k\in\mathbb{N}} \mathbb{R}^2$ which is the direct sum of copies of $\mathbb{R}^2$. Let us study the unbounded linear block-diagonal operator

$$A = \bigoplus_{k\in\mathbb{N}^*} \begin{pmatrix} -k & 4k+2 \\ 0 & -k-1 \end{pmatrix} : \quad X \to X,$$

which is built up from stable 2×2 matrices A_k. Each of these matrices satisfies a growth bound $1.5 \approx \frac{4}{e} \le \sup_{t\ge 0} \left\|e^{A_k t}\right\| \le \frac{\sqrt{37}}{3} \approx 2$, see Theorem 4.8, but the maximum value of

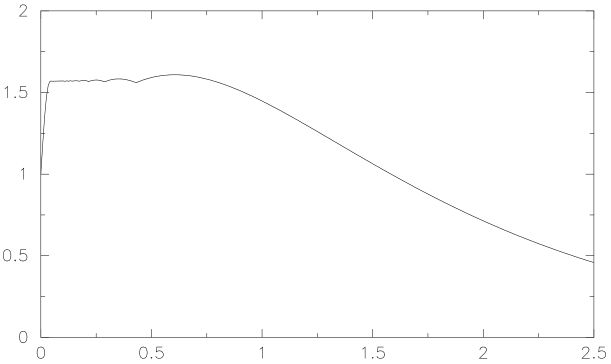

Figure 2.4: Norm of the matrix exponential for $A = \bigoplus_{k=1}^{20} A_k$.

$t \mapsto \left\|e^{A_k t}\right\|$ is attained at t_k^* for which $t_k^* \to 0$ holds as $k \to \infty$, which can be verified numerically. The semigroup $(T(t))_{t\in\mathbb{R}_+} = (e^{At})_{t\in\mathbb{R}^+}$ generated by A will also satisfy $1.5 \leq \sup_{t>0} \|T(t)\| \leq 2$, and therefore $\lim_{t\to 0} \|T(t)\| > 1$, hence T is not a uniformly continuous semigroup, and $\mu(A) = \alpha_0(T) = \infty$, see Corollary 2.17 and Theorem 2.19. Figure 2.4 shows the spectral norm $\left\|e^{At}\right\|$ for $A = \bigoplus_{k=1}^{20} A_k$. ■

Remark 2.39. Let A be the generator of a strongly continuous semigroup T. The initial growth rate $\mu(A)$ collects the microscopic effects of $t \mapsto \|T(t)\|$ for $t > 0$ near zero, while $\alpha(A)$ models the macroscopic or asymptotic effects. Hence if $\mu(A)$ differs significantly from $\alpha(A)$, say $\mu(A) > 0$ while $\alpha(A) < 0$, then we expect non-trivial *transient effects* in $t \mapsto \|T(t)\|$ for moderately sized $t > 0$ like multiple local maxima and minima. Note that μ depends on the used norm, while α is independent of the norm.

2.3.1 Initial Growth Rates in Finite-Dimensional Spaces

We will now turn to the matrix case and study the initial growth rate associated with a vector norm $\|\cdot\|$ of interest for matrices in $\mathbb{K}^{n\times n}$. In finite dimensions we identify $y \in \mathbb{K}^n$ with the linear form $f_y \in (\mathbb{K}^n)^* : x \mapsto y^*x$.[2] Hence the evaluation of a linear form $\langle f_y, x\rangle$ with $f_y \in (\mathbb{K}^n)^*$, $x \in \mathbb{K}^n$ is identified with the inner product $\langle x, y\rangle_2 = y^*x$ for $x, y \in \mathbb{K}^n$. Let us collect some of the properties of the initial growth rate.

Proposition 2.40. *Given matrices A, A' on $\mathbb{K}^{n\times n}$ and scalars $z \in \mathbb{K}, \alpha \in \mathbb{R}$. The initial growth rate $\mu(\cdot)$ satisfies*

(i) $-\mu(-A) \leq \operatorname{Re}\lambda \leq \mu(A)$, $\lambda \in \sigma(A)$,

(ii) $\mu(\alpha A) = |\alpha|\,\mu((\operatorname{sgn}\alpha)A)$,

(iii) $|\mu(A)| \leq \|A\|$,

(iv) $\mu(A + zI) = \mu(A) + \operatorname{Re} z$,

(v) $\mu(A + A') \leq \mu(A) + \mu(A')$,

(vi) $\mu(A) = \lim_{t\to\infty}(\|It + A\| - t)$.

[2] Note that $y \mapsto y^*$ is not $\mathbb{C}$-linear.

Proof. From Proposition 2.31 we know that

$$\mu(A) = \sup\{\operatorname{Re}\langle Ax, y\rangle_2 \mid \|x\|\,\|y\|^* = \langle x, y\rangle_2 = 1\}.$$

Moreover, $\mu(-A)$ satisfies $-\mu(-A) = \inf\{\operatorname{Re}\langle Ax, y\rangle_2 \mid \|x\|\,\|y\|^* = \langle x, y\rangle_2 = 1\}$. Hence by enlarging or restricting the conditions on the pair (x, y) we obtain the required statements. For item *(i)*, consider an eigenvector x corresponding to an eigenvalue $\lambda \in \sigma(A)$. Then for all dual vectors y of x, $\operatorname{Re}\langle Ax, y\rangle_2 = \operatorname{Re}\lambda\langle x, y\rangle_2 = \operatorname{Re}\lambda\,\|y\|^*\,\|x\|$, which shows *(i)*. Items *(ii)* and *(iv)* hold as $\mu(\alpha A) = \mu(\operatorname{sgn}\alpha\,|\alpha|\,A) = |\alpha|\,\mu((\operatorname{sgn}\alpha)A)$ and

$$\mu(A + zI) = \sup_{(x,y)\text{ NDP}} \operatorname{Re}\langle (A + zI)x, y\rangle_2 = \sup_{(x,y)\text{ NDP}} \operatorname{Re}\left(\langle Ax, y\rangle_2 + \langle zx, y\rangle_2\right) = \mu(A) + \operatorname{Re} z.$$

Formula *(vi)* is found in Lemma 2.30. For *(iii)* let us replace the unitary dual pair (x, y) by the normed pair (x, y) where $\|x\| = 1 = \|y\|^*$. Then

$$\mu(A) \le \sup_{\|x\|=1=\|y\|^*} \operatorname{Re}\langle Ax, y\rangle_2 \le \sup_{\|x\|=1=\|y\|^*} \|y\|^*\,\|A\|\,\|x\| = \|A\|\,.$$

Now $t - \|A\| \le \|It + A\|$ for all $t \ge 0$ and hence by *(vi)* we have $\mu(A) \ge -\|A\|$ which shows the lower bound in *(iii)*.
The subadditivity *(v)* is again verified using Proposition 2.31,

$$\begin{aligned}\mu(A + A') &= \sup_{(x,y)\text{ NDP}} \operatorname{Re}\langle (A + A')x, y\rangle_2\\ &\le \sup_{(x,y)\text{ NDP}} \operatorname{Re}\langle Ax, y\rangle_2 + \sup_{(x,y)\text{ NDP}} \operatorname{Re}\langle A'x, y\rangle_2 = \mu(A) + \mu(A').\end{aligned}$$

Hence all statements of Proposition 2.40 have been verified. □

Items *(ii)* and *(v)* of Proposition 2.40 imply that μ is a convex function with

$$\mu(\alpha A + (1-\alpha)A') \le \alpha\mu(A) + (1-\alpha)\mu(A') \quad \text{for all } A, A' \in \mathbb{K}^{n\times n} \text{ and } \alpha \in [0,1].$$

Items *(iii)* and *(v)* show that μ a continuous function, as $\mu(A + \Delta) \le \mu(A) + \|\Delta\|$ and $\mu(A) \le \mu(A + \Delta) + \mu(-\Delta) \le \mu(A + \Delta) + \|\Delta\|$.
A matrix $A \in \mathbb{C}^{n\times n}$ generates a contraction semigroup with respect to $\|\cdot\|$ if the closed unit ball $\mathbb{B} = \{x \in \mathbb{C}^n \mid \|x\| \le 1\}$ is forward-invariant under the flow of $\dot{x} = Ax$. Hence for every $t > 0$ the inclusion $e^{At}\mathbb{B} \subset \mathbb{B}$ holds. Note that this only needs to be checked for an infinitesimally small $t > 0$, i.e., we need a criterion which decides if for every initial value $x_0 \in \partial\mathbb{B}$ the derivative of the solution $x(t, x_0)$ of $\dot{x} = Ax$ in $t = 0$, $\dot{x}(0, x_0)$, points inside or is tangentially to the unit ball $\mathbb{B}$. And indeed this information is provided by the initial growth rate $\mu(A)$ as $\mu(x, A) \le 0$ for all $x \in \partial\mathbb{B}$ is equivalent to $\lim_{h\searrow 0} \frac{1}{h}\|x + hAx\| - \|x\| \le 0$ which shows that $\|x + hAx\| \le \|x\| = 1$ for $h \to \infty$, hence Ax points inside or along the unit ball. Thus $\mathbb{B}$ is forward-invariant under the flow of $\dot{x} = Ax$ if $\mu(A) \le 0$, where μ satisfies

$$\mu(A) = \lim_{t\searrow 0} t^{-1}\left(\|I + tA\| - 1\right). \tag{2.25}$$

We have seen in Proposition 2.31 and Proposition 2.40 that this limit is well-defined since $t^{-1}(\|I+tA\|-1)$ is monotonically decreasing for $t \searrow 0$ and since it is bounded from below by $-\|A\|$.

The following theorem recalls the formulas for some standard operator norms and gives the corresponding initial growth rates.

Theorem 2.41. *Let $x=(x_i)\in\mathbb{C}^n$ and $A=(a_{ij})\in\mathbb{C}^{n\times n}$. If $\mu_p(\cdot)$ denotes the initial growth rate with respect to the norm $\|\cdot\|_p$ $(p=1,2,\infty)$ then*

$$\begin{aligned}
\|x\|_1 &= \sum_i |x_i|, & \|A\|_1 &= \max_j \sum_i |a_{ij}|, & \mu_1(A) &= \max_j \left(\operatorname{Re} a_{jj} + \sum_{i\neq j} |a_{ij}|\right),\\
\|x\|_2 &= \sqrt{\sum_i |x_i|^2}, & \|A\|_2 &= \sqrt{\max_i \lambda_i(A^*A)}, & \mu_2(A) &= \tfrac{1}{2}\max_i \lambda_i(A+A^*),\\
\|x\|_\infty &= \max_i |x_i|, & \|A\|_\infty &= \max_i \sum_j |a_{ij}|, & \mu_\infty(A) &= \max_i \left(\operatorname{Re} a_{ii} + \sum_{j\neq i} |a_{ij}|\right).
\end{aligned}$$

The operator norms $\|\cdot\|_1$, $\|\cdot\|_\infty$ and $\|\cdot\|_2$ are called (absolute) column-sum norm, (absolute) row-sum norm and spectral norm, respectively.

Proof. We will show the formulas for the initial growth rates. Note that the spectral norm is self-dual, therefore

$$\mu_2(A) = \sup_{\|x\|=1} \operatorname{Re}\langle x, Ax\rangle_2 = \tfrac{1}{2}\sup_{\|x\|=1} x^*(A+A^*)x = \tfrac{1}{2}\lambda_{\max}(A+A^*),$$

where the last equality follows from the Rayleigh-Ritz Theorem for Hermitian matrices, see [70]. For the 1- and ∞-norm case we use the fact that the real part of $z\in\mathbb{C}$ can be represented by $\operatorname{Re} z = \lim_{r\to\infty}|z+r|-r$. By setting $r=t^{-1}$ in (2.25) we obtain

$$\begin{aligned}
\mu_1(A) &= \lim_{t\to 0} t^{-1}\|I+At\| - t^{-1} = \lim_{r\to\infty}\|rI+A\|_1 - r\\
&= \max_j \left(\lim_{r\to\infty}|a_{jj}+r| - r + \sum_{i\neq j}|a_{ij}|\right) = \max_j\left(\operatorname{Re} a_{jj} + \sum_{i\neq j}|a_{ij}|\right).
\end{aligned}$$

Analogously, $\mu_\infty = \max_i(\operatorname{Re} a_{ii} + \sum_{j\neq i}|a_{ij}|)$. □

The following proposition is a direct consequence of the characterization of dissipativity in Lemma 2.32 and the rule $\mu(A-\beta I)=\mu(A)-\beta$ of Proposition 2.40.

Proposition 2.42. *Suppose $\|\cdot\|$ is an operator norm on $\mathbb{K}^{n\times n}$. Then $\mu(A)$ is the least upper exponential bound for $\|e^{At}\|$, $\mu(A)=\min\left\{\mu\in\mathbb{R}\,\middle|\,\forall t\geq 0:\ \|e^{At}\|\leq e^{\mu t}\right\}$.*

This characterization also holds for uniformly continuous semigroups on a Banach space X, as $\|e^{At}\|\leq e^{\|A\|t}$ for $t\in\mathbb{R}_+$, see also (2.9) and Theorem 2.19.

If the matrix norm under consideration is not an operator norm, dissipativity and $\mu(A)\leq 0$ are not equivalent as the following example shows.

Example 2.43. Suppose that $\mathbb{C}^{n\times n}$ is endowed with the Frobenius norm $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$ which is a matrix norm, but not an operator norm. Furthermore, if A is a matrix of rank 1 then $\|A\|_F = \|A\|_2$. Hence if $A \in \mathbb{C}^{n\times n}$ has a simple uniquely determined rightmost eigenvalue then $\left\|e^{At}\right\|_F \approx \left\|e^{At}\right\|_2$ for t large since the dominant eigenmotion "survives" asymptotically, see Proposition 3.14. For $A = \left(\begin{smallmatrix} -5 & 36 \\ 0 & -20 \end{smallmatrix}\right)$ the transient behaviour $t \mapsto \left\|e^{At}\right\|$ is depicted in Figure 2.5 for both the Frobenius and the spectral norm. The initial growth rate of A with respect to $\|\cdot\|_F$ (which is negative here) is not an upper exponential bound for $\left\|e^{At}\right\|_F$. However, for large t, $\left\|e^{At}\right\|_F$ is a good approximation of $\left\|e^{At}\right\|_2$. ■

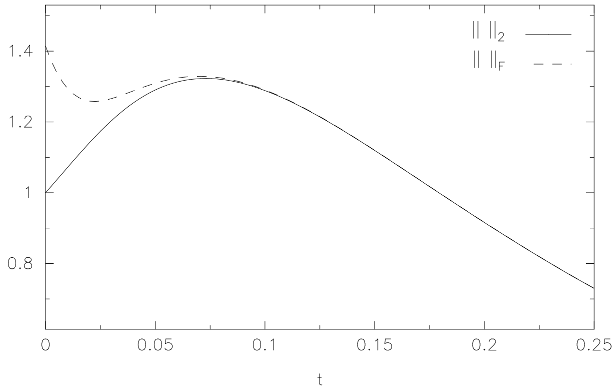

Figure 2.5: Transient motion and the Frobenius norm.

If the initial growth rate is negative then its absolute value can be interpreted as a *dissipativity radius*, as the following result implies.

Lemma 2.44. *Let $A \in \mathbb{K}^{n\times n}$ and $\|\cdot\|$ be an operator norm on $\mathbb{K}^{n\times n}$. Suppose that A is dissipative with $\mu(A) < 0$. If $\Delta \in \mathbb{K}^{n\times n}$, $\|\Delta\| \le \delta$ then*

$$\left\|e^{(A+\Delta)t}\right\| \le e^{(\mu(A)+\delta)t}, \qquad t \ge 0.$$

Hence $A + \Delta$ is dissipative, if $\delta \le |\mu(A)|$. On the other hand, if $\delta > |\mu(A)|$ then $A + \Delta$ with $\Delta = \delta I$ is not dissipative.

Proof. The subadditivity of the initial growth rate and Proposition 2.42 give the estimate

$$\left\|e^{(A+\Delta)t}\right\| \le e^{\mu(A+\Delta)t} \le e^{(\mu(A)+\mu(\Delta))t} \le e^{(\mu(A)+\delta)t}.$$

By Proposition 2.40 (iv) we have $\mu(A + \delta I) = \mu(A) + \delta I$ and for $\delta > |\mu(A)|$ and $\Delta = \delta I$, the initial growth rate $\mu(A + \Delta) > 0$, hence $A + \Delta$ is not dissipative. □

Hence if A is dissipative, matrix perturbations $\Delta \in \mathbb{K}^{n\times n}$ with norm $\|\Delta\| \le |\mu(A)|$ will not destroy dissipativity, $\mu(A + \Delta) \le 0$. On the other hand, for $\delta > |\mu(A)|$ the perturbation $\Delta = \delta I$ satisfies $\mu(A + \Delta) > 0$.

Let us return to the discussion of duality issues related to dissipativity. In the following $\langle\cdot,\cdot\rangle_2$ denotes the standard Euclidean inner product in $\mathbb{K}^n$ while $\|\cdot\|$ is an arbitrary vector norm on $\mathbb{K}^n$. Let us denote the unit sphere by $\mathbb{S} = \{x \in \mathbb{K}^n \mid \|x\| = 1\}$.

Figure 2.6 illustrates the unit balls of dual norms and a pair of dual vectors $(x, y) \in \mathbb{S} \times \mathbb{S}^*$. A small calculation shows that the line connecting x with ${}^{y}/_{\|y\|_2^2}$ is tangent to the unit ball $\mathbb{B}$ in x, as y is an outer normal of $\mathbb{B}$ in x.

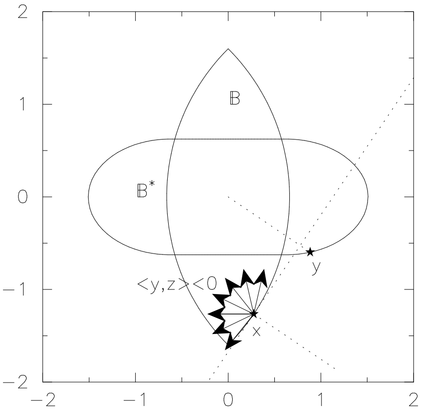

Figure 2.6: Dissipativity.

Now if A is strictly dissipative with

$$\max_{(x,y)\ \mathrm{NDP}} \mathrm{Re}\, \langle y, Ax \rangle_2 < 0$$

then all $z = Ax = \dot{x}$ point inside the unit ball $\mathbb{B}$. In other words, if (x, y) is a unitary dual pair with respect to $\|\cdot\|$, the angle spanned by $z = Ax$ and the outer normal y is obtuse. Figure 2.6 depicts allowed directions for a given x. If A is a dissipative matrix with $\max_{(x,y)\ \mathrm{NDP}} \mathrm{Re}\, \langle y, Ax \rangle_2 \leq 0$ then there may exist $x \in \partial\mathbb{B}$ such that Ax spans a right angle with the outer normal y, hence Ax is tangentially to the unit ball $\mathbb{B}$.

As a different interpretation of Proposition 2.31, the initial growth rate for a general norm corresponds to a rightmost point of the *numerical range* of A (also called *field of values*). The numerical range of A with respect to the norm $\|\cdot\|$ is defined as follows,

$$W_{\|\cdot\|}(A) = \left\{ \frac{\langle Ax, y \rangle_2}{\langle x, y \rangle_2} \,\middle|\, (x, y) \in \mathbb{C}^n \times \mathbb{C}^n \ \text{ is a dual pair of } \|\cdot\| \right\} \subset \mathbb{C}. \tag{2.26}$$

$\Delta W_{\|\cdot\|}(A)$ is the set of all *Rayleigh quotients* of dual pairs, its *numerical abscissa* is defined by the initial growth rate with respect to the norm $\|\cdot\|$,

$$n_{\|\cdot\|}(A) = \sup \left\{ \mathrm{Re}\, w \,\middle|\, w \in W_{\|\cdot\|}(A) \right\}$$

and equals the initial growth rate with respect to the norm $\|\cdot\|$, $n_{\|\cdot\|} = \mu_{\|\cdot\|}$. There exists a well-studied object which is closely related to the numerical range associated with $\|\cdot\|_\infty$, $W_\infty(\cdot)$. This is the object of the following theorems and propositions.

Theorem 2.45 (Gershgorin's Theorem)**.** *For $A \in \mathbb{K}^{n \times n}$ set $R_i = \sum_{j \neq i} |a_{ij}|$, $i = 1, \ldots, n$, and define the ith Gershgorin disk by $\mathcal{G}_i(A) = \{ z \in \mathbb{C} \mid |z - a_{ii}| \leq R_i \}$. Then*

$$\sigma(A) \subset \mathcal{G}(A) := \bigcup_{i=1}^{n} \mathcal{G}_i(A).$$

Each connected component of $\mathcal{G}(A)$ contains at least one eigenvalue of A.

For a proof, see [70]. We will now have a closer look at the Gershgorin set $\mathcal{G}(A)$.

Proposition 2.46. *For a given matrix $A \in \mathbb{C}^{n \times n}$ the Gershgorin set $\mathcal{G}(A)$ is contained in the numerical range $W_\infty(A)$ of A associated with $\|\cdot\|_\infty$, and the numerical range is contained in the convex hull of the Gershgorin set, that is,*

$$\mathcal{G}(A) \subset W_\infty(A) \subset \mathrm{conv}\, \mathcal{G}(A). \tag{2.27}$$

Proof. Let us first describe dual vectors of $x \in \mathbb{C}^n$ with respect to $\|\cdot\|_\infty$. Note that its dual norm is $\|\cdot\|_\infty^* = \|\cdot\|_1$. We introduce the index set $I(x) := \{i \in \{1, \ldots, n\} \mid \|x\|_\infty = |x_i|\}$ which collects all critical indices of x. Then the set of all dual vectors of x with respect to $\|\cdot\|_\infty$ is given by

$$\mathcal{D}(x) := \left\{ \sum_{i \in I(x)} \alpha_i x_i e^i \,\middle|\, \sum_{i \in I(x)} \alpha_i > 0, \alpha_i \geq 0 \text{ for all } i \in I(x) \right\},$$

where e^i is the i-th unit vector in $\mathbb{C}^n$. This can be seen as follows. For every $y \in \mathcal{D}(x)$, its dual norm is given by $\|y\|_\infty^* = \|y\|_1 = \sum_{i \in I(x)} \alpha_i |x_i| = \|x\|_\infty \sum_{i \in I(x)} \alpha_i > 0$ and $\langle y, x \rangle_2 = \sum_{i \in I(x)} \alpha_i \bar{x}_i x_i = \|x\|_\infty^2 \sum_{i \in I(x)} \alpha_i = \|x\|_\infty \|y\|_1$. Hence the pair (x, y) is a dual pair by Definition 2.26. Conversely, assume that y is a dual vector of x. Then $\langle x, y \rangle_2 = \sum_{i=1}^n \bar{y}_i x_i = \|x\|_\infty \|y\|_1 = \max_i |x_i| \sum_{i=1}^n |y_i|$ has to hold. But generally, we only have for $x, y \in \mathbb{C}^n$ with $\langle x, y \rangle_2 \in \mathbb{R}_+$ that

$$\sum_{i=1}^n \bar{y}_i x_i = \left| \sum_{i=1}^n \bar{y}_i x_i \right| \leq \sum_{i=1}^n |y_i| \, |x_i| \leq \max_i |x_i| \sum_{i=1}^n |y_i| . \tag{2.28}$$

To obtain equality in (2.28), we must have $y_i = 0$ for all indices i with $\|x\|_\infty \neq |x_i|$. Collecting all those indices i with $|x_i| = \|x\|_\infty$ in the set $I(x)$, we rewrite (2.28) as $\sum_{i \in I(x)} \bar{y}_i x_i \leq \sum_{i \in I(x)} |y_i| \, |x_i|$. Again, to obtain equality, y_i must be a nonnegative multiple of x_i. As y is a dual vector of x, at least one $y_i \neq 0$ which shows that $y \in \mathcal{D}(x)$. In the case that $I(x) = \{i\}$ the dual vectors y of x satisfy $y = \alpha x_i e^i$ for $\alpha > 0$. We now show that every $z \in \mathcal{G}(A)$ can be represented by a Rayleigh quotient $z = \langle Ax, y \rangle_2 / \langle x, y \rangle_2$ where (x, y) is a dual pair with respect to $\|\cdot\|_\infty$. For $z \in \mathcal{G}(A)$ there exist an index i_0 and $\zeta \in \mathbb{C}$, $|\zeta| \leq 1$ such that $z = a_{i_0 i_0} + (\sum_{j \neq i_0} |a_{i_0 j}|)\zeta$. By introducing $\zeta_j \in \mathbb{C}$ such that $a_{i_0 j} = |a_{i_0 j}| \bar{\zeta}_j$, $|\zeta_j| = 1$, we have

$$z = a_{i_0 i_0} + \sum_{j \neq i_0} a_{i_0 j} (\zeta \zeta_j), \qquad |\zeta \zeta_j| \leq 1 \quad \text{for all} \quad j = 1, \ldots, n.$$

The pair (x, e^{i_0}) of vectors $x = (\zeta\zeta_1, \ldots, 1, \ldots, \zeta\zeta_n)^\top$ with a 1-entry in the i_0-th component is a normed dual pair associated with $\|\cdot\|_\infty$ because $\|x\|_\infty = 1$ as $|\zeta\zeta_j| \leq 1$, and because $\|e^{i_0}\|_1 = 1$, $\langle e^{i_0}, x \rangle_2 = 1$ by construction. Hence $z = \langle e^{i_0}, Ax \rangle_2 / \langle e^{i_0}, x \rangle_2 \in W_\infty(A)$ and therefore $\mathcal{G}(A) \subset W_\infty(A)$.

On the other hand, for every $z \in W_\infty(A)$ there exists a dual pair (x, y) such that $z = \frac{\langle Ax, y \rangle_2}{\langle x, y \rangle_2}$. Since $y \in \mathcal{D}(x)$, it is given by $y = \sum_{i \in I(x)} \alpha_i x_i e^i$ for the index set $I(x)$ defined above and $\alpha_i \geq 0$, $\sum_{i \in I(x)} \alpha_i > 0$. For each $i_0 \in I(x)$ the vector $y^{i_0} = x_{i_0} e^{i_0}$ is a dual vector of x. Since $|\frac{x_j}{x_{i_0}}| \leq 1$ for all $j \in \{1, \ldots, n\}$ the associated Rayleigh quotient satisfies

$$\frac{\langle Ax, y^{i_0} \rangle_2}{\langle x, y^{i_0} \rangle_2} = \frac{\bar{x}_{i_0}}{\bar{x}_{i_0} x_{i_0}} \left(\sum_{j=1}^n a_{i_0 j} x_j \right) = a_{i_0 i_0} + \sum_{j \neq i_0} a_{i_0 j} \frac{x_j}{x_{i_0}} \in \mathcal{G}(A). \tag{2.29}$$

Hence we obtain for $z = \frac{\langle Ax,y\rangle_2}{\langle x,y\rangle_2} \in W_\infty(A)$,

$$z = \frac{\langle Ax,y\rangle_2}{\langle x,y\rangle_2} = \frac{\sum_{i\in I(x)} \langle Ax, \alpha_i y^i\rangle_2}{\sum_{j\in I(x)} \langle x, \alpha_j y^j\rangle_2} = \sum_{i\in I(x)} \alpha_i \frac{\langle x, y^i\rangle_2}{\sum_{j\in I(x)} \alpha_j \langle x, y^j\rangle_2} \frac{\langle Ax, y^i\rangle_2}{\langle x, y^i\rangle_2},$$

where $\frac{\langle Ax,y^i\rangle_2}{\langle x,y^i\rangle_2} \in \mathcal{G}(A)$ by (2.29). Therefore, each $z \in W_\infty(A)$ is given by a convex combination of elements in $\mathcal{G}(A)$ and thus $W_\infty(A) \subset \operatorname{conv}\mathcal{G}(A)$. □

The numerical range of $A \in \mathbb{C}^{n\times n}$ associated with any norm always contains the spectrum of A, hence the first statement of Theorem 2.45 follows immediately when x in the Rayleigh quotient is set to an eigenvector of A.

Remark 2.47. We conclude from (2.29) that we obtain the Gershgorin set if we consider Rayleigh quotients of dual pairs (x, y) where the dual vector y of x is a scalar multiple of a unit vector, that is,

$$\mathcal{G}(A) = \left\{ \frac{\langle Ax,y\rangle_2}{\langle x,y\rangle_2} \,\middle|\, (x,y)\,\text{DP of } \|\cdot\|_\infty, y = x_{i_0} e^{i_0} \text{ for some } i_0 \right\}. \tag{2.30}$$

This equation allows the following interpretation. Let us consider the unit sphere $\mathbb{S}_\infty = \{x \in \mathbb{C}^n \mid \|x\|_\infty = 1\} \subset \mathbb{C}^n$ as a CW-complex, see [75]. If we delete all its components of dimensions less than $n-1$ we get a set of open faces. These faces consist of points with uniquely determined dual vectors. Using only those points for the Rayleigh quotients, we arrive at (2.30). Hence each Gershgorin disk $\mathcal{G}_i(A)$ corresponds to those Rayleigh quotients which correspond to dual pairs with $y = e^i$ as dual vector.

From Proposition 2.31 and Proposition 2.46 we get the following characterization of the initial growth rate with respect to the ∞-norm.

Corollary 2.48. *For all $A \in \mathbb{C}^{n\times n}$,*

$$\mu_\infty(A) = \max\{\operatorname{Re} z \mid z \in W_\infty(A)\} = \max\{\operatorname{Re} z \mid z \in \mathcal{G}(A)\}. \tag{2.31}$$

Proof. We have $\sup\{\operatorname{Re} z \mid z \in \mathcal{G}(A)\} = \sup\{\operatorname{Re} z \mid z \in \operatorname{conv}\mathcal{G}(A)\}$, and hence (2.31) follows from (2.27). □

Definition 2.49. A matrix $A = (A_{ij}) \in \mathbb{C}^{n\times n}$ is called *diagonally dominant* (with negative real parts of the diagonal elements) if

$$\text{for all} \quad i = 1, \ldots, n: \qquad \operatorname{Re} a_{ii} + \sum_{j\neq i} |a_{ij}| \le 0, \tag{2.32}$$

it is called *strictly diagonally dominant* if the strict inequality holds in (2.32).

An application of Corollary 2.48 gives the following result.

Corollary 2.50. *A (strictly) diagonally dominant matrix $A \in \mathbb{C}^{n\times n}$ is (strictly) dissipative with respect to $\|\cdot\|_\infty$, $\mu_\infty(A) \le 0$ ($\mu_\infty(A) < 0$, respectively). Moreover, its Gershgorin disks are located in the open (closed) left half-plane, $\mathcal{G}(A) \subset \bar{\mathbb{C}}_-$ ($\mathcal{G}(A) \subset \mathbb{C}_-$).*

Proof. From $z \in \mathcal{G}(\mathcal{A})$ it follows that $\operatorname{Re} z \le \max_i \left(\operatorname{Re} a_{ii} + \sum_{j\neq i} |a_{ij}|\right) = \mu_\infty(A)$. Hence if A is (strictly) diagonally dominant, then $\mu_\infty(A) \le 0$ ($\mu_\infty(A) < 0$), such that $\mathcal{G}(\mathcal{A}) \subset \bar{\mathbb{C}}_-$ ($\mathcal{G}(\mathcal{A}) \subset \mathbb{C}_-$, respectively). □

Example 2.51. Let us consider the matrix $A = \begin{pmatrix} -4 & 1 & 1 \\ 2 & -1 & 0 \\ 1 & 0 & -7 \end{pmatrix}$. Its Gershgorin set contains two disks of radius 2 centered at -4 and -1 and a disk of radius 1 centered at -7, while the spectrum is given by $\sigma(A) = \{-0.4171, -4.251, -7.332\}$. Figure 2.7 shows the spectrum, the Gershgorin set, and the numerical range with respect to the ∞-norm shaded in gray. Note that, unlike the Euclidean numerical range W_2, the set W_∞ is not necessarily a convex set. ■

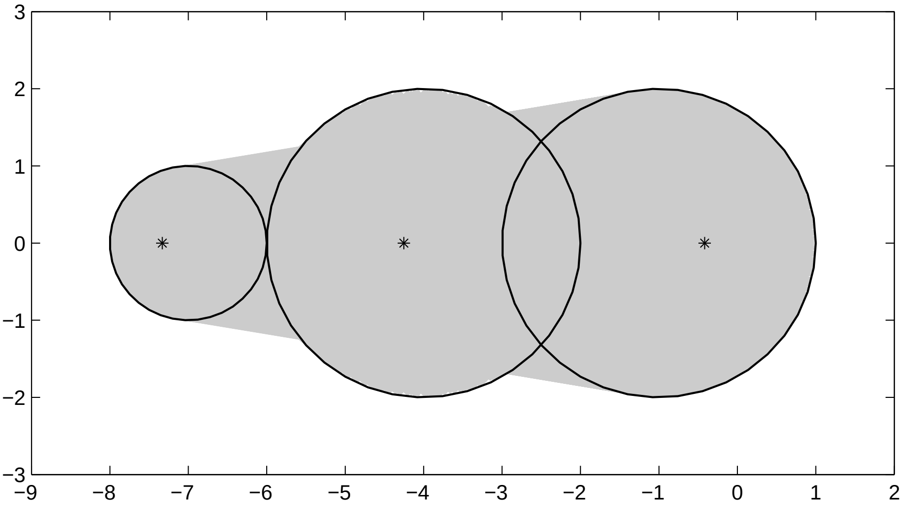

Figure 2.7: Numerical range $W_\infty(A)$ and Gershgorin disks $\mathcal{G}(A)$.

For later references, we collect the characterizations of the initial growth rate in the following corollary.

Corollary 2.52. *The initial growth rate $\mu(A)$ of $A \in \mathbb{K}^{n\times n}$ associated with the vector norm $\|\cdot\|$ is characterized as follows*

$$\mu(A) = \tfrac{d}{dt^+} \left\|e^{At}\right\| \big|_{t=0} = \lim_{h\searrow 0} \tfrac{1}{h}(\left\|e^{Ah}\right\| - 1) = \lim_{h\searrow 0} \tfrac{1}{h} \log \left\|e^{At}\right\| \tag{2.33}$$

$$= \lim_{h\searrow 0} \tfrac{1}{h} \left(\|I + Ah\| - 1\right) = \lim_{r\to\infty} \left(\|rI + A\| - r\right) \tag{2.33a}$$

$$= \lim_{h\searrow 0} \tfrac{1}{h} \left(\left\|(I + \tfrac{h}{k}A)^k\right\| - 1\right) = \lim_{h\searrow 0} \tfrac{1}{h} \left(\left\|(I - \tfrac{h}{k}A)^{-k}\right\| - 1\right), \quad k = 1, 2, 3, \ldots, \tag{2.33b}$$

$$\mu(A) = \max_{(x,y)\,DP} \frac{\operatorname{Re}\langle Ax, y\rangle_2}{\langle x, y\rangle_2} = \max_{(x,y)\,DP} \operatorname{Re}\frac{y^* Ax}{y^* x}, \tag{2.33c}$$

$$\mu(A) = \inf\left\{\omega \in \mathbb{R} \,\middle|\, \forall t \geq 0 \ \left\|e^{At}\right\| \leq e^{\omega t}\right\} \tag{2.33d}$$

$$= \inf\left\{\omega \in \mathbb{R} \,\middle|\, \forall \alpha > \omega\, \forall z \in \mathbb{C} : \ \left\|(\alpha I - A)^{-1} z\right\| \leq \frac{1}{\alpha - \omega}\left\|z\right\|\right\}. \tag{2.33e}$$

Proof. The characterizations (2.33) and (2.33a) are given in Theorem 2.19 and Definition 2.29. Equation (2.33b) follows from (2.33) by replacing e^{At} with the product formulation $(I - \frac{t}{k}A)^{-k}$ from Theorem 2.7 and with $(I + \frac{t}{k}A)^k$. We can identify the terms of (2.33b) with the initial growth rate as for all $k \in \mathbb{N}$, $k \geq 1$, $(I - \frac{t}{n}A)^{-n} = I + At + O(t^2)$ and $(I + \frac{t}{n}A)^n = I + At + O(t^2)$. Equation (2.33$c$) is due to Proposition 2.31 while (2.33d) is derived in Proposition 2.42. The last characterization (2.33e) is an application of Theorem 2.34 to $A - \omega I$, namely, $A - \omega I$ is dissipative if $\|(\tilde{\lambda} + \omega)I - A)x\| \geq \tilde{\lambda}\,\|x\|$ for all $\tilde{\lambda} > 0$. Setting $\alpha = \tilde{\lambda} + \omega$ and $z = (\alpha I - A)^{-1}x$ gives (2.33e). □

2.4 Liapunov Norms

Liapunov theory plays an important role in many fields of applied mathematics. Here the initial growth rate serves as an indicator if the semigroup $T = (e^{At})_{t \in \mathbb{R}_+}$ forms a contraction with respect to the norm under consideration. The same fact can also be interpreted in the following way: The norm is a Liapunov function for the system $\dot{x} = Ax$. We will prove this and related facts in the current section.

Definition 2.53. Let $A \in \mathbb{K}^{n \times n}$. If $\|\cdot\|$ is a vector norm on $\mathbb{K}^n$ such that the associated initial growth rate satisfies $\mu_{\|\cdot\|}(A) \leq 0$ then $\|\cdot\|$ is called a *Liapunov norm* for A. It is called a *strict Liapunov norm* if $\mu_{\|\cdot\|}(A) < 0$.

If $\|\cdot\|$ is a Liapunov norm for A then it generates a contraction semigroup, hence A is marginally stable, and if the norm is a strict Liapunov norm, then A is exponentially stable and generates a uniform contraction semigroup which follows from Proposition 2.42. Let us recall the definition of a Liapunov function.

Definition 2.54. A *Liapunov function* for the linear system $\dot{x} = Ax$ is a continuous function $V : \mathbb{K}^n \to \mathbb{R}$ for which the following properties hold:

1. V is *proper* at 0, i.e., the set $\{x \in \mathbb{K}^n \,|\, V(x) \leq \varepsilon\}$ is compact for all $\varepsilon > 0$.
2. V is positive definite, $V(0) = 0$ and $V(x) > 0$ for all $x \neq 0$.
3. For each initial value $x_0 \neq 0$ there exists a time $\tau > 0$ so that the solution $x(t, x_0)$ of $\dot{x} = Ax$ satisfies $V(x(t, x_0)) \leq V(x_0)$ for $t \in (0, \tau)$ and $V(x(\tau, x_0)) < V(x_0)$.

It is well-known that the existence of a Liapunov function for $\dot{x} = Ax$ implies that this system is asymptotically stable, see Sontag [128]. We now have a canonical candidate for a Liapunov function, namely, the Liapunov norm.

Lemma 2.55. *If $\|\cdot\|$ is a strict Liapunov norm for $A \in \mathbb{K}^{n\times n}$ then it is a Liapunov function for the system $\dot{x} = Ax$.*

Proof. Since $\|\cdot\|$ is a norm, it is clearly positive definite and has compact level sets. Now by the characterization of μ in (2.33*d*), we have $\left\|e^{At}x\right\| \leq e^{\mu(A)t}\|x\|$ for all $x \in \mathbb{K}^n$. As $\mu(A) < 0$, A generates a uniform contraction with respect to $\|\cdot\|$, or in other words, $\|\cdot\|$ is strictly decaying along the solutions of $\dot{x} = Ax$. Thus the norm is a Liapunov function for $\dot{x} = Ax$. □

In most cases, however, the norm of interest is not a Liapunov norm for the system under investigation. We therefore have to deal with two different norms, a given one and a suitable Liapunov norm. To compare these different norms on $\mathbb{C}^n$ we introduce the following notion.

Definition 2.56. Suppose ν and $\|\cdot\|$ are norms on $\mathbb{C}^n$. The *eccentricity of norms* of $\nu(\cdot)$ with respect to $\|\cdot\|$ is given by

$$\operatorname{ecc}(\nu) = \operatorname{ecc}(\nu, \|\cdot\|) = \frac{\max_{\|x\|=1}\nu(x)}{\min_{\|x\|=1}\nu(x)}. \tag{2.34}$$

The eccentricity measures the deformation of the unit balls of these two norms with respect to each other. Clearly,

$$\operatorname{ecc}(\nu, \|\cdot\|) = \frac{\max_{x\neq 0}\frac{\nu(x)}{\|x\|}}{\min_{x\neq 0}\frac{\nu(x)}{\|x\|}} = \frac{\max_{x\neq 0}\frac{\|x\|}{\nu(x)}}{\min_{x\neq 0}\frac{\|x\|}{\nu(x)}} = \operatorname{ecc}(\|\cdot\|, \nu). \tag{2.35}$$

This notion can now be employed to compare the transient behaviour under different norms. We obtain from Proposition 2.42 the following exponential bound.

Corollary 2.57. *Let $A \in \mathbb{C}^{n\times n}$ and given two norms $\|\cdot\|, \nu(\cdot)$ on $\mathbb{C}^n$. If $\mu_\nu(\cdot)$ denotes the initial growth rate with respect to $\nu(\cdot)$ we obtain*

$$\left\|e^{At}\right\| \leq \operatorname{ecc}(\nu, \|\cdot\|)e^{\mu_\nu(A)t},\ t \geq 0. \tag{2.36}$$

Proof. Proposition 2.42 gives the exponential estimate $\nu(e^{At}) \leq e^{\mu_\nu(A)t}$ for $t \geq 0$. For all $y \in \mathbb{C}^n$, $y \neq 0$ we obtain by considering the ν-norm of the normed vector $y/\|y\|$ that

$$\|y\| \min_{\|x\|=1}\nu(x) \leq \nu(y) \leq \|y\| \max_{\|x\|=1}\nu(x). \tag{2.37}$$

This implies for the associated operator norms $\|T\|, \nu(T)$ of any $T \in \mathbb{C}^{n\times n}$

$$\|T\| = \sup_{x\neq 0}\frac{\|Tx\|}{\|x\|} \leq \sup_{x\neq 0}\frac{\left(\min_{\|x\|=1}\nu(x)\right)^{-1}\nu(Tx)}{\left(\max_{\|x\|=1}\nu(x)\right)^{-1}\nu(x)} = \operatorname{ecc}(\nu, \|\cdot\|)\nu(T). \tag{2.38}$$

Setting $T = e^{At}$ gives the desired result. □

In particular, if ν is a strict Liapunov norm for A, then $\mu_\nu(A) < 0$, and (2.36) guarantees asymptotic stability. Let us study the following special setup for Corollary 2.57.

Proposition 2.58. *Given a vector norm $\|\cdot\|$ on $\mathbb{K}^n$ and an invertible matrix $W \in \mathrm{Gl}_n(\mathbb{K})$. Define $\nu(\cdot) = \|W\cdot\|$. The eccentricity of ν is given by the condition number of W,*

$$\kappa(W) := \mathrm{ecc}(\nu, \|\cdot\|) = \|W\| \left\|W^{-1}\right\|, \tag{2.39}$$

and the weighted initial growth rate satisfies

$$\mu_{\|\cdot\|,W}(A) := \mu_\nu(A) = \mu_{\|\cdot\|}(WAW^{-1}). \tag{2.40}$$

Proof. The eccentricity of ν is given by $\mathrm{ecc}(\nu) = \frac{\max_{\|x\|=1}\|Wx\|}{\min_{\|x\|=1}\|Wx\|}$. Now, $\max_{\|x\|=1}\|Wx\|$ is the operator norm of W and $\left(\min_{\|x\|=1}\|Wx\|\right)^{-1} = \max_{\|Wx\|=1}\|x\| = \max_{\|y\|=1}\|W^{-1}y\|$ is the operator norm of W^{-1} such that (2.39) holds. For the initial growth let us determine the operator norm associated with ν,

$$\nu(A) = \sup_{x\neq 0}\frac{\nu(Ax)}{\nu(x)} = \sup_{x\neq 0}\frac{\|WAx\|}{\|Wx\|} = \sup_{y\neq 0}\frac{\|WAW^{-1}y\|}{\|y\|} = \left\|WAW^{-1}\right\|,$$

where we used $y = W^{-1}x$. The characterization (2.33a) of the initial growth rate provides us with

$$\begin{aligned}\mu_\nu(A) &= \lim_{h\to 0} h^{-1}(\nu(I+Ah)-1) = \lim_{h\to 0} h^{-1}(\left\|W(I+Ah)W^{-1}\right\| - 1)\\ &= \lim_{h\to 0} h^{-1}(\left\|I + WAW^{-1}h\right\| - 1) = \mu_{\|\cdot\|}(WAW^{-1}),\end{aligned}$$

which shows (2.40). □

Surprisingly, there always exists a norm which realizes the best possible exponential bound. To see this, let us define the following constants, which measure transient motions.

Definition 2.59. Suppose $A \in \mathbb{K}^{n\times n}$ and $\|\cdot\|$ is a given operator norm on $\mathbb{K}^{n\times n}$. For any $\beta \geq \alpha(A)$ the *transient growth* or *transient amplification* of $(e^{At})_{t\geq 0}$ corresponding to the exponential rate β is defined by

$$M_\beta(A) = \inf\left\{M \in \mathbb{R} \,\middle|\, \forall t \geq 0 : \|e^{At}\| \leq Me^{\beta t}\right\}. \tag{2.41}$$

We set $M_\beta(A) = \infty$ if there is no M which satisfies the inequality in (2.41).

Now, $M_\beta(A) = M_0(A-\beta I)$ so that there is no loss of generality by only considering $\beta = 0$.

Definition 2.60. Given a norm $\|\cdot\|$ on $\mathbb{K}^n$ and a stable matrix $A \in \mathbb{K}^{n\times n}$.

1. A norm $\nu(\cdot)$ on $\mathbb{K}^n$ is called *transient norm* of A if $\mu_\nu(A) \leq 0$ and $\mathrm{ecc}(\nu, \|\cdot\|) = M_0(A) = \sup_{t\geq 0}\left\|e^{At}\right\|$.

2. The *Feller* norm on $\mathbb{K}^n$ induced by the matrix A is defined by $\|x\|_A = \sup_{t\geq 0}\left\|e^{At}x\right\|$.

This norm is named after W. Feller who used such a norm construction in his proof [41] of the Hille-Yosida Generation Theorem 2.6. Comparing with Definition 2.53, we see that each transient norm is also a Liapunov norm.

Lemma 2.61. *The Feller norm $\|\cdot\|_A$ induced by a stable matrix $A \in \mathbb{K}^{n\times n}$ is a transient norm of A.*

Proof. It is easily verified that $\|\cdot\|_A$ is indeed a norm, the triangle inequality holds because

$$\|x+y\|_A = \sup_{t\geq 0}\left\|e^{At}(x+y)\right\| \leq \sup_{t\geq 0}\left(\left\|e^{At}x\right\| + \left\|e^{At}y\right\|\right) \leq \|x\|_A + \|y\|_A \quad \text{for all} \quad x,y\in\mathbb{K}^n.$$

The eccentricity of $\|\cdot\|_A$ is given by $\mathrm{ecc}(\|\cdot\|_A) = \frac{\sup_{\|x\|=1}\sup_{t\geq 0}\left\|e^{At}x\right\|}{\inf_{\|x\|=1}\sup_{t\geq 0}\left\|e^{At}x\right\|}$. We now show that $\inf_{\|x\|=1}\sup_{t\geq 0}\left\|e^{At}x\right\| = 1$. If A is an exponentially stable matrix then for an arbitrary $x\in\mathbb{R}^n$, $\sup_{t\geq 0}\left\|e^{At}x\right\|$ is attained in finite time, say in $t_0 \geq 0$. Then we have for $y = e^{At_0}x$ that $\left\|e^{At}y\right\| \leq \|y\|$ for all $t\in\mathbb{R}_+$. Now consider the case that A is only marginally stable. If $\mathbb{K} = \mathbb{C}$ then we choose an eigenvector corresponding to a purely imaginary eigenvalue $i\omega \in \sigma(A)$. Then $\left\|e^{At}x\right\| = \left\|e^{i\omega t}x\right\| = \|x\|$. If $\mathbb{K} = \mathbb{R}$ and $A\in\mathbb{R}^{n\times n}$ is marginally stable then there exists a complex conjugate pair $\pm i\omega$ of eigenvalues of A. Let $x\in\mathbb{C}^n$ be an eigenvector associated with $i\omega$. For all $t\geq 0$ we have

$$2\left\|e^{At}\mathrm{Re}\,x\right\| = \left\|e^{At}(x+\bar{x})\right\| = \left\|e^{i\omega t}x + e^{i\omega t}\bar{x}\right\| = 2\left\|\cos(\omega t)\mathrm{Re}\,x - \sin(\omega t)\mathrm{Im}\,x\right\|,$$

which is a periodic oscillation, hence it attains its maximum in finite time t_0. Arguing as above, the trajectory starting in $y = e^{At_0}\mathrm{Re}\,x$ now satisfies $\left\|e^{At}y\right\| \leq \|y\|$ for all $t\geq 0$. Hence the eccentricity of $\|\cdot\|_A$ equals the transient amplification,

$$M_0(A) = \sup_{t\geq 0}\left\|e^{At}\right\| = \mathrm{ecc}\,\|\cdot\|_A\,. \tag{2.42}$$

To determine the initial growth of A with respect to $\|\cdot\|_A$ note that for all $t\geq 0$

$$\left\|e^{At}x\right\|_A = \sup_{s\geq 0}\left\|e^{A(s+t)}x\right\| = \sup_{s\geq t}\left\|e^{As}\right\| \leq \sup_{s\geq 0}\left\|e^{As}\right\| = \|x\|_A\,,$$

thus A generates a contraction with respect to $\|\cdot\|_A$ and the initial growth rate satisfies $\mu_A(A) \leq 0$. □

More precisely, we have the following result for the initial growth rate with respect to the Feller norm.

Corollary 2.62. *Given a stable matrix $A\in\mathbb{K}^{n\times n}$. Then the initial growth rate of A with respect to the Feller norm $\|\cdot\|_A$ satisfies $\mu_A(A) = \min\{\mu(A), 0\}$.*

Proof. If $\mu(A) \leq 0$ then by Proposition 2.42, $\left\|e^{At}x\right\| \leq e^{\mu(A)t}\|x\| \leq \|x\|$ for all $x\in\mathbb{K}^n$ and all $t\geq 0$. Hence $\|x\|_A = \sup_{t>0}\left\|e^{At}x\right\| = \|x\|$ and therefore $\mu_A(A) = \mu(A)$. Now, for $\mu(A) > 0$ Lemma 2.61 shows that $\mu_A(A) \leq 0$ Moreover, if $\mu(A) > 0$ there exist $x_0\in\mathbb{K}^n$ and $t_0 > 0$ such that $\left\|e^{At_0}x_0\right\| = \|x_0\|_A > \|x_0\|$. But for $h > 0$ with $h < t_0$, $\sup_{t>h}\left\|e^{At}x_0\right\| = \sup_{t>0}\left\|e^{At}x_0\right\| = \|x_0\|_A$ and hence $\mu_A(A) = 0$. □

Therefore if $\mu(A) \geq 0$ then the resulting Feller norm is a Liapunov norm, but not a strict Liapunov norm. Otherwise, if $\mu(A) < 0$ then the original norm $\|\cdot\|$ which coincides with $\|\cdot\|_A$ is already a strict Liapunov norm. The following lemma shows that the unit ball of a Feller norm is of simple structure.

Lemma 2.63. *Suppose that $A \in \mathbb{K}^{n\times n}$ is stable. Then the closed unit ball $\mathbb{B}_A$ of the associated Feller norm $\|\cdot\|_A$ is given by*

$$\mathbb{B}_A = \bigcap_{t\geq 0} e^{-At}\mathbb{B}, \tag{2.43}$$

where $\mathbb{B}$ is the closed unit ball of $\|\cdot\|$.

Proof. By definition, $x \in \mathbb{B}_A$ holds if and only if for all $t \geq 0$, $e^{At}x \in \mathbb{B}$, or equivalently, $x \in e^{-At}\mathbb{B}$ which gives (2.43). □

2.4.1 Transient Norms and Duality

Let us now investigate duality issues for transient norms. For dual norms we obtain the following result.

Theorem 2.64. *Suppose that $\|\cdot\|$ is a vector norm on $\mathbb{K}^n$ with associated initial growth rate $\mu(\cdot)$ and let $\mu^*(\cdot)$ denote the initial growth rate with respect to the dual norm $\|\cdot\|^*$ on $\mathbb{K}^n$. Then for all matrices $A \in \mathbb{K}^{n\times n}$ the following statements hold*

1. $\mu(A) = \mu^*(A^*)$.
2. $\mu_2(A) \leq \frac{1}{2}(\mu(A) + \mu^*(A))$.

Proof. The first statement follows directly from Proposition 2.31,

$$\mu(A) = \max_{\|x\|=1} \max_{\|y\|^*=1, \langle x,y\rangle_2=1} \operatorname{Re}\langle Ax, y\rangle_2, \qquad \mu^*(A^*) = \max_{\|y\|^*=1} \max_{\|x\|=1, \langle x,y\rangle_2=1} \operatorname{Re}\langle A^*y, x\rangle_2.$$

Now as $\operatorname{Re}\langle Ax, y\rangle_2 = \operatorname{Re}\langle A^*y, x\rangle_2$ the equality $\mu^*(A) = \mu(A^*)$ is proved. The second statement follows from the first, because

$$\begin{aligned}\mu(A) + \mu^*(A) &= \mu(A) + \mu(A^*) \geq \mu(A + A^*) \geq \alpha(A + A^*)\\ &= \lambda_{\max}(A + A^*) = 2\mu_2(A) = \mu_2(A + A^*),\end{aligned}$$

where we used that μ is a subadditive function, which is bounded from below by the spectral abscissa $\alpha(B)$, see Proposition 2.40 *(i)* and *(v)*. In case of Hermitian matrices the spectral abscissa is an eigenvalue. □

This theorem shows that given any norm, the initial growth rate for the spectral norm is the best lower bound for all mean values between the initial growth rate of a norm and the initial growth rate of its dual norm. For the following pair of dual norms, 1-norm and ∞-norm, we have $\mu_1^*(A) = \mu_\infty(A)$ for all $A \in \mathbb{K}^{n\times n}$. Part 2 of Theorem 2.64 has the following consequence.

Corollary 2.65. *Suppose that $A \in \mathbb{K}^{n\times n}$ satisfies $\mu_1(A) + \mu_\infty(A) \le 0$. Then A generates a contraction semigroup with respect to the spectral norm, such that*

$$\mu_2(A) \le \tfrac{1}{2}(\mu_1(A) + \mu_\infty(A)) \le 0.$$

Estimates involving μ_1 and μ_∞ will be studied in more detail in Chapter 5.
Now that we have treated the initial growth with respect to dual norms let us consider the eccentricities of dual norms.

Proposition 2.66. *For vector norms $\nu(\cdot)$, $\|\cdot\|$ on $\mathbb{K}^n$ it holds that*

$$\operatorname{ecc}(\nu, \|\cdot\|) = \operatorname{ecc}(\nu^*, \|\cdot\|^*).$$

Proof. It suffices to show that $\operatorname{ecc}(\nu^*, \|\cdot\|^*) \le \operatorname{ecc}(\nu, \|\cdot\|)$ since the bidual norms equal the original norms, hence

$$\operatorname{ecc}(\nu, \|\cdot\|) = \operatorname{ecc}(\nu^{**}, \|\cdot\|^{**}) \le \operatorname{ecc}(\nu^*, \|\cdot\|^*) \le \operatorname{ecc}(\nu, \|\cdot\|) \tag{2.44}$$

implies equality. To this end, let us prove that $\max_{\|y\|^*=1} \nu^*(y) = \max_{\nu(x)=1} \|x\|$ and $\min_{\|y\|^*=1} \nu^*(y) \ge \min_{\nu(x)=1} \|x\|$. To show the first of these claims note that by definition

$$\max_{\|y\|^*=1} \nu^*(y) = \max_{\|y\|^*=1} \max_{\nu(x)=1} |y^*x| = \max_{\nu(x)=1} \max_{\|y\|^*=1} |y^*x| = \max_{\nu(x)=1} \|x\| . \tag{2.45}$$

To show the second claim, we define $\alpha = \max\{\beta \,|\, \nu(\beta z) \le 1 \text{ for all } \|z\| = 1\}$. Then we have $\alpha = \min_{\nu(u)=1} \|u\|$. Now consider $\min_{\|y\|^*=1} \nu^*(y) = \min_{\|y\|^*=1} \max_{\nu(x)\le 1} |y^*x|$. Let us choose a special x in the previous formula. For this, let z be a dual vector of y which satisfies $\|z\| = 1$ and $y^*z = \|y\|^*$. By definition of α we have $\nu(\alpha z) \le 1$. Hence the special choice $x = \alpha z$ yields

$$\min_{\|y\|^*=1} \nu^*(y) = \min_{\|y\|^*=1} \max_{\nu(x)\le 1} |y^*x| \ge \min_{\|y\|^*=1} \alpha\, |y^*z| = \alpha = \min_{\nu(u)=1} \|u\| . \tag{2.46}$$

Combining (2.45), (2.46) and (2.35) we get

$$\operatorname{ecc}(\nu^*, \|\cdot\|^*) = \frac{\max_{\|y\|^*=1} \nu^*(y)}{\min_{\|y\|^*=1} \nu^*(y)} \le \frac{\max_{\nu(x)\le 1} \|x\|}{\min_{\nu(x)\le 1} \|x\|} = \operatorname{ecc}(\|\cdot\|, \nu) = \operatorname{ecc}(\nu, \|\cdot\|).$$

Hence equality follows in (2.44). □

Now, if $\|\cdot\| = \|\cdot\|_2$ is given then the dual norm of a transient norm ν satisfies by Proposition 2.66 $\operatorname{ecc}(\nu^*, \|\cdot\|_2) = \operatorname{ecc}(\nu, \|\cdot\|_2)$. Hence we can expect that ν^* is also a transient norm, but now for A^*. This is indeed true as the following corollary shows.

Corollary 2.67. *Suppose that $\|\cdot\| = \|\cdot\|_2$, i.e., all eccentricities are measured with respect to the Euclidean norm. Then the dual of a transient norm $\nu(\cdot)$ of A is a transient norm of A^*.*

Proof. By Proposition 2.66 we have $\mathrm{ecc}(\nu, \|\cdot\|_2) = \mathrm{ecc}(\nu^*, \|\cdot\|_2)$. Part 1 of Theorem 2.64 shows that $\mu^*_\nu(A^*) = \mu_\nu(A) \leq 0$. The transient amplification satisfies

$$M_0(A) = \sup_{t\geq 0} \left\|e^{At}\right\|_2 = \sup_{t\geq 0} \left\|e^{A^*t}\right\|_2 = M_0(A^*).$$

Hence ν^* is a transient norm of A^*. □

Suppose that $A \in \mathbb{K}^{n\times n}$ is stable and let us consider the norm $\widetilde{\|\cdot\|_A} := ((\|\cdot\|^*)_{A^*})^*$, that is, we start with the dual norm of $\|\cdot\|$ and construct the Feller norm with respect to A^*. This is a transient norm for A^*, hence by Corollary 2.67 its dual $\widetilde{\|\cdot\|_A}$ is a transient norm. This provides a second method of creating transient norms besides $\|\cdot\|_A$ itself. Let us now analyse this alternative method for the construction of transient norms. The following proposition shows how the unit ball $\tilde{\mathbb{B}}_A$ of $\widetilde{\|\cdot\|_A}$ is obtained from the trajectories of the system $\dot{x} = Ax$.

Proposition 2.68. *Suppose that $A \in \mathbb{K}^{n\times n}$ is a stable matrix. Let $\mathbb{B}$ be the closed unit ball of $\|\cdot\|$. The closed unit ball $\widetilde{\mathbb{B}}_A$ of the norm $\widetilde{\|\cdot\|_A} := ((\|\cdot\|^*)_{A^*})^*$ is given by*

$$\widetilde{\mathbb{B}}_A = \overline{\mathrm{conv}}\ \left\{e^{At}x \,\middle|\, t \geq 0,\, x \in \mathbb{B}\right\} = \overline{\mathrm{conv}} \bigcup_{t\geq 0} e^{At}\mathbb{B}, \tag{2.47}$$

where $\overline{\mathrm{conv}}$ denotes the closed convex hull and $\mathbb{B}$ is the closed unit ball of $\|\cdot\|$.

Proof. Recall that the dual set of a convex set $K \subset \mathbb{K}^n$ is given by

$$K^* = \{y \in \mathbb{K}^n \,|\, \forall\, x \in K:\ \mathrm{Re}\,\langle y, x\rangle_2 \leq 1\}.$$

Hence the dual of the unit ball $\mathbb{B}$ is the unit ball $\mathbb{B}^*$ of the dual norm. For a fixed $t \geq 0$ the dual set of $e^{At}\mathbb{B}$ is therefore given by $e^{-A^*t}\mathbb{B}^*$, as $x \in e^{At}\mathbb{B}$, $y \in e^{-A^*t}\mathbb{B}^*$ satisfy $\mathrm{Re}\,\langle y, x\rangle_2 = \mathrm{Re}\left\langle e^{A^*t}y, e^{-At}x\right\rangle_2 \leq 1$ by duality of $\mathbb{B}$ and $\mathbb{B}^*$. The closed unit ball of the norm $(\|\cdot\|^*)_{A^*}$ is given by $\mathbb{B}^*_{A^*} = \bigcap_{t\geq 0} e^{-A^*t}\mathbb{B}^*$, see Lemma 2.63. Its dual set can now be computed using [120, Corollary 16.5.2], which shows that the dual of a closed convex hull of the union of convex sets C_i is given by the intersection of the dual convex sets C_i^*, and therefore

$$\mathbb{B}^*_{A^*} = \bigcap_{t\geq 0} e^{-A^*t}\mathbb{B}^* = \left(\overline{\mathrm{conv}} \bigcup_{t\geq 0} \left(e^{-A^*t}\mathbb{B}^*\right)^*\right)^* = \left(\overline{\mathrm{conv}} \bigcup_{t\geq 0} e^{At}\mathbb{B}\right)^* = \left(\widetilde{\mathbb{B}}_A\right)^*.$$

Hence the unit ball of $\widetilde{\|\cdot\|_A}$ is given by (2.47). □

Let us compare the unit balls for both transient norms $\|\cdot\|_A$ and $\widetilde{\|\cdot\|_A}$. They are given by

$$\mathbb{B}_A = \left\{x \in \mathbb{B} \,\middle|\, e^{At}x \in \mathbb{B} \text{ for all } t \geq 0\right\}, \qquad \widetilde{\mathbb{B}}_A = \overline{\mathrm{conv}}\ \left\{e^{At}x \,\middle|\, x \in \mathbb{B}, t \geq 0\right\} \tag{2.48}$$

hence the first unit ball consists of all initial vectors for which the trajectory remains entirely in $\mathbb{B}$ while the latter unit ball is the smallest ball containing all trajectories starting in $\mathbb{B}$, i.e., the first is the largest A-invariant ball contained in $\mathbb{B}$, while the latter is the smallest A-invariant ball containing $\mathbb{B}$. It is easy to see that the following inclusions hold, $\mathbb{B}_A \subset \mathbb{B} \subset \widetilde{\mathbb{B}}_A$.

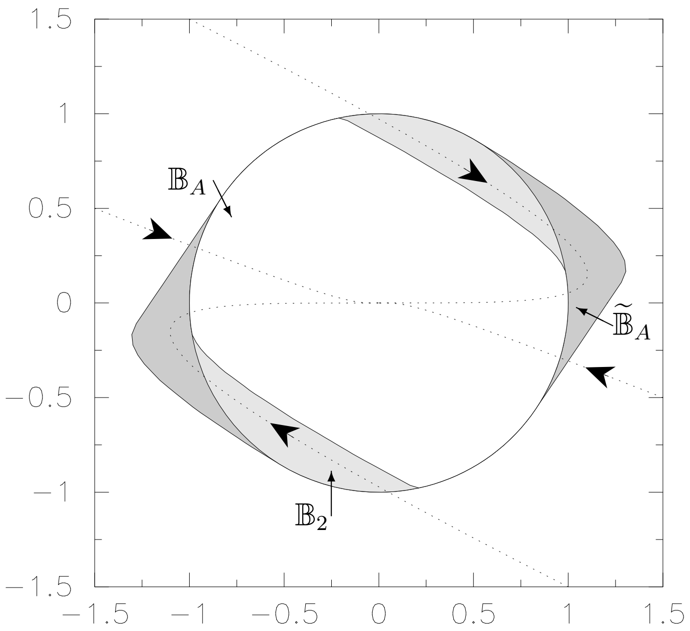

Figure 2.8: Unit balls of transient norms.

Example 2.69. Consider the stable matrix $A = \begin{pmatrix} -5 & 36 \\ 0 & -20 \end{pmatrix}$. The unit ball for its Feller norm, $\mathbb{B}_A$, and the unit ball of the transient norm, $\widetilde{\mathbb{B}}_A$, when starting from a Euclidean norm are shown in Figure 2.8. Both norms form Liapunov norms for $\dot{x} = Ax$ since their unit balls are invariant under the flow of A. Note that parts of the unit ball $\mathbb{B}_A$ and of the unit ball $\widetilde{\mathbb{B}}_A$ consist of trajectories of $\dot{x} = Ax$. Hence, these norms are not analytic as they contain segments of the unit circle and of trajectories as parts of their boundaries. ■

2.4.2 Common Liapunov Norms

We already noted in Lemma 2.44 that if $\mu(A) \leq -\delta < 0$ then $\mu(A+\Delta) < 0$ for all $\Delta \in \mathbb{K}^{n\times n}$, $\|\Delta\| < \delta$. This implies that the norm $\|\cdot\|$ is a Liapunov function for all perturbed systems $\dot{x} = (A + \Delta)x$ as long as the norm of a perturbation is bounded by δ. To generalize this concept, we introduce *linear time-invariant differential inclusions*, see Smirnov [126] and Vinter [144, Chapter 2]. We consider a set of matrices $\mathcal{A} \subset \mathbb{K}^{n\times n}$. The differential inclusion generated by this set is written formally as

$$\dot{x} \in \mathcal{A}x. \tag{2.49}$$

An absolute continuous function $x : \mathbb{R}_+ \to \mathbb{K}^n$ is called a solution of (2.49) if there exists a locally integrable function $v \in L^1_{loc}(\mathbb{R}_+, \mathbb{K}^n)$ with $v(t) \in \{Ax(t) \mid A \in \mathcal{A}\}$ almost everywhere in $\mathbb{R}_+$ such that $x(t) = x(t_0) + \int_{t_0}^{t} v(s)ds$ for $t, t_0 \in \mathbb{R}_+$.

A linear differential inclusion is *exponentially stable* if there exist constants $M \geq 1$ and $\beta < 0$ such that $\|x(t)\| \leq Me^{\beta t}\|x(0)\|$ for all $t \in \mathbb{R}_+$ and all solutions $x(\cdot)$.

It is well-known that the closure of the solution set of (2.49) (with respect to the norm $\|f\|_\infty = \sup_{t\geq 0}\|f(t)\|$) coincides with the solution set of the differential inclusion $\dot{x} \in (\overline{\operatorname{conv}}\,\mathcal{A})x$ (Theorem of Filippov-Ważewski).
We will now investigate under which condition we can switch between the different system matrices in $\mathcal{A}$ without loosing stability, or in other words, under which conditions the differential inclusion $\dot{x} \in \mathcal{A}x$ is stable. If the Liapunov function is a Liapunov norm, we can answer this question affirmatively. Before we present a proof of this fact, let us extend the inequalities of Proposition 2.42 *(i)* to time-varying differential equations. The following theorem allows us to compute that solutions of the differential equation (2.49) exist on $\mathbb{R}_+$ if $\sup_{A\in\mathcal{A}}\mu(A)$ is finite.

Theorem 2.70 (Ważewski inequalities)**.** *Consider the differential equation $\dot{x}(t) = A(t)x(t)$ on $t \in \mathbb{R}_+$ where $A(t) : \mathbb{R}_+ \to \mathbb{K}^{n\times n}$ is measurable matrix-valued function. If μ is the initial growth rate associated with a vector norm $\|\cdot\|$ on $\mathbb{K}^n$ and $\sup_{t\geq 0}\mu(A(t))$ is finite, we have for all $t \in \mathbb{R}^+$ and all initial values $x(0) = x_0 \in \mathbb{K}^n$*

$$e^{\int_0^t -\mu(-A(\theta))d\theta}\,\|x_0\| \leq \|x(t,x_0)\| \leq e^{\int_0^t \mu(A(\theta))d\theta}\,\|x_0\|\,. \tag{2.50}$$

Proof. Suppose that $x(t)$ is a solution of $x(t) = x(0)+\int_0^t A(s)x(s)ds$ with $x(0) = x_0$ and life span $I_{\max} = [0, t_{\max})$. Then x is absolutely continuous on $I_{\max}$, hence $v(t) = \dot{x}(t) = A(t)x(t)$ is a locally integrable function. Starting with the integral formulation of a solution, we obtain for $t \in I_{\max}$ and for small enough $h > 0$ that

$$x(t+h) = x(t) + \int_0^h v(t+\theta)d\theta = x(t) + \int_0^h A(t+\theta)x(t+\theta)d\theta$$

Hence taking norms,

$$\begin{aligned}\|x(t+h)\| = \left\|x(t) + \int_0^h A(t+\theta)x(t+\theta)d\theta\right\| &= \left\|x(t) + \int_0^h A(t+\theta)(x(t)+O(h))d\theta\right\|\\ \leq \left\|I + \int_0^h A(t+\theta)d\theta\right\| \|x(t)\| + O(h^2),&\\ \|x(t+h)\| - \|x(t)\| \leq \left(\left\|I + \int_0^h A(t+\theta)d\theta\right\| - 1\right)\|x(t)\| + O(h^2).&\end{aligned}$$

As A is a measurable function, $\lim_{h\searrow 0}\int_0^h A(t+\theta)d\theta = A(t)$ almost everywhere. Exploiting the monotonicity of $h^{-1}(\|I + hA(t)\| - 1)$ as $h \searrow 0$ we get

$$\tfrac{d}{dt^+}\|x(t)\| \leq \mu(A(t))\,\|x(t)\| \qquad \text{a.e.}$$

Analogously, the left derivative of $\|x(t)\|$ satisfies

$$\tfrac{d}{dt^-}\|x(t)\| \geq -\mu(-A(t))\,\|x(t)\| \qquad \text{a.e.}$$

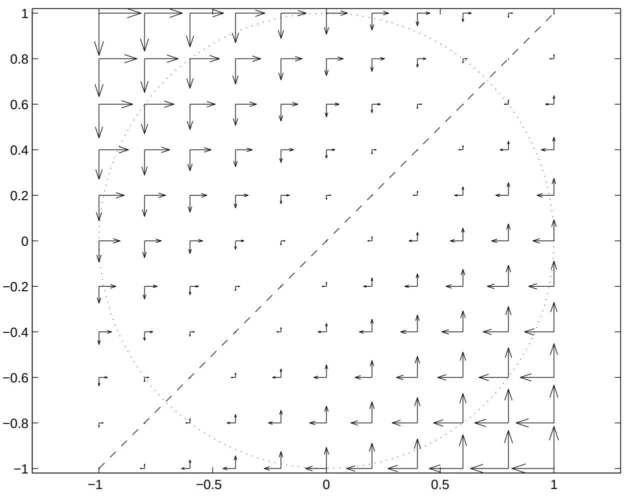

Figure 2.9: Vectorfields of $\dot{x} = A_1 x$ and $\dot{x} = A_2 x$.

Using integrating factors we obtain for $t \in I_{\max}$

$$\tfrac{d}{dt^-}\left[e^{\int_0^t \mu(-A(\theta))d\theta}\,\|x(t)\|\right] \le 0, \qquad \tfrac{d}{dt^+}\left[e^{-\int_0^t \mu(A(\theta))d\theta}\,\|x(t)\|\right] \ge 0,$$

from which (2.50) follows for all $t \in I_{\max}$. As $\sup_{t\ge 0}\mu(A(t)) < \infty$, we have $I_{\max} = \mathbb{R}_+$, whence (2.50) holds on $\mathbb{R}_+$. □

Corollary 2.71. *Given a closed set of matrices $\mathcal{A} \subset \mathbb{K}^{n\times n}$ and suppose that there exists a vector norm $\|\cdot\|$ such that the associated initial growth rate satisfies $\mu(A) < 0$ for all $A \in \mathcal{A}$. Then the differential inclusion*

$$\dot{x} \in (\overline{\operatorname{conv}}\,\mathcal{A})x \tag{2.51}$$

is exponentially stable and all solutions x satisfy the contraction property $\|x(t)\| < \|x(0)\|$ for $t > 0$.

Proof. As $\sup_{A\in\mathcal{A}}\mu(A)$ is bounded, a solution $x(t)$ of (2.51) exists on $\mathbb{R}_+$. Then we find an integrable function $v(\cdot)$ such that $x(t) = x(0) + \int_0^t v(\theta)d\theta$. We find a measurable selection $A(t) \in \operatorname{conv}\mathcal{A}$ such that $v(t) = A(t)x(t)$ for almost all $t \ge 0$, see [144, Theorem 2.3.11]. By Theorem 2.70 any solution is exponentially bounded with a negative decay rate, since by convexity $\mu(A(t)) < 0$ holds almost everywhere on $\mathbb{R}_+$, and therefore $\int_0^t \mu(A(\theta))d\theta < 0$. □

Example 2.72. Consider the two matrices $A_1 = \left(\begin{smallmatrix} 0 & 0 \\ 1 & -1 \end{smallmatrix}\right)$ and $A_2 = \left(\begin{smallmatrix} -1 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. Then one can easily see that any solution $x(t, x_0)$ of the differential inclusion $\dot{x} \in \{A_1, A_2\}x$ satisfies $\|x(t; x^0)\| \le \sqrt{2}\,\|x^0\|$ with respect to the Euclidean norm, see Figure 2.9. A common Liapunov norm is given by the maximum norm $\|x\|_\infty = \max_i |x_i|$. ■

Remark 2.73. The convex hull of a set of exponentially stable matrices does not necessarily contain only exponentially stable matrices. In particular, if $A \in \mathbb{K}^{n\times n}$ is an exponentially stable matrix with $\mu_2(A) > 0$ then $\operatorname{conv}\{A, A^*\}$ contains the instable matrix $\frac{1}{2}(A + A^*)$.

The following result which extends Corollary 2.71 can be found in Molchanov and Pyatnitskij [107].

Theorem 2.74. *The differential inclusion* (2.49) *is exponentially stable if and only if there exists a common Liapunov norm for $\mathcal{A} \subset \mathbb{K}^{n\times n}$. A suitable Liapunov norm is given by*

$$\nu(x) = \max\left\{ \left|\langle x, y^i\rangle_2\right| \,\middle|\, i = 1, \dots, m \right\}$$

for a set of vectors $y^i \in \mathbb{K}^n$, $i = 1, \dots, m$ with $\operatorname{span}\{y^i \,|\, i = 1, \dots m\} = \mathbb{K}^n$ *such that there exists $\gamma > 0$ with*

$$\sup_{A\in\mathcal{A}} \mu_\nu(x, A) \le -\gamma \left\|x\right\|_2 \quad \textit{for all } x \in \mathbb{K}^n.$$

2.5 Notes and References

Most of the material on semigroup theory used here can be found in the extensive literature on one-parameter semigroups, see e.g. [38, 113, 59]. The asymptotic growth rate is discussed in all of these references. Theorem 2.22 is a consequence of the following theorem.

Theorem 2.75 (Prüss)**.** *Let X be a Hilbert space and A a closed linear operator on X. If A is the generator of a strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$ then*

$$\omega_0(T) = \lim_{\varepsilon\to 0} \alpha_\varepsilon(A).$$

In this formulation, the result is due to Trefethen [137]. Prüss [118] uses the following characterization of the asymptotic growth bound,

$$\omega_0(T) = \inf\left\{ \omega > \alpha(A) \,\middle|\, \sup_{\operatorname{Re}\lambda>\omega} \left\|(\lambda I - A)^{-1}\right\| < \infty \right\},$$

hence for each $\omega > \omega_0$ the resolvent is uniformly bounded on $\mathbb{C}_{\ge\omega} = \{z \in \mathbb{C} \,|\, \operatorname{Re} z \ge \omega\}$. Note that Theorem 2.75 does not hold in arbitrary Banach spaces, see the comments on [38, Theorem V.1.11].

The discussion of the initial growth rate is not a standard topic, see [31, Exercises I.9.17–21]. The concept of the initial growth rate originates with works of Dahlquist [30] and Lozinskii [100], where it is coined *logarithmic norm* or *logarithmic derivative* but ideas for the spectral norm can already be found in Ważewski [145]. Bauer [11] discusses the relation to generalized fields of values. An interesting result connecting the resolvent with the numerical range is the following,

Theorem 2.76 ([113, Theorem I.3.9])**.** *Let A be a closed linear operator with dense domain in a Banach space X. If $\lambda \in \mathbb{C}$, $\lambda \notin W_{\|\cdot\|}(A)$ then $\lambda I - A$ is one-to-one and has closed range. Moreover if Σ_0 is a component of $W_{\|\cdot\|}(A)^{\mathrm{C}}$ satisfying $\varrho(A) \cap \Sigma_0 \ne \emptyset$, then the spectrum of A is contained in Σ_0^{C} and*

$$\|R(\lambda, A)\| \le \operatorname{dist}(\lambda, \bar{W}_{\|\cdot\|}(A))^{-1}.$$

More properties of the initial growth rate are given in Ström [134]. Vidyasagar [143] uses the initial growth rate under the name *matrix measure* and shows Ważewski's inequalities for an arbitrary norm. For an application of the initial growth rate to DAE systems, see Higueras and Söderlind [58]. The description of the initial growth rate via the dual norm is new, although a description of dissipative operators in terms of dual vectors is given in Engel and Nagel [38]. The related concept of *semi-scalar products* is used in Yosida [151] to characterize contraction semigroups. If A is a dissipative operator then $-A$ is sometimes called *accretive* for which characterizations are available in Kato [77].

The book of Arendt et al. [4] is devoted to the study of Laplace transformations and offers lots of additional material. For example, Theorem 2.7 is only a special case of the Post-Widder inversion formula. The notion of a Liapunov norm is introduced in [83]. For a discussion of dual vectors in finite dimensions see Horn and Johnson [70]. The generalized numerical ranges are introduced in [11]. For a relation between the Gershgorin set and the spectral numerical range $W_2(\cdot)$ see [71].

The transient amplification $M_0(A)$ has been introduced in Pritchard [117]. This article contains the ideas of many topics we discuss in the following chapters.

As already mentioned, we have traced back the usage of the transient norm $\|x\|_A$ to Feller [41]. However, as Daleckiĭ and Kreĭn [31, pp. 29, 68] note, the family of Liapunov norms

$$\|x\|_{A,p} = \left(\int_0^\infty \left\|e^{At}x\right\|^p \right)^{1/p}, \qquad p \geq 1,$$

has been introduced in lectures given by Kreĭn in 1947–1948, but these ideas were published as late as 1964 in [87]. Here the Feller norm is just a special case, $\|\cdot\|_A = \|\cdot\|_{A,\infty}$. The dual concept, the transient norm $\widetilde{\|\cdot\|}_A$, remains as of now unnamed. The notion of eccentricity for ellipses is found in classical geometry. For an application to stability issues, see for example Sarybekov [123], where the condition number of a quadratic Liapunov matrix is introduced as *quality of stability* of the associated system matrix. Wirth [150] introduces the concept of eccentricity for general norms.

For a discussion of differential inclusions see [6, 126]. A proof of the Ważewski inequalities (2.50) is found in Vidyasagar [142, Theorem 3.5.1] and Gil' [46, Corollary 4.2.5], see also the original article of Ważewski [145]. Note that the estimates for linear systems obtained from Theorem 2.70 perform better than estimates based upon Gronwall's Lemma as the latter works with $\|A\|$ which is always larger than $\mu(A)$.

This thesis only discusses continuous-time linear dynamical systems. For results on discrete-time systems, see Varga [140] and Higham [56].

Chapter 3

Bounds for the Transient Amplification

The matrix exponential of $A \in \mathbb{K}^{n\times n}$ carries all the information of the solutions of the linear time-invariant differential equation $\dot{x} = Ax$, information on both the short-term or transient behaviour and on the long-term or asymptotic behaviour. In this chapter we introduce a concept of stability that takes transient effects into account as we do not only prescribe a growth rate β but also a transient bound M, hence expanding the notion of exponential stability.

Moreover, we present old and new results for bounding the matrix exponential. We consider some upper bounds for the norm of the matrix exponential, $\left\|e^{At}\right\|$. These bounds presented here may be roughly grouped into three types:

- bounds using the spectrum of A,
- bounds using quadratic Liapunov functions, and
- bounds using the resolvent of A.

We show that bounds which depend on the spectrum of A are relatively weak when the matrix under consideration is highly nonnormal. After that we consider some results which deal with the singular value decompositions of A and of e^{At}. As a third method we consider quadratic Liapunov functions, where we show how the theory derived in Chapter 2 fits into the classical results on quadratic Liapunov functions. Finally, we take a look at bounds obtained from the resolvent.

3.1 (M, β)-Stability

We introduce a stability definition which does not only take asymptotic effects, but also transient effects into account. One can argue that this is a suitable requirement in the presence of physical constraints. Moreover, it is important to detect and handle overshoot phenomena.

Definition 3.1. Suppose $M \geq 1$, $\beta \in \mathbb{R}$ are given constants. The system matrix $A \in \mathbb{K}^{n\times n}$ of a linear time-invariant system

$$\dot{x}(t) = Ax(t), \qquad t \geq 0, \tag{3.1}$$

is said to be (M,β)-*stable* with respect to the operator norm $\|\cdot\|$ if it satisfies

$$\left\|e^{At}\right\| \leq Me^{\beta t} \quad \text{for } t \geq 0. \tag{3.2}$$

It is called *strictly* (M,β)-stable, if

$$\left\|e^{At}\right\| < Me^{\beta t} \quad \text{for } t > 0,$$

and *uniformly* (M,β)-stable if there exists $\beta' < \beta$ such that (3.2) holds with β replaced by β'. The set of all (M,β)-stable generators in $\mathbb{K}^{n\times n}$ is denoted by $\mathcal{G}(M,\beta)$.

In the case $M = 1, \beta \leq 0$ every matrix $A \in \mathcal{G}(M,\beta)$ generates a contraction semigroup. This has already been studied in Chapter 2. Unlike asymptotic stability or marginal stability in the sense of Liapunov, these stability notions depend on the chosen norm on $\mathbb{K}^{n\times n}$.
Using a transient norm as defined in Definition 2.60 we get the following description of (M,β)-stability.

Proposition 3.2. *The matrix $A \in \mathbb{K}^{n\times n}$ is (M,β)-stable if and only if there exists a Liapunov norm ν for A such that*

$$\mu_\nu(A) \leq \beta, \qquad \operatorname{ecc}\nu \leq M.$$

Proof. If $A \in \mathcal{G}(M,\beta)$ then $A - \beta I_n$ is stable and the function $\nu(x) := \|x\|_{A-\beta I} = \sup_{t\geq 0} e^{-\beta t}\left\|e^{At}x\right\|$ is finite and defines a norm. The eccentricity of this norm is given by $\operatorname{ecc}\nu = \sup\left\|e^{(A-\beta I)t}\right\| \leq M$, see (2.42). Moreover $\mu_\nu(A-\beta I) \leq 0$ by Corollary 2.62, hence $\mu_\nu(A) \leq \beta$. Hence we have found a suitable norm satisfying the conditions of the lemma. The converse implication is clear from Corollary 2.57. □

Let us now discuss uniform (M,β)-stability. Alternative proofs for the following propositions can be found in [67].

Proposition 3.3. *Given $M \geq 1$, $\beta \in \mathbb{R}$. The matrix $A \in \mathbb{K}^{n\times n}$ is uniformly (M,β)-stable if and only if* $\left\{\begin{array}{ll} \text{A is strictly (M,β)-stable with $\alpha(A) < \beta$} & \text{for $M > 1$,} \\ \mu(A) < \beta & \text{for $M = 1$.}\end{array}\right.$

Proof. Let us first study the case $M = 1$. If $\mu(A) < \beta$ then A is uniformly $(1,\beta)$-stable: It suffices to choose $\beta' = \mu(A)$. Conversely, if $\left\|e^{At}\right\| \leq e^{\beta' t}$, $t > 0$, and $\beta' < \beta$ then $\mu(A) \leq \beta' < \beta$, see Proposition 2.42. In case $M > 1$, uniform (M,β)-stability implies strict (M,β)-stability. Clearly, $\alpha(A) \leq \mu(A) < \beta$ holds by Proposition 2.40. Conversely, if A is strictly (M,β)-stable with $\alpha(A) < \beta$ then we set

$$M^* = \inf\{M \geq 1 \mid \left\|e^{(A-\beta I)t}\right\| < M \text{ for all } t > 0\}.$$

By construction we have $M^* \le M$. Let us suppose that $M^* = M$ holds. To obtain strict (M,β)-stability $\lim_{t\to\infty} \left\| e^{(A-\beta I)t} \right\| = M$ has to hold, i.e., the supremum of $\left\| e^{(A-\beta I)t} \right\|$ is obtained for $t \to \infty$. This contradicts the exponential stability of $A - \beta I$. Hence the maximum of $\left\| e^{(A-\beta)t} \right\|$ is attained for a finite $t^* \ge 0$. By continuity of the norm and of the matrix exponential there exists $\beta' < \beta$ such that $\left\| e^{At} \right\| \le M e^{\beta' t}$, $t \ge 0$. Hence A is uniformly (M,β)-stable. □

Let us investigate the topological properties of the set of (M,β)-stable matrices, see [75] for the used topological notions.

Proposition 3.4. *Suppose that $M \ge 1$, $\beta < 0$ are given constants. Then the set $\mathcal{G}(M,\beta)$ of complex (M,β)-stable matrices is closed and its interior is given by*

$$\mathring{\mathcal{G}}(M,\beta) = \left\{ A \in \mathbb{C}^{n\times n} \,\middle|\, A \text{ is uniformly } (M,\beta)\text{-stable } \right\} = \bigcup_{\beta' < \beta} \mathcal{G}(M,\beta').$$

Especially, for $A \in \mathcal{G}(M,\beta)$ we have the following perturbation result for all $\Delta \in \mathbb{C}^{n\times n}$

$$\left\| e^{(A+\Delta)t} \right\| \le M e^{\beta t} e^{M\|\Delta\| t}, \qquad t \ge 0. \tag{3.3}$$

Proof. For every converging sequence $A_k \in \mathcal{G}(M,\beta)$ with $\lim_{k\to\infty} A_k = A$, the continuity of the operator norm and the exponential gives $\left\| e^{At} \right\| = \lim_{k\to\infty} \left\| e^{A_k t} \right\| \le M e^{\beta t}$ for all $t \ge 0$, hence $A \in \mathcal{G}(M,\beta)$ and therefore $\mathcal{G}(M,\beta)$ is closed. If A is not uniformly (M,β)-stable then A belongs to the boundary of $\mathcal{G}(M,\beta)$ because $A + \varepsilon I \notin \mathcal{G}(M,\beta)$ for all $\varepsilon > 0$. Let us therefore assume that A is uniformly (M,β)-stable, i.e., there exists $\beta' < \beta$ with $\left\| e^{At} \right\| \le M e^{\beta' t}, t \ge 0$. By Proposition 3.2 we find a norm ν such that $\operatorname{ecc}\nu \le M$ and the associated growth rate satisfies $\mu_\nu(A) \le \beta'$. Then for $\Delta \in \mathbb{K}^{n\times n}$ we obtain using properties of the initial growth rate, see Proposition 2.40,

$$\left\| e^{(A+\Delta)t} \right\| \le \operatorname{ecc}\nu \cdot e^{\mu_\nu(A+\Delta)t} \le \operatorname{ecc}\nu \cdot e^{(\mu_\nu(A)+\mu_\nu(\Delta))t} \le \operatorname{ecc}\nu \cdot e^{(\mu_\nu(A)+\nu(\Delta))t}, \quad t \ge 0.$$

Now by (2.35) and (2.38) we can bound the operator norm $\nu(\Delta)$ by

$$\nu(\Delta) \le \operatorname{ecc}(\|\cdot\|, \nu) \|\Delta\| = \operatorname{ecc}(\nu, \|\cdot\|) \|\Delta\| \le M \|\Delta\|,$$

which shows (3.3). Therefore for every $\Delta \in \mathbb{K}^{n\times n}$ with $\|\Delta\| \le M^{-1}(\beta - \beta')$, the matrix $A + \Delta$ is (M,β)-stable. Therefore A is an interior point of $\mathcal{G}(M,\beta)$. Thus each uniformly (M,β)-stable matrix is contained in $\mathring{\mathcal{G}}(M,\beta)$, and clearly $\mathcal{G}(M,\beta') \subset \mathring{\mathcal{G}}(M,\beta)$ for $\beta' < \beta$. □

If $A \in \mathcal{G}(M,\beta')$ and $\Delta \in \mathbb{K}^{n\times n}$ commute then we can improve (3.3), namely,

$$\left\| e^{(A+\Delta)t} \right\| = \left\| e^{At} e^{\Delta t} \right\| \le \left\| e^{At} \right\| \left\| e^{\Delta t} \right\| \le M e^{\beta' t} e^{\mu(\Delta)t} \le M e^{(\beta' + \|\Delta\|)t}, \quad t \ge 0.$$

Here $\mu(\cdot)$ is the initial growth rate with respect to $\|\cdot\|$.

Example 3.5. Consider the matrix $A = \left(\begin{smallmatrix} -1 & 2 \\ 0 & -1 \end{smallmatrix} \right)$ studied in Example 2.13. We have seen that $\alpha(A) = -1$, $\mu_2(A) = 0$, and $\left\| e^{At} \right\|_2 < 1$ holds for all $t > 0$. Hence A is strictly $(1,0)$-stable. But it is not uniformly $(1,0)$-stable by Proposition 3.3 since $\mu_2(A) = 0$. Thus $A \in \mathcal{G}(1,0)$ lies on the boundary of $\mathcal{G}(1,0)$. In particular, for every $\varepsilon > 0$ the initial growth rate of $A_\varepsilon = \left(\begin{smallmatrix} -1 & 2 \\ \varepsilon & -1 \end{smallmatrix} \right)$ is given by $\mu_2(A_\varepsilon) = \varepsilon > 0$ which has been computed using Theorem 2.41. Therefore for $t > 0$ close to 0, $\left\| e^{A_\varepsilon t} \right\|_2 \not< 1$. ■

We now have a closer look at the transient growth $M_\beta(A)$ of A defined in Definition 2.59. It is easy to see that $M_\beta(A) = M_0(A-\beta I)$. For a stable $A \in K^{n\times n}$ the transient growth equals the eccentricity of the Feller norm associated with A, see Lemma 2.61 and equation (2.42). Let us first note the following monotonicity property.

Lemma 3.6. *Let $A \in \mathbb{K}^{n\times n}$. Then for $\beta' \geq \beta > \alpha(A)$,*

$$1 \leq M_{\beta'}(A) \leq M_\beta(A) < \infty.$$

Proof. If $\beta > \alpha(A)$ then $A-\beta I$ is exponentially stable, and $\left\|e^{(A-\beta I)t}\right\|$ is uniformly bounded for all $t \geq 0$. Thus $M_\beta(A) = \sup_{t\geq 0}\left\|e^{(A-\beta I)t}\right\|$ is finite. For $\beta \leq \beta'$ we have

$$M_{\beta'}(A) = \sup_{t\geq 0}\left\|e^{(A-\beta' I)t}\right\| = \sup_{t\geq 0}\left(e^{-(\beta'-\beta)t}\left\|e^{(A-\beta I)t}\right\|\right) \leq \sup_{t\geq 0}\left\|e^{(A-\beta I)t}\right\| = M_\beta(A)$$

as $\beta' \geq \beta$, and thus $e^{-(\beta'-\beta)t} \leq 1$ for all $t \geq 0$. □

Unfortunately, $M_0(A)$ does not depend continuously on A, as the following example shows.

Example 3.7. Consider the sequence of marginally stable matrices

$$A_k = \frac{1}{k}\begin{pmatrix} 0 & -1 \\ \mu^2 & 0 \end{pmatrix} \quad \text{for} \quad k \to \infty, \quad \mu > 1.$$

Its transient growth associated with the spectral norm is $M_0(A_k) = \mu$ as

$$e^{A_k t} = \cos(\tfrac{\mu}{k}t)I + \sin(\tfrac{\mu}{k}t)\begin{pmatrix} 0 & -\mu^{-1} \\ \mu & 0 \end{pmatrix},$$

and therefore $\sup_{t\geq 0}\left\|e^{A_k t}\right\|_2 = \left\|\left(\begin{smallmatrix} 0 & -\mu^{-1} \\ \mu & 0 \end{smallmatrix}\right)\right\|_2 = \mu$. But for $k \to \infty$, we have $A_k \to 0$ and $M_0(\lim_k A_k) = M_0(0) = 1 < \mu = \lim_k M_0(A_k)$. Starting with the Euclidean norm, the Feller norm associated with A_k is given by $\|x\|_{A_k} = \sqrt{x^*Px}$ with $P = \left(\begin{smallmatrix} \mu^2 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. As the Feller norm associated with the zero matrix is the Euclidean norm, we also have a discontinuity with respect to the formation of transient norms. If we consider the norm of the trajectories over a finite time interval and define the norm $\nu_{A,T}(x) = \sup_{t\in[0,T]}\left\|e^{At}\right\|$ for some $T < \infty$, then the Feller norm is obtained by $\lim_{T\to\infty}\nu_{A,T} = \nu_A$. However, for fixed T, $\lim_{k\to\infty}\nu_{A_k,T} = \|\cdot\|_2$. Hence the discontinuity is due to the fact that we consider an infinite time horizon. This only creates problems if we deal with marginally stable matrices. ■

One can show that $A \mapsto M_0(A)$ is lower semicontinuous on the set of stable matrices. The situation changes if we only consider exponentially stable matrices, as then $M_0(A)$ is finite and the map $A \mapsto M_0(A)$ is continuous.

Theorem 3.8 ([67, Proposition 5.5.5])**.** *The transient growth $A \mapsto M_0(A)$ is lower semicontinuous on the set of all stable matrices, and continuous on the set of all exponentially stable matrices.*

3.2 Bounds from the Spectrum

Strictly speaking, there are no bounds on $M_0(A)$ which only depend on the spectrum, some additional information from the eigenvectors is always needed. If an eigenvector basis is available, then we obtain the following classical bound. Suppose that there exists $V \in \mathbb{C}^{n\times n}$ such that $AV = V\Lambda$ where $\Lambda = \operatorname{diag}(\lambda_i)$, i.e., A is *diagonalizable*. Then

$$e^{At} = e^{V\Lambda V^{-1}t} = Ve^{\Lambda t}V^{-1} = V\operatorname{diag}(e^{\lambda_i t})V^{-1}, \qquad t \geq 0.$$

If $\|\cdot\|$ is a matrix norm on $\mathbb{C}^{n\times n}$ which satisfies

$$\|\Lambda\| = \max_i |\lambda_i| \quad \text{for every diagonal matrix } \Lambda = \operatorname{diag}(\lambda_i) \tag{3.4}$$

(especially, if $\|\cdot\|$ is an operator norm induced from a monotonic norm, see Lemma 1.9) then we have

$$\left\|e^{At}\right\| \leq \|V\| \left\|V^{-1}\right\| e^{\alpha(A)t}. \tag{3.5}$$

Hence the transient growth with respect to the asymptotic growth rate $\alpha(A)$, $M_{\alpha(A)}(A) = \sup_{t\geq 0}\left\|e^{(A-\alpha(A)I)t}\right\|$ is bounded by the *condition number* $\kappa(V) := \|V\| \left\|V^{-1}\right\|$ of an eigenvector basis $V \in \mathbb{C}^{n\times n}$. Note that it is not required that V consists of unit length eigenvectors of A. By introducing a suitable diagonal scaling matrix D the condition number $\kappa(VD)$ can be reduced. For a discussion of this topic, see Balakrishnan and Boyd [9] where an optimization strategy involving linear matrix inequalities (LMI) is presented.

The bound in (3.5) has the advantage that it is readily computable, but if A is not diagonalizable, this bound is of no use. Moreover, as we are mostly interested in nonnormal matrices, the condition numbers of the eigenvector matrix V tend to be large. Nevertheless, if $\alpha(A)$ is negative and of large modulus this upper bound quickly decays and can be used to identify an interval $I = [0, t_1]$ which has to contain the maximum of $t \mapsto \left\|e^{At}\right\|$.

Corollary 3.9. *Let $A \in \mathbb{K}^{n\times n}$ be stable and $AV = V\Lambda$ with $\Lambda = \operatorname{diag}(\lambda_i)$, $\lambda_i \in \sigma(A)$. If $t_1 = -\frac{\log \kappa(V)}{\alpha(A)}$ then $\left\|e^{At}\right\| \leq 1$ for $t \geq t_1$.*

Let us generalize the bound (3.5) to non-diagonalizable matrices where we now fix the norm to be the spectral norm. Instead of diagonalizing $A \in \mathbb{C}^{n\times n}$ itself we transform a scalar multiple $\delta^{-1}A$ with $\delta > 0$ into Jordan canonical form, whence $A = \delta V_\delta J_\delta V_\delta^{-1}$. Let us split J_δ into the diagonal matrix $\delta^{-1}\Lambda$, where $\operatorname{diag}(\Lambda)$ contains the eigenvalues of A, and the nilpotent matrix $N = (n_{ij})$ which only contains non-zero entries in the first off-diagonal $n_{i,i+1} \in \{0, 1\}$, $i = 1, \ldots, n-1$. Then we have $A = V_\delta(\Lambda + \delta N)V_\delta^{-1}$. The matrix $\Lambda + \delta N$ has the same Jordan structure as A. Hence we can choose V_δ in such way that the order of the Jordan blocks stays the same regardless of δ. Then N is independent of δ. As N is nilpotent there exists $k \leq n$ such that $N^{k-1} \neq 0$ and $N^k = 0$. Moreover, $\left\|N^\ell\right\|_2 = 1$ for all $\ell = 1, \ldots, k-1$. As the matrices Λ and N commute, we have for all $\delta > 0$ that

$$\begin{aligned}\left\|e^{At}\right\|_2 = \left\|e^{V_\delta(\Lambda+\delta N)V_\delta^{-1}t}\right\|_2 &\leq \|V_\delta\|_2 \left\|V_\delta^{-1}\right\|_2 \left\|e^{\Lambda t}\right\|_2 \left\|e^{\delta N t}\right\|_2 \\ &\leq \kappa_2(V_\delta)e^{\alpha(A)t}\sum_{\ell=0}^{k-1}(\delta t)^\ell, \qquad t \geq 0.\end{aligned} \tag{3.6}$$

Example 3.10. Consider the matrix $A = \left(\begin{smallmatrix} -1 & \gamma \\ 0 & -1 \end{smallmatrix}\right)$. Then $V_\delta = \operatorname{diag}(\gamma/\delta, 1)$ and $\kappa(V_\delta) = \max(|\gamma|/\delta, \delta/|\gamma|)$. Hence the minimal condition number $\kappa_2(V_\delta) = 1$ is attained at $\delta = |\gamma|$. Figure 3.1 illustrates some bounds for $\gamma = 5$. ■

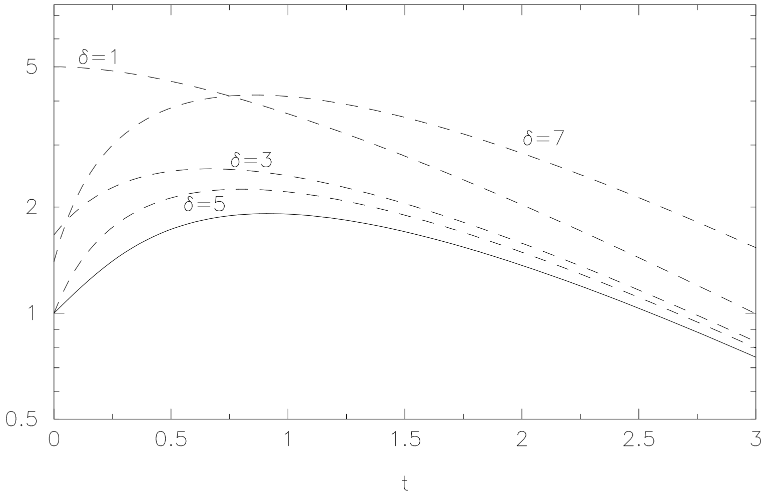

Figure 3.1: Growth bounds (3.6) for a non-diagonalizable matrix.

Unfortunately, the computation of a Jordan normal form is numerically intractable so that bounds of this type are of little practical use. We therefore need a different approach for non-diagonalizable matrices. Let us assume that the matrix norm is invariant under unitary transformations. Then we can replace A by its Schur form without loosing information. The following bound utilizes the upper triangular structure of the Schur form. Let us first introduce a measure of nonnormality based upon the Schur form.

Definition 3.11. The *departure from normality* of a matrix $A \in \mathbb{C}^{n\times n}$ with respect to a unitarily invariant norm $\|\cdot\|$ on $\mathbb{C}^{n\times n}$ is defined by

$$\operatorname{dep}(A) := \min\left\{ \|N\| \;\middle|\; \begin{array}{l} \text{There exists an unitary } U \in \mathbb{C}^{n\times n} \text{ such that } U^*AU = D + N, \\ \text{where } D \text{ is diagonal and } N \text{ is strictly upper triangular.} \end{array} \right\}.$$

This measure of normality was introduced by Henrici [54]. The following bound can be found in [138] without direct reference to the departure from normality.

Proposition 3.12. *Let $A \in \mathbb{C}^{n\times n}$ and $\|\cdot\|$ be a monotonic unitarily invariant norm on $\mathbb{C}^n$. Then the associated operator norm satisfies for all $t \geq 0$,*

$$\left\|e^{At}\right\| \leq e^{\alpha(A)t} \sum_{k=0}^{n-1} \frac{(t \operatorname{dep}(A))^k}{k!}. \tag{3.7}$$

Proof. The result is based on the fact that the matrix exponential of a perturbed matrix $A_1 + \Delta_1$ may be interpreted as a solution of the matrix-valued differential equation $\dot{X}(t) =$

$A_1X(t)+\Delta_1X(t)$ with $X(0)=I_n$. The variation-of-constants formula then gives

$$e^{(A_1+\Delta_1)t}=e^{A_1t}+\int_0^t e^{A_1(t-\theta)}\Delta_1 e^{(A_1+\Delta_1)\theta}d\theta. \tag{3.8}$$

This equation can be expanded recursively. But if Δ_1 is nilpotent then this process terminates after finitely many steps. As the norm is unitarily invariant, we may assume without loss of generality that A is given in a Schur form where the strictly upper triangular part N has the smallest norm with respect to all Schur forms of A. Hence its norm is the departure of normality of A. Writing $A=D+N$, we have decomposed A into a diagonal and a nilpotent part where $D=\mathrm{Diag}(A)$ is the diagonal matrix with the diagonal entries of A and $N=A-\mathrm{Diag}(A)$ is strictly upper triangular. Repeated use of (3.8) with $A_1=D,\Delta_1=N$ gives

$$\begin{aligned}e^{(D+N)t}&=e^{Dt}+\int_0^t e^{D(t-t_1)}Ne^{(D+N)t_1}dt_1\\&=e^{Dt}+\int_0^t e^{D(t-t_1)}Ne^{Dt_1}dt_1+\int_0^t e^{D(t-t_1)}N\int_0^{t_1}e^{D(t_1-t_2)}Ne^{(D+N)t_2}dt_2dt_1.\end{aligned}$$

Continuing this process we obtain

$$\begin{aligned}e^{(D+N)t}&=e^{Dt}+\sum_{k=1}^{n-1}A_k+R_n,\ \text{where} \qquad (3.9)\\A_k(t)&=\int_0^t\int_0^{t_1}\dots\int_0^{t_{k-1}}e^{D(t-t_1)}Ne^{D(t_1-t_2)}N\dots Ne^{Dt_k}dt_k\dots dt_1,\\R_n(t)&=\int_0^t\int_0^{t_1}\dots\int_0^{t_{n-1}}e^{D(t-t_1)}Ne^{D(t_1-t_2)}N\dots Ne^{(D+N)t_n}dt_n\dots dt_1.\end{aligned}$$

As all the factors $e^{D(t_i-t_j)}N$ are strictly upper triangular, the product of n of these terms is 0, and so $R_n(t)=0$. By Lemma 1.9, $\left\|e^{Dt}\right\|=e^{\alpha(A)t}$.
The norm of the innermost integral of A_k is bounded by

$$\begin{aligned}\left\|\int_0^{t_{k-1}}e^{D(t_{k-1}-t_k)}Ne^{Dt_k}dt_k\right\|&\le\int_0^{t_{k-1}}e^{\alpha(A)(t_{k-1}-t_k)}\left\|N\right\|e^{\alpha(A)t_k}dt_k\\&=e^{\alpha(A)(t_{k-1})}\left\|N\right\|\int_0^{t_{k-1}}dt_k=e^{\alpha(A)(t_{k-1})}\left\|N\right\|t_{k-1}.\end{aligned}$$

Hence $\|A_k(t)\|\le e^{\alpha(A)t}(k!)^{-1}(\|N\|\,t)^k$. Taking norms in (3.9) therefore gives (3.7). □

The advantage of this bound is that it equals 1 in $t=0$, hence it is better suited for the approximation of $\left\|e^{At}\right\|$ than purely exponential estimates of the form $Me^{\beta t}$. Moreover, every Schur form of A gives rise to such a bound (3.7). However, the computation of the best bound via the departure from normality may not be tractable for higher dimensions as all possible Schur forms have to be tested. A simplification occurs when considering the Frobenius norm $\|A\|_F = (\sum_{ij} |a_{ij}|^2)^{1/2}$. Here the departure from normality is constant over all Schur forms. Henrici [54] derives the following formula

$$\operatorname{dep}_F(A) = \sqrt{\|A\|_F^2 - \sum_{\lambda_i \in \sigma(A)} |\lambda_i|^2},$$

which in case of a real spectrum reduces to $\operatorname{dep}_F(A) = \sqrt{\|A\|_F^2 - \operatorname{trace} A^2}$.

We also note that for $A \in \mathbb{C}^{2\times 2}$, the Frobenius-departure from normality coincides with the spectral departure from normality, $\operatorname{dep}_F(A) = \operatorname{dep}_2(A)$ as the strictly upper triangular matrix is of rank 1, and so its Frobenius and spectral norm are equal. However, the bound in Proposition 3.12 is not valid for the Frobenius norm, as $\|I\|_F \neq 1$. An estimate of the spectral norm of e^{At} in terms of the Frobenius-departure from normality is found in [46, Corollary 2.1.6].

Example 3.13. Let us reconsider the matrices A_γ discussed in Example 3.10. The departure from normality is given by $\operatorname{dep}_2(A_\gamma) = |\gamma|$ and Proposition 3.12 yields the estimate $\left\|e^{A_\gamma t}\right\|_2 \leq e^{-t}(1 + |\gamma|\, t)$ which coincides with the best bound obtained from (3.6). Let us now take a look at the matrix

$$A = \begin{pmatrix} -4 & 32 & -72 \\ & -2 & 6 \\ & & -1 \end{pmatrix} \quad \text{with a departure from normality } \operatorname{dep}_2(A) = 78.90.$$

Note that this moderate departure from normality leads to an upper bound in (3.7) which is way off, see Figure 3.2 (note the logarithmic scale). ■

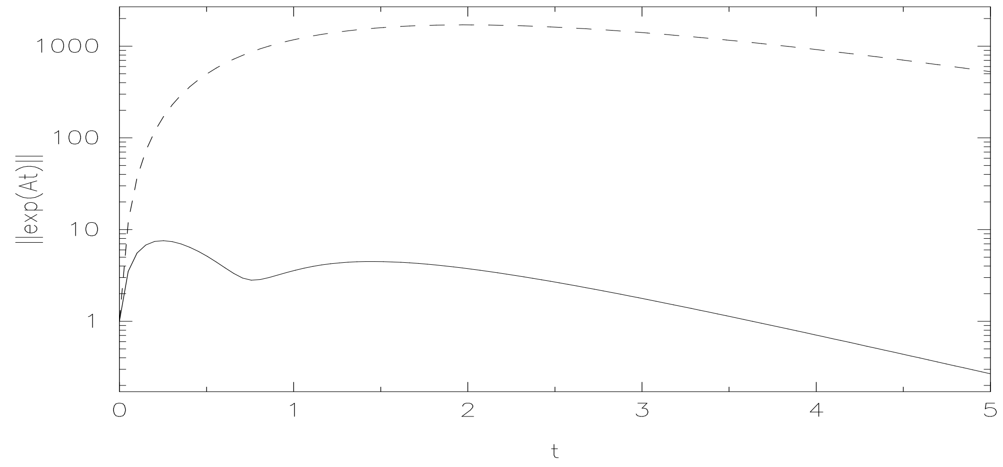

Figure 3.2: Bound based upon the departure of normality.

We assume now that $A \in \mathbb{C}^{n\times n}$ has n linearly independent eigenvectors and that the eigenvalue with the largest real part is uniquely determined, i.e.

$$\alpha(A) = \operatorname{Re}\lambda_1 > \operatorname{Re}\lambda_2 \geq \operatorname{Re}\lambda_3 \geq \cdots \geq \operatorname{Re}\lambda_n.$$

Let us denote the eigenpairs of A by (λ_i, v^i) where $v^i \in \mathbb{C}^n$, $\|v^i\| = 1$ is an eigenvector of A associated with the eigenvalue λ_i. Then for each initial value $x^0 = \sum_{i=1}^n a_i v^i$ the solution $x(t, x^0)$ of $\dot{x} = Ax$, $x(0) = x^0$ satisfies

$$\begin{aligned} \left\|x(t,x^0)\right\| &= \left\| e^{At} \sum_{i=1}^n a_i v^i \right\| = \left\| \sum_{i=1}^n a_i e^{\lambda_i t} v^i \right\| \\ &\le e^{\alpha(A)t} \left(|a_1| + \sum_{i=2}^n e^{-(\alpha(A) - \operatorname{Re}\lambda_i)t} |a_i| \right) \to e^{\alpha(A)t} |a_1| \quad \text{as } t \to \infty. \end{aligned} \tag{3.10}$$

From this calculation we immediately get the following result.

Proposition 3.14. *Let $A \in \mathbb{C}^{n\times n}$ such that A is diagonalizable with $AV = V\Lambda$, $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ and the leading eigenvalue λ_1 with $\operatorname{Re}\lambda_1 = \alpha(A)$ is uniquely determined. Then*

$$\left\|e^{At}\right\| \approx e^{\alpha(A)t} \sup_{\|x\|=1} \left|e_1^\top V^{-1} x\right|, \qquad t \gg 0.$$

Proof. From (3.10) we conclude that we have to extract the first coordinate of $x \in \mathbb{C}^n$ with respect to the transformation induced by the matrix $V = [v^1 \dots v^n]$. This projection is given by $\pi_1 : x = \sum_{i=1}^n \alpha_i v^i \mapsto \alpha_1$, or, equivalently, $\pi_1(x) = e_1^\top V^{-1} x$. Maximization over all x with $\|x\| = 1$ yields that $\left\|e^{At}\right\| - e^{\alpha(A)t} \sup_{\|x\|=1} \left|e_1^\top V^{-1} x\right| \to 0$ as $t \to \infty$. □

Clearly, the rate of this approximation is influenced by the difference $\operatorname{Re}(\lambda_1 - \lambda_2) > 0$, the larger this value the more dominant the eigenmotion corresponding to $\alpha(A)$ becomes compared to the eigenmotions of smaller eigenvalues.

Example 3.15. Consider the matrix $A = \left(\begin{smallmatrix} -5 & 36 \\ 0 & -20 \end{smallmatrix}\right)$ which we studied in Example 2.69. Then $V = \left(\begin{smallmatrix} 1 & -12 \\ 0 & 5 \end{smallmatrix}\right)$ is a matrix consisting of eigenvectors of A. Here $\sup_{\|x\|_2=1} \left|e_1^\top V^{-1} x\right| = 2.6$ while the spectral condition number of V is 33.97. This condition number can be reduced by renormalizing V, but then still $\kappa(V) = 5$. Figure 3.3 shows that the norm of the matrix exponential of A is approximated well by the bound of Proposition 3.14 for large t. ■

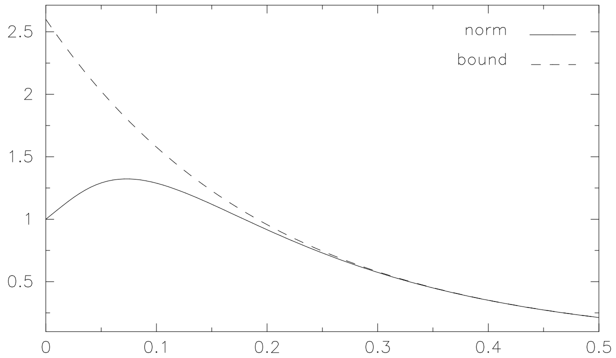

Figure 3.3: Bound based upon the dominant eigenvalue.

We have seen in this section that estimates for the norm of the matrix exponential not only require knowledge of (parts of) the spectrum, but also information about the eigenvectors. Moreover, the bounds derived in this section are mostly of interest for asymptotic approximations.

3.3 Bounds from Singular Value Decompositions

In this section we fix the matrix norm to be the spectral norm. Let us recall the definition of the singular value decomposition.

Theorem 3.16 (Singular Value Decomposition)**.** *If $A \in \mathbb{K}^{m\times n}$ is a matrix of rank r then there exist unitary or orthogonal matrices (if $\mathbb{K}=\mathbb{C}$ or $\mathbb{K}=\mathbb{R}$, respectively)*

$$U = [u_1,\dots,u_m] \in \mathbb{K}^{m\times m} \quad \textit{and} \quad V = [v_1,\dots,v_n] \in \mathbb{K}^{n\times n}$$

such that

$$U^*AV = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}_{m\times n} \qquad \textit{where} \quad \Sigma = \operatorname{diag}(\sigma_1,\dots,\sigma_r) \quad \textit{with} \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0.$$

The σ_k are called the *singular values* of A, and u_k and v_k are the kth *left singular vector* and the kth *right singular vector* of A, respectively. Here we are only interested in the case $n=m$. The singular value decomposition (SVD) allows us to decompose each matrix $A \in \mathbb{K}^{n\times n}$ into a sum of rank-one matrices,

$$A = \sum_{k=1}^{n} \sigma_k u_k v_k^*, \qquad \text{where} \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0 \quad \text{and} \quad v_i^* v_j = u_i^* u_j = \delta_{ij}. \tag{3.11}$$

Hence $Av_k = \sigma_k u_k$ and $A^*u_k = \sigma_k v_k$ hold for $k = 1,\dots,n$. Moreover, this implies that v_k is an eigenvector corresponding to the eigenvalue σ_k^2 of A^*A and analogously that u_k is an eigenvector corresponding to the eigenvalue σ_k^2 of AA^*. The spectral norm of A is given by $\|A\|_2 = \sigma_1(A)$. The *dyadic decomposition* (3.11) can now be used in two possible ways for obtaining exponential bounds.

- We decompose A and derive results from the Campbell-Baker-Hausdorff Theorem.
- We use a SVD of e^{At} and get conditions for local maxima of $\left\|e^{At}\right\|_2$.

Before we enter this analysis let us consider the case when A is a scalar multiple of an idempotent matrix $P = P^2$ (especially if A is a rank-one matrix). We need the following lemma.

Lemma 3.17. *Suppose that $f : \mathbb{C} \to \mathbb{C}$ is an entire function defined by its Taylor series $f(s) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} s^k$. Then the associated matrix-valued function $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$, $A \mapsto \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k$ satisfies for idempotent matrices $P \in \mathbb{C}^{n\times n}$, $P = P^2$, and $s \in \mathbb{C}$*

$$f(Ps) = f(0)(I-P) + Pf(s).$$

Proof. The matrix-valued function $s \mapsto f(Ps)$ is defined on $\mathbb{C}$. It is given by

$$\begin{aligned} f(Ps) &= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}(Ps)^k = f(0)I + P\sum_{k=1}^{\infty} \frac{f^{(k)}(0)}{k!} s^k \\ &= f(0)I + P(f(s) - f(0)) = (I-P)f(0) + Pf(s). \end{aligned}$$

□

Application of this lemma to the matrix exponential gives the following result.

Proposition 3.18. *Let $\lambda \in \mathbb{C}$ and $P \in \mathbb{C}^{n\times n}$ with $P = P^2$. Then*

$$e^{\lambda P t} = (I - P) + Pe^{\lambda t}, \quad t \geq 0.$$

If $\operatorname{Re}\lambda < 0$ *this implies that*

$$\lim_{t\to\infty} e^{\lambda P t} = I - P.$$

Every idempotent matrix P defines a projection $x \mapsto Px$ from $\mathbb{K}^n$ onto $\operatorname{im} P$ along the complementary subspace $\ker P$.
The following corollary gathers some facts for rank-one matrices.

Corollary 3.19. *Let $A = \sigma u v^* \in \mathbb{C}^{n\times n}$ be the SVD of a rank-one matrix. Then A has only one non-trivial eigenvalue given by $\lambda := \operatorname{trace} A = \sigma v^* u$. Its associated right eigenvector is given by the left singular vector u, and the left eigenvector is given by the right singular vector v. The matrix exponential of A is given by*

$$e^{At} = (I - \tfrac{A}{\operatorname{trace} A}) + \tfrac{A}{\operatorname{trace} A} e^{t \operatorname{trace} A}.$$

Proof. The trace is the sum of all eigenvalues. But if there is only one nonzero eigenvalue, then for $n \geq 2$ we have $\operatorname{trace} A \in \sigma(A) = \{0, \operatorname{trace} A\}$. Now $\operatorname{trace} A = \sigma \operatorname{trace} uv^* = \sigma \sum_{i=1}^n u_i \bar{v}_i = \sigma v^* u$. The right eigenvector corresponding to $\lambda = \operatorname{trace} A$ is given by u, as $Au = (\sigma u v^*)u = \lambda u$, and, analogously, the left eigenvector is given by v. The spectrum of $P = \frac{A}{\operatorname{trace} A}$ is given by $\{0, 1\}$, and P is idempotent, $P^2 = \frac{uv^*}{v^*u}\frac{uv^*}{v^*u} = \frac{v^*u}{v^*u}\frac{uv^*}{v^*u} = P$. For the matrix exponential of A, we have $A = (\operatorname{trace} A)P$ and hence by Proposition 3.18, $e^{At} = (I-P) + Pe^{t \operatorname{trace} A}$. □

Hence the matrix exponential e^{At} is a continuous deformation from $e^{A\cdot 0} = I_n$ to the projection onto the complement, $\lim_{t\to\infty} e^{At} = I - \frac{A}{\operatorname{trace} A}$, if A is of rank 1.

Corollary 3.20. *Suppose that $A = \sigma u v^* \in \mathbb{C}^{n\times n}$ is a matrix of rank one where $\sigma > 0$ and $u, v \in \mathbb{C}^n$, $\|u\|_2 = 1 = \|v\|_2$ satisfy* $\operatorname{Re} v^* u < 0$. *Then $\left\|e^{At}\right\|_2$ is a monotonously increasing function as $t \to \infty$ and*

$$\sup_{t\geq 0} \left\|e^{At}\right\|_2 = |v^* u|^{-1}.$$

Proof. We set $P = \frac{A}{\operatorname{trace} A}$. To show that the norm of the matrix exponential of A is convex, we first show that the function

$$g(a) = \|Pa + (I-P)\|_2 = \|aI + (1-a)(I-P)\|_2\,, \qquad a \in \mathbb{R}, \tag{3.12}$$

is convex. Namely, for $\theta \in (0,1)$ and $a \neq b$, $a, b \in \mathbb{R}$

$$\begin{aligned} g(\theta a + (1-\theta)b) &= \|(\theta a + (1-\theta)b)\, I + (\theta + (1-\theta) - \theta a - (1-\theta)b)\,(I-P)\|_2 \\ &= \|\theta\,(aI + (1-a)(I-P)) + (1-\theta)\,(bI + (1-b)(I-P))\|_2 \\ &\leq \theta g(a) + (1-\theta)g(b). \end{aligned}$$

Let us now determine the minimum of g. Note that $g(1+b) = \|I + bP\|_2 \geq 1$ for all $b \in \mathbb{R}$. Let us assume that $\|I + bP\|_2 < 1$ then $I - (I + bP) = bP$ would be invertible, which contradicts $\operatorname{rank} P = 1$. Hence a local minimum of g is attained in $b = 0$, as $g(1) = \|I\|_2 = 1$. With $\lambda = \operatorname{trace} A$ and $a = e^{\lambda t}$ we obtain from (3.12) and Corollary 3.19 that $\left\|e^{At}\right\|_2 = g(e^{\lambda t})$ holds by Proposition 3.18. The convexity of g implies that $\left\|e^{At}\right\|_2$ is a monotone increasing function for $t \geq 0$. By Proposition 3.18, $\lim_{t\to\infty} \left\|e^{At}\right\|_2 = \|I - P\|_2$ holds. Since P is idempotent there exists a unitary transformation U such that $UPU^* = \left(\begin{smallmatrix} I & P' \\ 0 & 0 \end{smallmatrix}\right)$. Now by Corollary 4.3 (see below) we have $\|I - P\|_2 = \|P\|_2$. For the norm of $P = \frac{uv^*}{v^*u}$, consider $P^*P = \frac{vv^*}{|u^*v|^2}$ from which we see that v is an eigenvector for the sole nonzero eigenvalue $|v^*u|^{-2} \geq (\|u\|\,\|v\|)^{-2} = 1$. Hence if $\operatorname{Re} \operatorname{trace} A < 0$ then $\sup_{t\geq 0} \left\|e^{At}\right\|_2 = \lim_{t\to\infty} \left\|e^{At}\right\|_2 = \|I - P\|_2 = |v^*u|^{-1} \geq 1$. □

Hence the transient amplification $M_0(A)$ is given by the inverse of the cosine of the angle spanned by the left and right singular vectors, which are also eigenvectors associated with the nonzero eigenvalue of A, see Corollary 3.19. This quantity $|v^*u|^{-1}$ is also called the condition number of the associated eigenvalue, see [56].

3.3.1 Decomposing A

Let us now return to general matrices and consider the spectral norm of the matrix exponential. Given two matrices $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{n\times n}$, $[A, B] = AB - BA$ denotes their *Lie bracket* or *commutator*. We may write the product of the matrix exponential e^{At} and e^{Bt} as the matrix exponential of a matrix-valued function $C(t)$. This result is known as the *Campbell-Baker-Hausdorff formula* [124, Theorem I.IV.7.4]. The first terms of the Taylor series of $C(t)$ with $e^{C(t)} = e^{At}e^{Bt}$ are given by

$$C(t) = (A+B)t + \tfrac{1}{2}[A,B]t^2 + \tfrac{1}{12}([A,[A,B]] + [B,[B,A]])t^3 + \tfrac{1}{24}[A,[[A,B],B]]t^4 + O(t^5). \tag{3.13}$$

Then we find the following approximation.

Proposition 3.21. *Given $A \in \mathbb{K}^{n\times n}$. Then*

$$\begin{aligned} \log \left\|e^{At}\right\|_2 = \tfrac{1}{2}\lambda_{\max}\big((A + A^*)t &+ \tfrac{1}{2}[A^*, A]t^2 \\ &+ \tfrac{1}{12}([A^*,[A^*,A]] + [A,[A,A^*]])t^3 + \tfrac{1}{24}[A^*,[[A^*,A],A]]t^4 + O(t^5)\big). \end{aligned} \tag{3.14}$$

Proof. The spectral norm of the matrix exponential of A is given by the square root of the largest eigenvalue of $e^{A^*t}e^{At}$. There exists a Hermitian matrix function $C(t)$ such that $e^{A^*t}e^{At} = e^{C(t)}$ where $C(t)$ is obtained from (3.13) by replacing A with A^* and B with A. Hence the spectral norm of e^{At} is given by

$$\sqrt{\lambda_{\max}(e^{C(t)})} = e^{1/2\lambda_{\max}(C(t))},$$

which proves (3.14). □

For convenience, let us compute all the Lie brackets in (3.14),

$$\begin{aligned}
[A^*, A] &= A^*A - AA^*,\\
[A^*, [A^*, A]] + [A, [A, A^*]] &= A^{*2}A + A^*A^2 - 2A^*AA^* + A^2A^* + AA^{*2} - 2AA^*A,\\
[A^*, [[A^*, A], A]] &= A^{*2}A^2 - 2(A^*A)^2 - A^2A^{*2} + 2(AA^*)^2.
\end{aligned}$$

If we now partition $A = A_0 + A_1$ with $A_0 = \sigma_1 u_1 v_1^*$ and $A_1 = \sum_{k>1} \sigma_k u_k v_k^*$ where σ_k, u_k, v_k stem from a singular value decomposition (Theorem 3.16), then $[A_0^*, A_1] = 0 = [A_1^*, A_0]$. The Lie bracket $[A^*, A]$ then simplifies to

$$[A^*, A] = [(A_0 + A_1)^*, A_0 + A_1] = A_0^*A_0 - A_0A_0^* + A_1^*A_1 - A_1A_1^* = [A_0^*, A_0] + [A_1^*, A_1].$$

Iteration of this decomposition on the tail A_1 gives us the following result.

Proposition 3.22. *Suppose that $A \in \mathbb{K}^{n\times n}$ has a singular value decomposition given by $A = \sum_{i=1}^n \sigma_i u_i v_i^*$. Then $\left\|e^{At}\right\|_2 = e^{1/2\lambda_{\max}C(t)}$ where*

$$C(t) = \sum_{k=1}^n \sigma_k(u_kv_k^* + v_ku_k^*)t + \sigma_k^2(v_kv_k^* - u_ku_k^*)t^2 + O(t^3).$$

Proof. Consider the dyadic decomposition $A = \sum_{k=1}^n \sigma_k u_k v_k^*$ and set $A_k = \sigma_k u_k v_k^*$, $k = 1, \dots, n$. Then

$$A + A^* = \sum_{k=1}^n A_k + A_k^* = \sum_{k=1}^n \sigma_k(u_kv_k^* + v_ku_k^*). \tag{3.15}$$

The Lie bracket $[A^*, A]$ now satisfies

$$[A^*, A] = \sum_{k=1}^n [A_k^*, A_k] = \sum_{k=1}^n \sigma_k^2(v_kv_k^* - u_ku_k^*) \tag{3.16}$$

as $[A_k^*, A_j] = 0$ for $k \neq j$. Using these explicit formulas (3.15) and (3.16) in Proposition 3.21 gives the required result. □

The bounds derived in Propositions 3.21 and 3.22 are only valid for small $t > 0$. Moreover, the expansion (3.14) is an extension of the growth bound $\left\|e^{At}\right\|_2 \le e^{\mu_2(A)t}$ presented in Proposition 2.42.

However, there seems to be more to this topic. When trying to generalize Corollary 3.20 to matrices of full rank, numerical experiments show the following remarkable behaviour.

Conjecture 3.23. Let $A \in \mathbb{K}^{n\times n}$ be a stable matrix with SVD $A = \sum_{i=1}^{n} \sigma_i u_i v_i^*$. Then under suitable conditions $\max_{t\geq 0} \left\| e^{At} \right\| \approx |v_i^* u_i|^{-1}$ where $i \in \{1, \ldots, n\}$ minimizes $|\sigma_i v_i^* u_i|$.

There always seems to be an index $i \in \{1, \ldots, n\}$ such that the term $|v_i^* u_i|^{-1}$ is of the right order of magnitude when compared with $M_0(A)$. It is not clear how to choose this index to achieve a good match. The method given in Conjecture 3.23 works quite well, but may fail miserably, if the singular vectors become perpendicular. Let us illustrate the problems related to this conjecture with the following example.

Example 3.24. We consider the following parameterized family of matrices

$$A_\tau = \begin{pmatrix} -1.5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -0.5 \end{pmatrix} + \tau \begin{pmatrix} 0 & 5 & -12 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \tau \in \mathbb{R}. \tag{3.17}$$

The spectrum of A_τ is constant for all τ. As τ enters linearly into the departure from normality $\operatorname{dep}(A)$ we expect some interesting transient behaviour. Figure 3.4 shows some experiments. The dashed line is the bound predicted by the Conjecture 3.23, while the dotted lines provide the values $|u_i^* v_i|^{-1}$ for all other indices. From these images we deduce that in this example the approximation performs well for $\tau \in [0, 1.5]$ and $\tau \geq 5$. However, in the suboptimal regions it seems that some dotted line can take over the role of the best approximation. Near $\tau = 2.18$ and $\tau = 2.52$, $|v_i^* u_i|^{-1}$ is infinite, as the left and right singular values become orthogonal. ■

In accordance with the notions for the sensitivity analysis of eigenvectors, [48, Section 7.2.2], we may call the term $|v_i^* u_i|^{-1}$ the *condition* of the singular value σ_i.

3.3.2 Decomposing e^{At}

Let us now consider singular decompositions of e^{At}. As $\sigma_1(e^{At}) = \left\| e^{At} \right\|_2$ we can find conditions for critical points of $t \mapsto \left\| e^{At} \right\|_2$. Let us first note the following fact about the SVD of parameter-dependent matrices.

Theorem 3.25. *Let $T : I \to \mathbb{K}^{n\times n}$ be an analytical function and $I \subset \mathbb{R}$ an open interval. Then there exist continuous and piecewise real analytic functions $\sigma_i : I \to \mathbb{R}_+$, $i = 1, \ldots, n$, with*

$$\sigma_1(t) \geq \sigma_2(t) \geq \cdots \geq \sigma_n(t) \geq 0, \qquad i = 1, \ldots, n,\ t \in I, \tag{3.18}$$

and piecewise analytic functions $u_i, v_i : I \to \mathbb{K}^n$, $i = 1, \ldots, n$ with

$$u_i(t)^* u_j(t) = \delta_{ij} \text{ and } v_i(t)^* v_j(t) = \delta_{ij} \quad \text{for all } i, j = 1, \ldots, n,\ t \in I,$$

such that

$$T(t) = \sum_{k=1}^{n} \sigma_k(t) u_k(t) v_k(t)^*, \qquad t \in I. \tag{3.19}$$

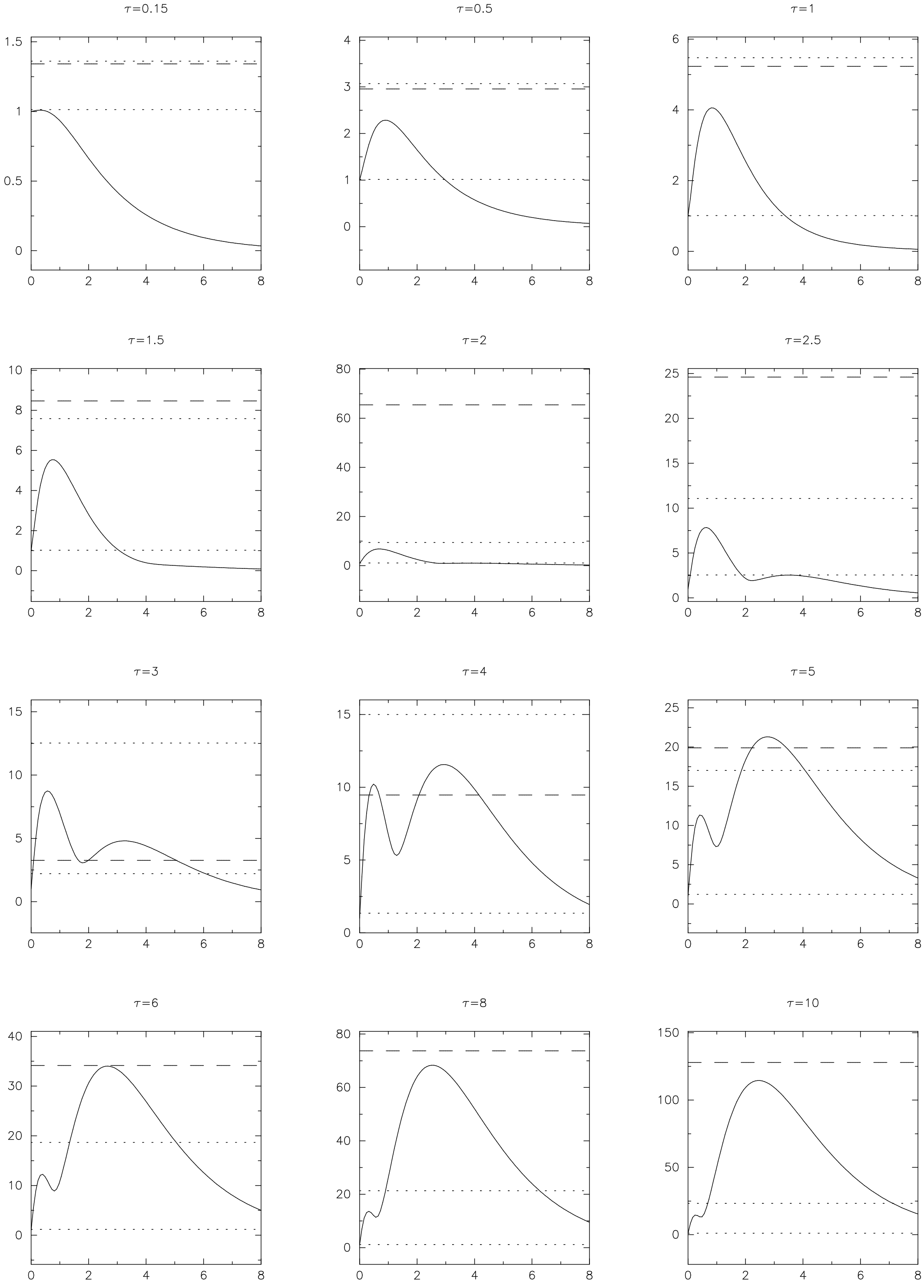

Figure 3.4: SVD approximations for $\sup_{t\geq 0} \left\| e^{A_\tau t} \right\|_2$.

Moreover for each $t_0 \in I$, the one-sided limits and derivatives

$$\lim_{t\searrow t_0} f(t), \lim_{t\nearrow t_0} f(t), \quad \lim_{t\searrow t_0} \dot{f}(t), \lim_{t\nearrow t_0} \dot{f}(t)$$

exist for all functions $f = \sigma_i, u_i, v_i$, $i = 1, \dots, n$.

Proof. In [67, Theorem 4.3.17] *(iii)* it was shown that there exists an analytical pseudo-SVD on any open interval I. Those functions differ from the functions defined in the theorem by relaxing (3.18), the pseudo-singular values only need to satisfy $\tilde{\sigma}_i : I \to \mathbb{R}$ without any restriction on the ordering on the positivity. Enforcing the positivity of σ_i by replacing one of the singular vectors by its negative value, and enforcing the ordering by resorting the indices, we obtain piecewise analytic functions. □

In the following, if we use the term SVD for parameter-dependent functions, we always associate it with a piecewise analytic dyadic decomposition of the form (3.19). This decomposition is not necessarily uniquely determined.

Lemma 3.26. *For $A \in \mathbb{K}^{n\times n}$ let the SVD of $T(t) = e^{At}$, $t > 0$, be given by* (3.19)*. Then we have*

$$\mu_k(t) = \mu_k(t, A) := u_k(t)^*(A^* + A)u_k(t) = v_k(t)^*(A^* + A)v_k(t).$$

Moreover, $\sum_{k=1}^n \mu_k(t) = 2\mathrm{Re}\,\mathrm{trace}\, A$. For $\beta \in \mathbb{R}$, the SVD of $S(t) = e^{(A-\beta I)t}$ is given by $S(t) = \sum_{k=1}^n e^{-\beta t}\sigma_k(t)u_k(t)v_k^*(t)$ and $\mu_k(t, A - \beta I) = \mu_k(t, A) - 2\beta$.

Proof. The vectors u_k, v_k satisfy the equations $T(t)v_k(t) = \sigma_k(t)u_k(t)$, $T(t)^*u_k(t) = \sigma_k(t)v_k(t)$ for all $k = 1, \dots, n$. Moreover, the matrices $T(t)$ and A commute for all $t \geq 0$. Hence, if we suppress the dependence on t we conclude from $T^*AT = T^*TA$ and $T^*A^*T = A^*T^*T$ that

$$\begin{aligned} u_k^*(A + A^*)u_k &= \sigma_k^{-2} v_k^* T^*(A + A^*)Tv_k \\ &= \sigma_k^{-2}(v_k^*T^*TAv_k + v_k^*A^*T^*Tv_k) = v_k^*(A + A^*)v_k. \end{aligned}$$

As the $(u_k)_{k=1,\dots,n}$ form an orthonormal basis of $\mathbb{K}^n$, we obtain for all $t > 0$

$$\begin{aligned} \sum_{k=1}^n \mu_k(t) &= \sum_{k=1}^n u_k(t)^*(A + A^*)u_k(t) = \sum_{k=1}^n (u_k(t)^*Au_k(t) + u_k(t)^*A^*u_k(t)) \\ &= \mathrm{trace}\, A + \mathrm{trace}\, A^* = 2\mathrm{Re}\,\mathrm{trace}\, A. \end{aligned}$$

It is easy to see that the dyadic decomposition of $S(t) = e^{(A-\beta I)t}$ is given by $S(t) = \sum_{k=1}^n (e^{-\beta t}\sigma_k(t))u_k(t)v_k(t)^*$. Therefore the singular vectors are invariant under scalar shifts $A \rightsquigarrow A - \beta I$ and $\mu_k(t, A - \beta I) = u_k(t)^*(A - \beta I + A^* - \beta I)u_k(t) = \mu_k(t, A) - 2\beta u_k(t)^*u_k(t) = \mu_k(t, A) - 2\beta$ shows the behaviour of μ_k under scalar shifts of A. □

By definition, we have that $\lambda_{\min}(A + A^*) \leq \mu_k(t) \leq \lambda_{\max}(A + A^*)$ for all $k = 1, \dots, n$ and all $t > 0$.

Let us now show how $\mu_k(t)$ can be used for the further analysis.

Proposition 3.27. *Given $A \in \mathbb{K}^{n\times n}$. The singular values $\sigma_k(t)$ of $T(t) = e^{At}$ are absolutely continuous and satisfy for almost all $t > 0$ the differential equation*

$$\tfrac{d}{dt}\sigma_k^2(t) = \mu_k(t)\sigma_k^2(t). \tag{3.20}$$

Proof. The functions $\sigma_k(t)$, $k = 1,\dots,n$, are absolutely continuous on $[0,t]$ for all $t > 0$, as they are continuous and piecewise analytic by Theorem 3.25. Again, to save space we drop the dependence on t. Almost everywhere on $\mathbb{R}_+$ the derivative of $\sigma_k^2 = \sigma_k v_k^* T^* u_k = v_k^* T^* T v_k$ is given by

$$\begin{aligned}\dot{\sigma}_k^2 &= \dot{v}_k^* T^* T v_k + v_k^* \dot{T}^* T v_k + v_k^* T^* \dot{T} v_k + v_k^* T^* T \dot{v}_k\\ &= 2\mathrm{Re}\,\sigma_k^2 v_k^* \dot{v}_k + v_k^* T^*(A^* + A) T v_k = v_k^* T^*(A^* + A) T v_k,\end{aligned}$$

since the singular vectors v_k are of unit length, $v_k^* v_k = 1$, so that $v_k^* \dot{v}_k = 0$. Now, σ_k^2 satisfies the differential equation $\frac{d}{dt}\sigma_k^2 = v_k^* T^*(A + A^*) T v_k = \sigma_k^2 u_k^*(A + A^*) u_k = \sigma_k^2 \mu_k$ for almost all $t > 0$. □

Note that the differential equation (3.20) is equivalent to

$$\dot{\sigma}_k(t) := \tfrac{d}{dt}\sigma_k(t) = {}^1\!/\!_2\,\mu_k(t)\sigma_k(t). \tag{3.20'}$$

Proposition 3.28. *If $\lambda_k(A)$ denote the eigenvalues of $A \in \mathbb{K}^{n\times n}$ with real parts decreasingly ordered for $k = 1,\dots,n$ then*

$$\lim_{t\to 0}\mu_k(t) = \lambda_k(A + A^*), \qquad \textit{for all } t_0 \geq 0, \quad \lim_{t\to\infty}\frac{1}{t}\int_{t_0}^{t}\mu_k(\theta)\,d\theta = 2\mathrm{Re}\,\lambda_k(A).$$

Proof. Let us first consider the case $t \to \infty$. The following result on the asymptotic behaviour of singular values of matrix powers is due to Yamamoto, for a proof see [71, Theorem 3.3.21],

$$\lim_{j\to\infty}\sigma_k(B^j)^{1/j} = |\lambda_{\hat{k}}(B)|, \qquad k = 1,\dots,n, \tag{3.21}$$

where $|\lambda_{\hat{1}}| \geq |\lambda_{\hat{2}}| \geq \cdots \geq |\lambda_{\hat{n}}|$ are sorted with respect to the modulus. Setting $B = e^A$ in (3.21) gives us

$$\lim_{j\to\infty}\tfrac{1}{j}\log\sigma_k(e^{Aj}) = \log\left|\lambda_{\hat{k}}(e^A)\right| = \mathrm{Re}\,\lambda_k(A). \tag{3.22}$$

For $t \in \mathbb{R}_+$ with $t = j + \tau$, $j \in \mathbb{N}$, $\tau \in [0,1)$, we obtain $\sigma_k(e^{At}) \leq \sigma_k(e^{Aj})\left\|e^{A\tau}\right\|$ using a Weyl inequality for singular values given in [71, Theorem 3.3.16 (d)]. As $\left\|e^{A\tau}\right\|$ is uniformly bounded for $\tau \in (-1,1)$, $\lim_{j\to\infty}\left\|e^{A\tau}\right\|^{1/j} = 1$. Hence

$$\begin{aligned}\limsup_{t\to\infty}\tfrac{1}{t}\log\sigma_k(e^{At}) &\leq \limsup_{j\to\infty}\tfrac{1}{j}\log\left(\sigma_k(e^{Aj})\sup_{\tau\in[0,1)}\left\|e^{A\tau}\right\|\right)\\ &= \lim_{j\to\infty}\tfrac{1}{j}\log\left(\sigma_k(e^{Aj})\sup_{\tau\in[0,1)}\left\|e^{A\tau}\right\|\right) = \mathrm{Re}\,\lambda_k(A),\end{aligned}$$

Writing $t = j - \tau$, $j \in \mathbb{N}$, $\tau \in [0,1)$ gives $\sigma_k(e^{At}) \geq \sigma_k(e^{Aj}) \left\|e^{A\tau}\right\|^{-1}$, thus

$$\begin{aligned}\liminf_{t\to\infty} \tfrac{1}{t}\log \sigma_k(e^{At}) &\geq \liminf_{j\to\infty} \tfrac{1}{j}\log\left(\sigma_k(e^{Aj}) \inf_{\tau\in[0,1)} \left\|e^{A\tau}\right\|^{-1}\right)\\ &= \lim_{j\to\infty} \tfrac{1}{j}\log\left(\sigma_k(e^{Aj}) \inf_{\tau\in[0,1)} \left\|e^{A\tau}\right\|^{-1}\right) = \operatorname{Re}\lambda_k(A).\end{aligned}$$

Therefore (3.22) is also valid for real t, and $\lim_{t\to\infty} \frac{1}{t}\log\sigma_k(e^{At}) = \operatorname{Re}\lambda_k(A)$, see also [47]. Rewriting (3.20) as an integral equation, we obtain $\sigma_k^2(t) = \sigma_k^2(t_0)e^{\int_{t_0}^t \mu_k(\theta)d\theta}$ for $t \geq t_0 \geq 0$. The asymptotic growth rate of $\sigma_k(t)$ is given by

$$\lim_{t\to\infty} \frac{1}{t}\log\sigma_k(t) = \lim_{t\to\infty} \frac{1}{2t}\log\sigma_k^2(t) = \lim_{t\to\infty} \frac{1}{2t}\int_{t_0}^t \mu_k(\theta)d\theta \tag{3.23}$$

for any $t_0 \in \mathbb{R}_+$ and $t \geq t_0$. By (3.22), equation (3.23) can be rewritten as

$$\lim_{t\to\infty} \frac{1}{t}\int_{t_0}^t \mu_k(\theta)d\theta = 2\lim_{t\to\infty} \frac{1}{t}\log\sigma_k(t) = 2\operatorname{Re}\lambda_k(A).$$

Let us now consider $t = 0$. The function $t \mapsto e^{At}$ is analytic for all $t \in \mathbb{R}$, hence by Theorem 3.25 $u_k(0)$ and $v_k(0)$ are well-defined, moreover we have $\sigma_1(0) = \cdots = \sigma_k(0) = \cdots = \sigma_n(0) = 1$. We show that there exists an eigenvector w_k of unit length corresponding to the kth largest eigenvalue of $A + A^*$, i.e., $(A + A^*)w_k = \lambda_k(A + A^*)w_k$, such that $w_k = u_k(0)$. To see this, note that the differential equation (3.20) is satisfied for the one-sided derivative $\frac{d}{dt^+}\sigma_k^2(t)$ in $t = 0$ and that the vector $u_k(t)$ is by definition an eigenvector of $T(t)T^*(t)$ corresponding to the eigenvalue $\sigma_k^2(t)$ for $t > 0$. For small $t > 0$ we can approximate $\sigma_k^2(t)$ by $1 + \mu_k(0)t + O(t^2)$ and $T(t)$ by $I + At + O(t^2)$. Then

$$\begin{aligned}T(t)T^*(t)u_k(t) &= ((I{+}At)(I{+}A^*t){+}O(t^2))u_k(t) = (I{+}(A{+}A^*)t{+}O(t^2))u_k(t)\\ T(t)T^*(t)u_k(t) &= \sigma_k^2(t)u_k(t) = (1 + \mu_k(0)t + O(t^2))u_k(t).\end{aligned} \tag{3.24}$$

Now, consider $\frac{1}{t}T(t)T^*(t)u_k(t) = \frac{1}{t}\sigma_k^2(t)u_k(t)$. In the limit for $t \searrow 0$ we obtain from (3.24) that the vector $w_k = u_k(0)$ satisfies $(A + A^*)w_k = \mu_k(0)w_k$. Hence, w_k is an eigenvector corresponding to an eigenvalue $\mu_k(0) = \lim_{t\to 0}\mu_k(t)$ of $A{+}A^*$. As $\sigma_k^2(t) = 1{+}\mu_k(0)t{+}O(t^2)$ the $\mu_k(t)$ are decreasingly ordered for small enough $t > 0$ to retain the order of the singular values. Hence $\mu_k(0) = \lambda_k(A + A^*)$ is an eigenvalue of $A + A^*$ with associated eigenvector $u_k(0) = w_k$. The analogous argument for v_k also shows that $\lim_{t\searrow 0} v_k(t) = w_k$. □

If the limits $\lim_{t\to\infty}\mu_k(t)$ exist, then we have $\lim_{t\to\infty}\mu_k(t) = 2\operatorname{Re}\lambda_k(A)$. Numerical experients suggest that this is always true, but a rigorous proof of the existence of these limits is still missing.

For $t > 0$, the vectors $v_k(t)$ and $\sigma_k(t)u_k(t)$ of e^{At} are the initial and final vectors of the trajectory for which the amplification in $[0,t]$ corresponds to the associated singular values $\sigma_k(t)$. Here we have $x(t, v_k(t)) = T(t)v_k(t) = \sigma_k(t)u_k(t)$, and so $\|x(t, v_k(t))\|_2 = \sigma_k(t)\|u_k(t)\|_2 = \sigma_k(t)$.

Especially for the largest singular value, $\sigma_1(t_0) = \left\|e^{At_0}\right\|_2$ and $T(t_0)v_1(t_0) = \sigma_1(t_0)u_1(t_0) = \left\|e^{At_0}\right\| u_1(t_0)$, so that the solutions $x(t, x_0)$ of $\dot{x} = Ax$ satisfy

$$\|x(t_0, v_1(t_0))\|_2 = \left\|e^{At_0}v_1(t_0)\right\|_2 = \sigma_1(t_0)\,\|u_1(t_0)\|_2 = \sigma_1(t_0) = \left\|e^{At_0}\right\|_2. \tag{3.25}$$

Let us take a closer look at the term $\mu_1(t) = u_1(t)^*(A + A^*)u_1(t)$. We get from Proposition 3.28 and Theorem 2.41 that $1/2\mu_1(0)$ equals the initial growth rate of A with respect to the spectral norm. Moreover, we can use it to detect local extrema of $\sigma_1(t)$.

Proposition 3.29. *If $t_0 > 0$ is an isolated local maximizer for $\sigma_1 : t \mapsto \left\|e^{At}\right\|_2$ then $\mu_1(t_0) = 0$ and there is a sign change from $\mu_1(t_0-) > 0$ to $\mu_1(t_0+) < 0$. If $t_0 > 0$ is a local minimizer for σ_1 then $0 \in [\mu_1(t_0-), \mu_1(t_0+)]$. Here $\mu_1(t_0-)$ and $\mu_1(t_0+)$ are the left and right limits of $\mu_1(t)$ in t_0.*

Proof. In a local maximum, the function $\sigma_1(t) = \left\|e^{At}\right\|_2$ is differentiable. To this end, note that for two continuously differentiable functions $f, g : I \to \mathbb{R}$ the function $h = \max\{f, g\}$ is differentiable in local maxima, as $f(t) > g(t)$ and $\dot{f} = 0$ implies $\dot{h} = 0$. If $f(t) = g(t)$ in a local maximum of h, then both f and g attain a local maximum in t, and $\dot{f}(t) = 0 = \dot{g}(t)$. Now $\sigma_1(t)$ is the maximum of n continuously differentiable functions by Theorem 3.25. Its derivative is given by $\dot{\sigma}_1(t) = \frac{1}{2}\sigma_1(t)\mu_1(t)$, as by Proposition 3.27, $\dot{\sigma}_1^2(t) = 2\sigma_1(t)\dot{\sigma}_1(t) = \sigma_1^2(t)\mu_1(t)$. This function is therefore well-defined in local maxima. In particular, for local maxima attained at $t_0 > 0$, $\mu_1(t_0) = 0$ since $\sigma_1(t) > 0$. As a necessary condition for isolated local maxima of σ_1 the sign of $\mu_1(t)$ changes from $+1$ to -1 when passing through $t = t_0$. Local minima of σ_1, however, may not be differentiable. They can only be detected by a sign change of $\mu_1(t)$ from -1 to $+1$ when $t > 0$ passes through a local minimum located at $t = t_0$. □

Example 3.30. We compute $\mu_1(t)$ for the matrix $A = \bigoplus_{k=1}^{8} \frac{-9}{k(k+1)} \left(\begin{smallmatrix} k+1 & 6k+3 \\ 0 & k \end{smallmatrix}\right)$. Figure 3.5 shows the norm of e^{At} and the function $\mu_1(t)$. Here the zeros of μ_1 correspond to local maxima of $\left\|e^{At}\right\|_2$ which are barely noticeable, while minima coincide with jumps of μ_1. In these minima the order of the singular values changes. ■

The brute-force computation of $M_0(A)$, $\alpha(A) < 0$, requires the knowledge of $\left\|e^{At}\right\|_2$ for all t from a sufficiently large interval $[0, T]$. A rough bound for T can be obtained from Corollary 3.9 or from proposition 3.14.

The results obtained in this section show that the singular vectors corresponding to the largest singular value of e^{At} provide enough information to compute a derivative of the singular value function $\sigma_1(t) = \left\|e^{At}\right\|_2$, and hence to implement a Newton method to determine local maxima. Moreover, Proposition 3.28 supplies us with an indicator that the transient phase is over when $\mu_k \approx 2\mathrm{Re}\,\lambda_k$ holds for all $k = 1, \ldots, n$.

Note that the singular vectors of e^{At} corresponding to $\sigma_1(t)$ are necessarily needed for the computation of $\sigma_1(t) = \left\|e^{At}\right\|_2$.

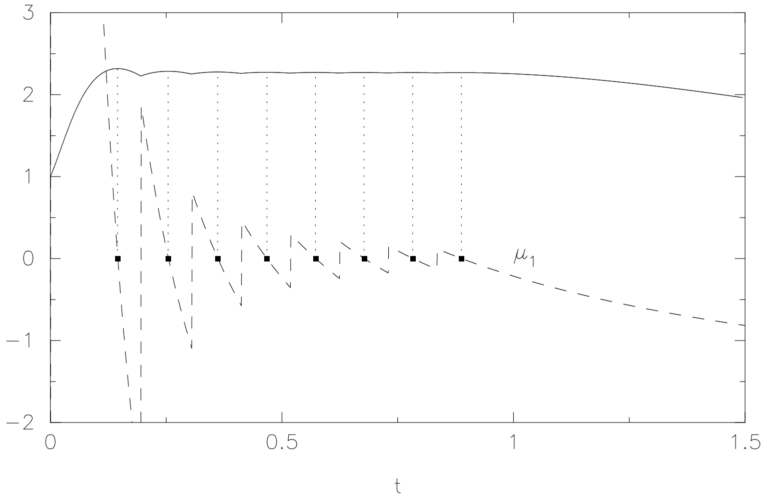

Figure 3.5: $\mu_1(t)$ and local maxima of $\left\|e^{At}\right\|_2$.

3.4 Bounds via Liapunov Functions

In this section we connect our previous discussion of the initial growth rate with some classical results on quadratic Liapunov functions. Let us denote the set of all complex Hermitian matrices by $\mathcal{H}^n(\mathbb{C}) \subset \mathbb{C}^{n\times n}$ and the set of real symmetric matrices by $\mathcal{H}^n(\mathbb{R}) \subset \mathbb{R}^{n\times n}$. Both cases are treated by considering $\mathcal{H}^n = \mathcal{H}^n(\mathbb{K})$, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Suppose that we have found a positive definite Hermitian solution $P \succ 0$ of the Liapunov equation

$$PA + A^*P = -Q \tag{3.26}$$

for given $A \in \mathbb{K}^{n\times n}$ and $Q \succeq 0$. We can use our knowledge of the initial growth rate to derive estimates of $\left\|e^{At}\right\|_2$ based upon (3.26). Let us associate with the matrix A the linear Liapunov operator $\mathcal{L}_A : \mathcal{H}^n \to \mathcal{H}^n, P \mapsto PA + A^*P = -Q$ and its inverse $\mathcal{L}_A^{-1}(-Q) = P$, which always exists when A is exponentially stable. The inner product with weight P, $\langle x, y\rangle_P = y^*Px$, defines the P-norm $\|\cdot\|_P = \sqrt{\langle\cdot,\cdot\rangle_P}$.

Lemma 3.31. *Given $P \succ 0$. If $R \in \mathbb{C}^{n\times n}$ satisfies $P = R^*R$ then the initial growth rate $\mu_P(A)$ corresponding to the elliptical norm $\|\cdot\|_P$ is given by*

$$\begin{aligned}\mu_P(A) = \max_{x\neq 0} \frac{\operatorname{Re}\langle Ax, x\rangle_P}{\langle x, x\rangle_P} &= -\tfrac{1}{2}\min_{x\neq 0}\frac{\langle x, Qx\rangle_2}{\langle x, Px\rangle_2} \\ &= \tfrac{1}{2}\lambda_{\max}\left((RAR^{-1}) + (RAR^{-1})^*\right).\end{aligned} \tag{3.27}$$

where $Q = -\mathcal{L}_A(P)$.

Proof. By Proposition 2.31 we have to determine the dual norm of $\|\cdot\|_P$ and the associated dual vectors of $x \in \mathbb{K}^n$. The dual norm of $\|\cdot\|_P$ is given by $\|\cdot\|_{P^{-1}}$, see (2.19), and a unitary dual pair is uniquely determined by (x, y) where $\|x\|_P = 1$ and $y = Px$. Hence

$\mu_P(A) = \max_{x \neq 0} \frac{\operatorname{Re}\langle Ax, Px\rangle_2}{\langle x, Px\rangle_2} = \max_{x \neq 0} \frac{\operatorname{Re}\langle Ax, x\rangle_P}{\langle x, x\rangle_P}$. Using $PA + A^*P = -Q$ we can write $\operatorname{Re}\langle x, Ax\rangle_P = \frac{1}{2}\langle x, (PA + A^*P)x\rangle_2 = -\frac{1}{2}\langle x, Qx\rangle_2$. If we set $y = Rx$ where R satisfies $R^*R = P$ (e.g. let R be a Cholesky factor or a symmetric square root of A) then

$$-\frac{\langle x, x\rangle_Q}{\langle x, x\rangle_P} = \frac{\langle (R^{-1})^*(PA + A^*P)R^{-1}y, y\rangle_2}{\langle y, y\rangle_2} = \frac{\langle (RAR^{-1} + (R^{-1})^*A^*R^*)y, y\rangle_2}{\langle y, y\rangle_2}.$$

By the Rayleigh-Ritz Theorem [70], maximizing the last quotient over all $y \neq 0$ gives the largest eigenvalue of $RAR^{-1} + (RAR^{-1})^*$ □

From the quotient (3.27) we obtain an estimate for the initial growth rate in the following situation.

Corollary 3.32. *Given an exponentially stable matrix $A \in \mathbb{C}^{n\times n}$, a positive definite matrix $P \succ 0$ and $\beta \in \mathbb{R}$ such that $PA + A^*P \preceq 2\beta P$. Then $\mu_P(A) \leq \beta$.*

The quotient $\min \frac{\langle x, Qx\rangle_2}{\langle x, Px\rangle_2}$ can also be interpreted as the generalized eigenvalue of a Hermitian matrix pencil, see [44, Chapter X].

Proposition 3.33 ([44, Theorem X.22])**.** *Given a Hermitian matrix pencil $(Q, P) \in \mathcal{H}^n \times \mathcal{H}^n$ with $P \succ 0$. Then the pencil is regular, i.e., $\det(Q - \lambda P) \not\equiv 0$ and its characteristic equation $\det(Q - \lambda P) = 0$ always has n real roots $\lambda_1, \ldots, \lambda_n$, counting multiplicities. Moreover, there exist $Z \in \mathbb{K}^{n\times n}$ and $\Lambda = \operatorname{diag}(\lambda_i) \in \mathbb{R}^{n\times n}$ such that $QZ = PZ\Lambda$ and $Z^*PZ = I_n$.*

We call $\sigma(Q, P) := \{\lambda \in \mathbb{C} \mid \det(Q - \lambda P) = 0\}$ the *spectrum* of the Hermitian pencil (Q, P), its elements are called generalized eigenvalues of (Q, P). For these pencils, a counterpart of the Rayleigh-Ritz Theorem holds true.

Proposition 3.34 ([44, Theorem X.13])**.** *For a Hermitian matrix pencil $(Q, P) \in \mathcal{H}^n \times \mathcal{H}^n$ with $P \succ 0$, the largest and smallest generalized eigenvalues are given by*

$$\lambda_{\max}(Q, P) = \max_{x \neq 0} \frac{x^*Qx}{x^*Px}, \qquad \lambda_{\min}(Q, P) = \min_{x \neq 0} \frac{x^*Qx}{x^*Px}. \tag{3.28}$$

Hence the initial growth rate $\mu_P(A)$ associated with the positive definite matrix $P \in \mathcal{H}^n$ is given by the maximal generalized eigenvalue of the matrix pencil $(-(PA + A^*P), P)$. As this pencil is regular, we can rewrite the spectrum of the pencil as the spectrum of a matrix, $\sigma(-(PA + A^*P), P) = \sigma(-(A + P^{-1}A^*P)) = \sigma(-(PAP^{-1} + A^*))$. The matrix $A + P^{-1}A^*P$ is not Hermitian any more. From these remarks about matrix pencils we extract yet another way of computing the initial growth rate with respect to P, namely,

$$\mu_P(A) = \tfrac{1}{2}\lambda_{\max}(-(A + P^{-1}A^*P)) = -\tfrac{1}{2}\lambda_{\min}(A + P^{-1}AP).$$

By properties of the initial growth rate, $\mu_P(A) \leq 0$ implies that $(e^{At})_{t\in\mathbb{R}_+}$ is a contraction semigroup in the P-norm since $\left\|e^{At}x\right\|_P \leq e^{\mu_P(A)t}$. For an estimate with respect to the spectral norm we have to compute the eccentricity of $\|\cdot\|_P$.

Theorem 3.35. *Let $A \in \mathbb{K}^{n\times n}$ and $Q \in \mathcal{H}^n$. Suppose that there exists $P \succ 0$ which solves* (3.26)*. Then*

$$\left\|e^{At}\right\|_2 \le \sqrt{\kappa_2(P)}e^{\mu_P(A)t}, \qquad t \ge 0.$$

Here $\kappa_2(P)$ denotes the condition number of P defined by $\kappa_2(P) = \|P\|_2 \left\|P^{-1}\right\|_2$.

Note that we do not assume that A is stable. Hence Q is not necessarily positive semidefinite, thus $\mu_P(A)$ may also be positive.

Proof. In order to apply Corollary 2.57, we only have to show that the eccentricity of $\|\cdot\|_P$ when compared with $\|\cdot\|_2$ is given by $\kappa_2(P)^{1/2}$. This follows from

$$\lambda_{\min}(P)\langle x,x\rangle_2 \le \langle x,x\rangle_P \le \lambda_{\max}(P)\langle x,x\rangle_2, \qquad x \in \mathbb{C}^n, x \ne 0, \tag{3.29}$$

where $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$ denote the minimal and maximal eigenvalue of P, respectively. However, for the eigenvectors corresponding to the maximal and minimal eigenvalues of P, equality holds in either of the two inequalities of (3.29). Hence, $\operatorname{ecc}\|\cdot\|_P = \frac{\lambda_{\max}(P)^{1/2}}{\lambda_{\min}(P)^{1/2}}$. The statement of the theorem then follows from Corollary 2.57. □

The following definition determines the set of matrices which satisfy Theorem 3.35.

Definition 3.36. A matrix $A \in \mathbb{K}^{n\times n}$ is called *quadratically* (M,β)*-stable* if there exists a positive definite $P \in \mathcal{H}^n$ with $\kappa(P)^{1/2} \le M$ and $\mu_P(A) \le \beta$.

If the norm $\|\cdot\|$ under consideration is the spectral norm, we can interpret (3.5) as a special case of Theorem 3.35.

Corollary 3.37. *Suppose that $A \in \mathbb{K}^{n\times n}$ is diagonalizable with an invertible matrix $V \in \mathbb{C}^{n\times n}$ of left eigenvectors satisfying $V^*A = \Lambda V^*$, $\Lambda = \operatorname{diag}(\lambda_i)$, $\lambda_i \in \sigma(A)$. Then $\left\|e^{At}\right\|_2 \le \kappa_2(V)e^{\alpha(A)t}$, $t \ge 0$.*

Proof. Setting $P = VV^*$ gives $PA + A^*P = V(\Lambda + \bar{\Lambda})V^* = -Q$. Hence for $y = V^*x$

$$\mu_P(A) = \tfrac{1}{2}\max_{x\ne 0}\frac{\langle x, -Qx\rangle_2}{\langle x, Px\rangle_2} = \tfrac{1}{2}\max_{y\ne 0}\frac{\left\langle y, (\Lambda+\bar{\Lambda})y\right\rangle_2}{\langle y, y\rangle_2} = \tfrac{1}{2}\lambda_{\max}(\Lambda+\bar{\Lambda}) = \alpha(A).$$

Moreover, the square root of the condition number of P is given by $\sqrt{\kappa_2(P)} = \kappa_2(V) = \|V\|_2 \left\|V^{-1}\right\|_2$. The corollary now follows from Theorem 3.35. □

Example 3.38. Consider the matrix $A = \left(\begin{smallmatrix}-5 & 36\\ 0 & -20\end{smallmatrix}\right)$ which we already studied in Example 3.15. Figure 3.6 shows an ellipse which is invariant under the flow of $\dot{x} = Ax$. The associated quadratic form is induced by the Hermitian matrix $P = \left(\begin{smallmatrix}125 & 40\\ 40 & 317\end{smallmatrix}\right) \succ 0$, hence the transient growth is bounded by $\kappa(P)^{1/2} = \frac{5}{3}$. Here the initial growth rate $\mu_P(A)$ equals 0 as there exist trajectories which enter the ellipse tangentially, and therefore $Q = -(PA + A^*P) = 50\left(\begin{smallmatrix}25 & -70\\ -70 & 196\end{smallmatrix}\right)$ is only semidefinite. ■

If both $P \in \mathcal{H}^n$ and $Q \in \mathcal{H}^n$ are positive definite and related via a Liapunov equation $\mathcal{L}_A(P) = -Q$ then we can compare the initial growth rates induced by the elliptical norms associated with P and Q, respectively.

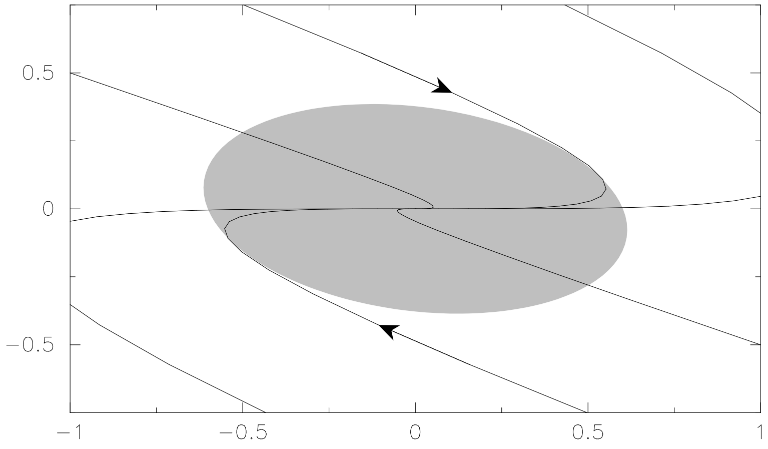

Figure 3.6: Flow-invariant ellipse.

Theorem 3.39. *Suppose that $A \in \mathbb{K}^{n\times n}$ is an exponentially stable matrix and $P \succ 0$ and $Q \succ 0$ solve $\mathfrak{L}_A(P) = -Q$. Then we have*

$$-\mu_Q(-A) \le -\mu_P(-A) \le \mu_P(A) \le \mu_Q(A).$$

Proof. Let us first consider the inequality $\mu_P(A) \le \mu_Q(A)$. Theorem 3.35 implies for the inner product with weight Q that

$$\left\langle e^{At}x, e^{At}x\right\rangle_Q \le e^{2\mu_Q(A)t}\langle x, x\rangle_Q, \tag{3.30}$$

where $\mu_Q(A)$ may also be positive. As both P and Q are positive definite, Lemma 3.31 shows that $\mu_P(A) = -\frac{1}{2}\min_{x\neq 0}\frac{\langle x,Qx\rangle}{\langle x,Px\rangle} < 0$ always holds. If therefore $\mu_Q(A) \ge 0$ then $\mu_Q(A) > \mu_P(A)$ is trivially satisfied. Let us therefore assume that $\mu_Q(A) < 0$. Note that

$$-\tfrac{d}{dt}\left\langle e^{At}x, e^{At}x\right\rangle_P = -2\mathrm{Re}\left\langle e^{At}x, Ae^{At}x\right\rangle_P = \left\langle e^{At}x, e^{At}x\right\rangle_Q, \qquad x \in \mathbb{K}^n, t \ge 0.$$

By using this equality in the integration of (3.30) we obtain

$$\left\langle e^{At}x, e^{At}x\right\rangle_P = \int_t^\infty \left\langle e^{As}x, e^{As}x\right\rangle_Q ds \le \int_t^\infty e^{2\mu_Q(A)s} ds \langle x, x\rangle_Q = -\frac{1}{2\mu_Q(A)} e^{2\mu_Q(A)t}\langle x, x\rangle_Q.$$

This integral is well-defined as $\mu_Q(A) < 0$. Hence for $t = 0$ and all $x \neq 0$

$$\langle x, x\rangle_P \le -\frac{1}{2\mu_Q(A)}\langle x, x\rangle_Q \iff \mu_Q(A) \ge -\tfrac{1}{2}\min_{x\neq 0}\frac{\langle x, x\rangle_Q}{\langle x, x\rangle_P} = \mu_P(A).$$

The lower bound follows analogously by considering

$$\left\langle e^{-At}x, e^{-At}x\right\rangle_Q \le e^{2\mu_Q(-A)t}\langle x, x\rangle_Q \iff \left\langle e^{At}x, e^{At}x\right\rangle_Q \ge e^{-2\mu_Q(-A)t}\langle x, x\rangle_Q. \tag{3.31}$$

If $\mu_Q(-A) > 0$ then an integration of (3.31) provides us with $\langle x, x\rangle_P \geq \frac{1}{2\mu_Q(-A)}\langle x, x\rangle_Q$. Hence $-\mu_Q(-A) \leq -\frac{1}{2}\frac{\langle x,x\rangle_Q}{\langle x,x\rangle_P}$. Taking the minimum of this quotient over all $x \neq 0$, we have

$$-\mu_Q(-A) \leq \tfrac{1}{2}\min_{x\neq 0}\frac{-\langle x, x\rangle_Q}{\langle x, x\rangle_P} = -\mu_P(-A) < 0.$$

The case $\mu_Q(-A) \leq 0$ is again trivial, as $-\mu_P(-A) = -\frac{1}{2}\max_{x\neq 0}\frac{\langle x,Q,x\rangle}{\langle x,Px\rangle} < 0$. The inequality $-\mu_P(-A) \leq \mu_P(A)$ is found in Proposition 2.40 (i). □

Let us now study the effect of using Theorem 3.39 iteratively.

Theorem 3.40. *Let $A \in \mathbb{K}^{n\times n}$ be exponentially stable. Consider the Hermitian matrix sequence $(P_i)_{i\in\mathbb{N}} \subset \mathcal{H}^n$ of Liapunov solutions $P_iA + A^*P_i = -P_{i-1}/\|P_{i-1}\|$. Then for a generically chosen initial value $P_0 \succ 0$*

$$\lim_{i\to\infty}\mu_{P_i}(A) = \alpha(A), \qquad \lim_{i\to\infty} -\mu_{P_i}(-A) = -\alpha(-A).$$

Proof. The construction of the matrix sequence (P_i) corresponds to an inverse power method without shifts applied to the linear operator $-\mathcal{L}_A$, see Wilkinson [148] and Stewart [132] for a general discussion. This method converges to some subspace spanned by eigenvectors which are associated with eigenvalues of A that minimize the distance to the origin. If such an eigenvalue λ_* which located nearest to the origin is uniquely determined, i.e. $\{\lambda_*\} = \{\lambda \in \sigma(A) \mid |\lambda| = \min_{\lambda'\in\sigma(A)}|\lambda'|\}$ then the convergent subspace is of dimension 1. Hence the inverse power method converges to an eigenvector P_* of $-\mathcal{L}_A$ corresponding to the eigenvalue λ_*. If λ_* is of higher geometric multiplicity then the convergent eigenvalue depends on the choice of the initial value P_0. Now, $P_0 \succ 0$ is positive definite and as $-\mathcal{L}_A^{-1} : \mathcal{H}_+^n \to \mathcal{H}_+^n$ retains the positive-definiteness, all $P_i \succ 0$. Thus if the limit $P_* = \lim_{i\to\infty} P_i$ exists it is a Hermitian matrix. But Hermitian eigenvectors $P \in \mathcal{H}^n$ of $\mathcal{L}(A)$ are associated with real eigenvalues $\lambda \in \mathbb{R}$ as $\lambda P = PA + A^*P = (PA + A^*P)^* = \bar{\lambda}P$. The spectrum of the Liapunov operator $\mathcal{L}_A$ as an operator on $\mathbb{K}^{n\times n}$ is given by $\sigma(\mathcal{L}_A) = \{\lambda_1 + \bar{\lambda}_2 \mid \lambda_1, \lambda_2 \in \sigma(A)\}$, see [90, Theorem 12.2.1]. As A is exponentially stable, $\min\operatorname{dist}(-\sigma(\mathcal{L}_A), 0)$ is attained for $\lambda_* = -2\alpha(A)$. Therefore the inverse power method converges. Let us now study the limit of the spectra $\frac{1}{2}\sigma(P_iA + A^*P_i)$ as $i \to \infty$. Let us assume that A is given in (complex) Schur form where the real parts of the eigenvalues are increasingly ordered along the diagonal. If R_i is a Cholesky decomposition of the positive definite Hermitian matrix $P_i = R_i^*R_i$ then R is an upper triangular matrix, and hence the product

$$R_iAR_i^{-1}, \qquad i \in \mathbb{N}, \tag{3.32}$$

is upper triangular, too. By construction its diagonal coincides with the diagonal of A. Now P_i converges to an eigenvector P_* associated with the eigenvalue $2\alpha(A)$ of $\mathcal{L}_A$. As the diagonal of (3.32) is constant, it must converge to the diagonal matrix of A, $R_iAR_i^{-1} \to \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Hence $\lim_{i\to\infty}\frac{1}{2}\sigma(P_iA + A^*P_i) = \{\operatorname{Re}\lambda \mid \lambda \in \sigma(A)\}$ and especially $\lim_{i\to\infty}\mu_{P_i}(A) = \alpha(A)$, $\lim_{i\to\infty} -\mu_{P_i}(-A) = -\alpha(-A)$. □

Unfortunately, if the optimal eigenvalue $\lambda_* = -2\alpha(A)$ is of simple multiplicity then P_* is of rank one, and $\kappa(P_i) \to \infty$ as $i \to \infty$ which is unacceptable. Instead, let us now try to optimize the condition number. We pose the following problem.

Problem 3.41. *For a given stable matrix $A \in \mathbb{K}^{n\times n}$ find a positive definite solution $P \in \mathcal{H}^n$ of the Liapunov inequality $PA + A^*P \preceq 0$ with minimal condition number,*

$$\kappa^* = \inf\{\kappa(P) \,|\, \mathcal{L}_A(P) \preceq 0\}.$$

As the condition number only fixes the ratio between the largest and smallest eigenvalue of P we cannot expect uniqueness (modulo scalar multiples) for dimensions $n \geq 3$.
The problem of finding a quadratic Liapunov norm with minimal eccentricity may be recast as a semidefinite program with linear matrix inequality constraints. This formulation can be readily used with available numerical solvers.

Problem 3.42. *For a given matrix $A \in \mathbb{K}^{n\times n}$ find a solution $(\kappa, P) \in \mathbb{R}_+ \times \mathcal{H}^n$ of the following semidefinite program*

$$\textit{Minimize } \kappa \geq 1 \textit{ under } I_n \preceq P \preceq \kappa I_n,\ P = P^*,\ PA + A^*P \preceq 0.$$

The solution set will be empty if A is not stable as the Liapunov inequality is never satisfied for positive definite $P \in \mathcal{H}^n$. Unfortunately, the numerical treatment of Problem 3.42 runs into difficulties even for moderate matrix dimensions. We use the following proposition to show that for the optimal solution pair (P', Q') of (3.26) P' is positive definite and Q' is only semidefinite which causes numerical problems.

Proposition 3.43. *Suppose that $A \in \mathbb{K}^{n\times n}$ is stable and that the Hermitian pairs (P_1, Q_1), (P_2, Q_2) satisfy*

$$P_1A + A^*P_1 = -Q_1, \qquad P_2A + A^*P_2 = -Q_2,$$

with $P_1 \succ 0$, $P_2 \succ 0$, $\kappa(P_2) < \kappa(P_1)$, $Q_1 \succeq 0, Q_1 + Q_2 \succeq 0$. Then $\kappa(P_1 + P_2) < \kappa(P_1)$.

Proof. Under the conditions of the proposition, both (P_1, Q_1) and $(P_1 + P_2, Q_1 + Q_2)$ are pairs of a positive definite matrix and a semidefinite matrix that satisfy the Liapunov equation (3.26). We therefore have to show that $\kappa(P_1 + P_2) \leq \kappa(P_1)$. As P_1 and P_2 are both positive definite Hermitian matrices we have

$$\kappa(P_1 + P_2) = \frac{\lambda_{\max}(P_1 + P_2)}{\lambda_{\min}(P_1 + P_2)} \leq \frac{\lambda_{\max}(P_1) + \lambda_{\max}(P_2)}{\lambda_{\min}(P_1) + \lambda_{\min}(P_2)}. \tag{3.33}$$

Now, from $\kappa(P_2) < \kappa(P_1)$ we obtain

$$\kappa(P_2) = \frac{\lambda_{\max}(P_2)}{\lambda_{\min}(P_2)} < \frac{\lambda_{\max}(P_1) + \lambda_{\max}(P_2)}{\lambda_{\min}(P_1) + \lambda_{\min}(P_2)} < \frac{\lambda_{\max}(P_1)}{\lambda_{\min}(P_1)} = \kappa(P_1). \tag{3.34}$$

This yields $\kappa(P_1 + P_2) < \kappa(P_1)$. Therefore $P' = P_1 + P_2$ yields a smaller condition number than P_1. □

In [82] it was noted that the choice $P_2 = \lambda I_n$ with an appropriate scaling factor λ leads to a condition number $\kappa(P_1+P_2)$ which is less or equal to the condition number $\kappa(P_1)$. Suppose that $A + A^* \not\prec 0$ and $Q_1 \succ 0$, then for $P_2 = \lambda I_n, \lambda > 0$, we have $Q_2 = -\mathcal{L}_A(P_2) = -\lambda(A + A^*)$. Since $\kappa(P_2) = 1$ and $\kappa(P_1) > 1$ Proposition 3.43 yields that the condition number estimate of the sum $P_1 + P_2$ is always improved provided $Q_1 + Q_2 = Q_1 - \lambda(A + A^*) \succeq 0$. Hence one should choose λ to be the smallest positive generalized eigenvalue of the matrix pencil $(Q_1, A + A^*)$. With this choice, $Q' = Q_1 - \lambda(A + A^*)$ is singular.
Let us generalize this procedure. We assume that (P_1, Q_1) is a pair of positive definite Hermitian matrices which satisfy the Lyapunov equation (3.26). If $Q' \in \mathcal{H}^n$ is some search direction then we have to determine $\lambda' \in \mathbb{R}$ such that the conditions of Proposition 3.43 hold for $Q_2 = \lambda' Q'$. We obtain the following update step.

Lemma 3.44. *Suppose that $A \in \mathbb{K}^{n\times n}$ is stable. Given Hermitian matrices Q_1 and Q_2 which satisfy the conditions of Proposition 3.43, we set $\tilde{\lambda}$ to the smallest nonnegative generalized eigenvalue of the matrix pencil $(Q_1, -(Q_1+Q_2))$, $\tilde{\lambda} = \min(\sigma\,(Q_1, -(Q_1 + Q_2))\cap\mathbb{R}_+)$. Then the positive definite Hermitian matrix*

$$\tilde{P} = P_1 + \tilde{\lambda} P_2$$

satisfies $\kappa(\tilde{P}) \leq \kappa(P_1)$, and yields a positive semidefinite $\tilde{Q} = -\mathcal{L}_A(\tilde{P}) \in \mathcal{H}^n$.

However, it is unclear how to determine a feasible search direction. Moreover, a different strategy should be used if Q_1 is singular. Concluding from this lemma an optimal solution of Problem (3.42) is attained for a singular Q. This may be one of the reasons for the bad performance of the numerical solvers.
Furthermore, there is a gap between the exponential estimates obtained from quadratic Liapunov functions and the transient amplification $M_0(A) = \sup_{t\geq 0} \left\|e^{At}\right\|$ as the following example shows.

Example 3.45. For a given $k \in \mathbb{N}$ consider the matrix $A = \left(\begin{smallmatrix} -1 & 50k \\ 0 & -51 \end{smallmatrix}\right)$. The spectral norm of the matrix exponential for a real 2×2 matrix in upper triangular form $A = \left(\begin{smallmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{smallmatrix}\right)$, $\lambda_1 \neq \lambda_2$, is given by

$$\left\|e^{At}\right\| = \tfrac{1}{2}\left|e^{\lambda_1 t} - e^{\lambda_2 t}\right| \left(\sqrt{\coth(\tfrac{1}{2}(\lambda_1 - \lambda_2)t)^2 + (\tfrac{\alpha}{\lambda_1-\lambda_2})^2} + \sqrt{1 + (\tfrac{\alpha}{\lambda_1-\lambda_2})^2}\right),$$

see Proposition 4.4. For $\beta = -1$ we get the monotonously increasing function

$$\left\|e^{(A-\beta I)t}\right\| = \tfrac{1}{2}(1 - e^{-50t}) \left(\sqrt{\coth(25t)^2 + k^2} + \sqrt{1 + k^2}\right) \xrightarrow{t\to\infty} \sqrt{1+k^2}$$

as $\lim_{x\to\infty} \coth(x) = 1$. Hence, $M = \sqrt{1+k^2}$ is the smallest possible bound for strict (M,β)-stability with $\beta = -1$. Now let us examine which bound can be obtained using Theorem 3.35. The strict Liapunov inequality $PA + A^*P + 2P \prec 0$ is unsolvable, but there exist matrices $P \succ 0$ which solve $PA + A^*P \preceq -2P$. The matrix $P = \left(\begin{smallmatrix} p_1 & p_3 \\ p_3 & p_2 \end{smallmatrix}\right)$ is a solution of this inequality if and only if

$$kp_1 - p_3 = 0, \qquad kp_3 - p_2 < 0.$$

If we fix $p_1 = 1$ then necessarily $p_3 = k$ and $p_2 > k^2$. With this choice, P is positive definite. Other solutions are positive scalar multiples of solutions representable in such a manner. Let us now compute the condition number of P. The condition number for a 2×2 real positive definite symmetric matrix P is given by

$$\kappa(P) = \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} = \frac{\operatorname{trace} P}{2 \det P}\left(\operatorname{trace} P + \sqrt{(\operatorname{trace} P)^2 - 4 \det P}\right) - 1, \tag{3.35}$$

which can be obtained by expressing $\lambda_{\max}(P)$ and $\lambda_{\min}(P)$ in terms of $\operatorname{trace}(P) = \lambda_{\max}(P) + \lambda_{\min}(P)$ and $\det(P) = \lambda_{\max}(P)\lambda_{\min}(P)$. By writing $p_2 = k^2 + \alpha$ we get

$$\begin{aligned} \kappa(\alpha) &= \frac{1 + k^2 + \alpha}{2\alpha}\left((1 + k^2 + \alpha) + \sqrt{(1 + k^2 + \alpha)^2 - 4\alpha}\right) - 1 \\ &= \frac{1 + k^2 + \alpha}{2\alpha}\left((1 + k^2 + \alpha) + \sqrt{(\alpha + (k^2 - 1))^2 + 4k^2}\right) - 1, \end{aligned}$$

which attains its minimum of $k^2 + 2k\sqrt{1 + k^2} + (1 + k^2)$ at $\tilde{\alpha} = 1 + k^2$. Therefore the best bound obtainable by Theorem 3.35 is $\sqrt{\kappa(\tilde{\alpha})} = k + \sqrt{1 + k^2}$. ■

In this example there is a gap of k between this Liapunov bound and the minimal bound M. More interestingly, the quotient of both bounds approaches 2. It is an open question if in general this "quadratic Liapunov performance" quotient is bounded by the dimension of the space,

$$\sup_{A \in \mathbb{K}^{n \times n} \text{ stable}} \left(\inf\left\{\sqrt{\kappa(P)} \,\middle|\, P \succ 0, PA + A^*P \preceq 0\right\}\right)\left(\sup_{t \geq 0}\left\|e^{At}\right\|\right)^{-1} \overset{?}{=} n.$$

3.5 Bounds from the Resolvent

The *resolvent* of $A \in \mathbb{C}^{n \times n}$ is given by $R(s, A) = (sI_n - A)^{-1}$. It may be used for an alternative definition of the matrix exponential via

$$e^{At} = \lim_{k \to \infty}\left(I - \frac{At}{k}\right)^{-k} = \lim_{k \to \infty}\left(\tfrac{k}{t} R\left(\tfrac{k}{t}, A\right)\right)^k, \ t > 0, \tag{3.36}$$

that is, e^{At} is defined as the limiting product of implicit Euler steps. This limit is defined for k large enough such that $I - A\frac{t}{k}$ is invertible, which is guaranteed for $k > t\rho(A)$. Let us recall the characterization (2.33*b*) of Corollary 2.52. We rephrase it in the following proposition.

Proposition 3.46. *Suppose that $A \in \mathbb{K}^{n \times n}$ then for each fixed $k \in \mathbb{N}^*$,*

$$\mu(A) = \tfrac{d}{dt^+}\left\|(I - A\tfrac{t}{k})^{-k}\right\|\Big|_{t=0}.$$

For an example, see Figure 2.1 where different resolvent approximations of e^{At} indeed have the same initial growth rate. Note also that the resolvent of A and the matrix exponential of A are connected via a Laplace transformation, see Corollary 2.9. We can rewrite the inverse Laplace transformation in form of a Cauchy integral formula for operators,

$$e^{At} = \tfrac{1}{2\pi i} \int_{\Gamma} e^{st} R(s, A) ds,$$

where Γ is any positively oriented, piecewise smooth simple closed curve encircling the spectrum of A.

Now consider a full block perturbation structure $\mathbf{\Delta} = (\mathbb{C}^{n\times n}, \|\cdot\|)$, cf. Section 1.3. If the operator norm $\|\cdot\|$ is induced from a semi-algebraic vector norm (for example, a p-norm with rational p) then the boundary of the ε-pseudospectrum for $\varepsilon > 0$ is piecewise analytic, see Karow [76, Corollary 3.2.2]. Hence the contour Γ of an ε-pseudospectrum is rectifiable and defines a piecewise smooth simple curve encircling the spectrum of A. This contour may contain several connected components, but this causes no problem for the following result. Namely, using this contour we obtain for all $\varepsilon > 0$

$$\left\|e^{At}\right\| \leq \frac{1}{2\pi} \int_{\partial\sigma_\varepsilon(A\,|\,\mathbf{\Delta})} e^{\mathrm{Re}\,st} \left\|R(s, A)\right\| ds = \frac{1}{2\pi\varepsilon} \int_{\partial\sigma_\varepsilon(A\,|\,\mathbf{\Delta})} e^{\mathrm{Re}\,st} ds, \qquad t \geq 0.$$

If the length of the contour is known this provides the basis of further estimates, see Embree and Trefethen [37]. We now shed some light on the theorems of Hille-Yosida and Kreiss-Spijker.

3.5.1 Kreiss Matrix and Hille-Yosida Generation Theorems

The Hille-Yosida-Theorem links the (M, β)-stability of A to properties of the resolvent $R(s, A)$, see [38] and the discussion in Chapter 2. One may be interested in the transient amplification only in certain directions of the state space. Hence we use structure matrices to take this into account. We now present a structured version of the Hille-Yosida Theorem for the matrix case.

Definition 3.47. Suppose that the structure matrices $B \in \mathbb{C}^{n\times \ell}$ and $C \in \mathbb{C}^{q\times n}$ are given. A matrix $A \in \mathbb{C}^{n\times n}$ is said to be *structured* (M, β)*-stable* if $\beta > \alpha(A)$ and

$$\left\|Ce^{At}B\right\| \leq Me^{\beta t}, \qquad t \geq 0.$$

The *structured transient bound* is given by

$$M_\beta(A, B, C) = \sup_{t\geq 0} \left\|Ce^{(A-\beta I)t}B\right\|.$$

Note that always $M_\beta(A, B, C) \geq \|CB\|$.

Theorem 3.48 (Hille-Yosida, structured version)**.** *The matrix $A \in \mathbb{C}^{n\times n}$ is structured (M,β)-stable for the given structure matrices $B \in \mathbb{C}^{n\times \ell}$ and $C \in \mathbb{C}^{q\times n}$ if and only if for all $k \in \mathbb{N}$. and for* $\operatorname{Re} s > \beta$

$$\left\|CR(s,A)^k B\right\| = \left\|C(sI-A)^{-k}B\right\| \leq \frac{M}{(\operatorname{Re} s-\beta)^k}. \tag{3.37}$$

Proof. Let A be structured (M,β)-stable. Then for $t=0$ we have $\|CB\| \leq M$ hence (3.37) holds for $k=0$. Now let $k \in \mathbb{N}^*$ be arbitrary. Using the Laplace transformation we have for all $\operatorname{Re} s > \beta > \alpha(A)$ that

$$(sI-A)^{-k} = \frac{1}{(k-1)!}\int_0^\infty t^{k-1}e^{(A-sI)t}dt.$$

Therefore we obtain for all $y \in \mathbb{C}^\ell$ that

$$\begin{aligned}\left\|C(sI-A)^{-k}By\right\| &\leq \frac{1}{(k-1)!}\int_0^\infty t^{k-1}e^{-\operatorname{Re} st}\left\|Ce^{At}By\right\|dt \\ &\leq \frac{M}{(k-1)!}\int_0^\infty t^{k-1}e^{(\beta-\operatorname{Re} s)t}\left\|y\right\|dt = \frac{M}{(\operatorname{Re} s-\beta)^k}\left\|y\right\|,\end{aligned}$$

where $\frac{1}{(k-1)!}\int_0^\infty t^{k-1}e^{-\gamma t}dt = \frac{1}{\gamma^k}$ follows from repeated partial integration. Hence (3.37) holds. Conversely, if (3.37) holds for all $k \in \mathbb{N}$ then we use the representation (3.36). To this end, fix $t > 0$ and set $s = \frac{k}{t} + \beta$. Then $\frac{sI-A}{s-\beta} = I - (A-\beta I)\frac{t}{k}$, and (3.37) now gives for all $k \in \mathbb{N}$

$$M \geq \left\|C\left(\frac{sI-A}{s-\beta}\right)^{-k}B\right\| = \left\|C\left(I-(A-\beta I)\frac{t}{k}\right)^{-k}B\right\| \xrightarrow[k\to\infty]{} \left\|Ce^{(A-\beta I)t}B\right\|.$$

Therefore A is structured (M,β)-stable. □

For a formulation which also works for operators in Banach spaces, see Theorem 2.6. The main issue in the proof of the operator-theoretic version is to establish (3.36).

In order to use Theorem 3.48 for the test whether a matrix is structured (M,β)-stable all powers of the resolvent have to be checked, so that this test is of little practical use. In contrast, the Kreiss-Spijker Theorem does not require higher powers of the resolvent to be known. We also present a structured version.

Theorem 3.49 (Kreiss-Spijker, structured version)**.** *Suppose $A \in \mathbb{C}^{n\times n}$ is a stable matrix and $B \in \mathbb{C}^{n\times\ell}$ and $C \in \mathbb{C}^{q\times n}$ are given structure matrices. Define the* Kreiss constant $k(A,B,C) = \sup_{\operatorname{Re} s>0}(\operatorname{Re} s)\left\|C(sI-A)^{-1}B\right\|$. *Then*

$$k(A,B,C) \leq M_0(A,B,C) \leq (en)k(A,B,C), \tag{3.38}$$

where $e = \exp(1) = 2.718\ldots$

The main part of this proof is Spijker's Lemma. We present here a version due to Aptekarev [3]. We start with some remarks about the *Riemannian sphere*, defined by $\mathbb{S}^2 = \{(x_1, x_2, x_3)^\top \in \mathbb{R}^3 \,|\, x_1^2 + x_2^2 + x_3^2 = 1\}$. The north pole of this sphere is $N = (0, 0, 1)^\top$ and the south pole is $S = (0, 0, -1)^\top$. For each $x = (x_1, x_2, x_3)^\top \in \mathbb{S}^2$ there exists a rotation $G \in SO_3$ such that $Gx = S$, i.e., the axis given by $(x, -x)$ becomes vertical. To see this, consider

$$G = \begin{pmatrix} 1 - \frac{x_1^2}{1-x_3} & -\frac{x_1 x_2}{1-x_3} & x_1 \\ -\frac{x_1 x_2}{1-x_3} & 1 - \frac{x_2^2}{1-x_3} & x_2 \\ -x_1 & -x_2 & -x_3 \end{pmatrix}. \tag{3.39}$$

Some computations show that $G^{-1} = G^\top$ and $\det(G) = 1$, hence $G \in SO_3$, and $Gx = (0, 0, -1)^\top = S$.
The sphere $\mathbb{S}^2$ can be identified with $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ via the *stereographic projection*

$$\varphi : \mathbb{S}^2 \setminus \{N\} : (x_1, x_2, x_3)^\top \mapsto \tfrac{1}{1-x_3}(x_1 + ix_2), \qquad \varphi(N) = \infty.$$

Let us study the map $\varphi(G\varphi^{-1}(\cdot)) : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$. We want to identify this map with a *linear fractional transformation* (LFT) $\mu(z) = \frac{\alpha z + \beta}{\gamma z + \delta}$ for suitable $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. Note that each LFT is uniquely determined by specifying the image of three points, see [28, Proposition III.3.9]. Hence let us determine the LFT μ which satisfies $\mu(1) = \varphi(Ge_1)$, $\mu(i) = \varphi(Ge_2)$ and $\mu(0) = \varphi(-Ge_3)$. From this equations we obtain after some calculations $\alpha = \delta = 1 - x_3$ and $\beta = -\bar{\delta} = -(x_1 + ix_2)$. Hence $\mu(z) = \frac{(1-x_3)z - (x_1 + ix_2)}{(x_1 - ix_2)z + (1-x_3)}$ is a LFT which corresponds to a rotation of the form (3.39), so that $c = \varphi(x)$ is mapped into $\mu(c) = 0$ and $\varphi(-x)$ into ∞. The map $c \mapsto \varphi(-\varphi^{-1}(c))$ in $\hat{\mathbb{C}}$ is called the *antipodal map*. It is given by $c \mapsto -\bar{c}^{-1}$ as

$$\varphi((-x_1, -x_2, -x_3)^\top) = \tfrac{1}{1+x_3}(-x_1 - ix_2) = -\overline{(\varphi((x_1, x_2, x_3)^\top))}^{-1},$$

since the inverse of $\frac{1}{1-x_3}(x_1 + ix_2)$ for $(x_1, x_2, x_3)^\top \in \mathbb{S}^2$ is given by $\frac{1}{1+x_3}(x_1 - ix_2)$.
The following lemma provides us with the main tool for the proof of Spijker's Lemma. A *rational function* of degree n is the quotient of two coprime polynomials for which at least one has the maximal degree n.

Lemma 3.50. *Suppose that q is a complex rational function of degree $n \geq 1$. Then there exist linear fractional transformations μ_1 and μ_2, such that*

$$\mu_2 \circ q \circ \mu_1(z) = zr(z),$$

where r is a complex rational function of degree $n-1$. These LFTs are given by $\mu_1(z) = \frac{z-c}{\bar{c}z-1}$ and $\mu_2(z) = \frac{\alpha z - \beta}{\bar{\beta} z + \bar{\alpha}}$ for suitable $\alpha, \beta, c \in \mathbb{C}$.

Proof. A solution $c \in \mathbb{C}$ of

$$q(\bar{c}^{\,-1}) = -\overline{q(c)}^{\,-1}. \tag{3.40}$$

always exists as (3.40) contains only rational expressions in $\bar{c}$. Hence after expanding the left hand side of (3.40) with powers of $\bar{c}$ and then expanding with the denominators we obtain a non-trivial polynomial equation in $\bar{c}$. Note that if c solves (3.40) then so does $\bar{c}^{-1}$. Thus this equation always has a solution c with $|c| < 1$. The case $|c| = 1$ is impossible, as this would imply $q(c)\overline{q(c)} = -1$. For $\mu_1(z) = \frac{c-z}{1-\bar{c}z}$ we therefore have $\mu_1 : 0 \mapsto c$ and $\infty \mapsto \bar{c}^{-1}$. Now by (3.40), $\zeta := q \circ \mu_1(0)$ and $q \circ \mu_1(\infty)$ are antipodal points on the Riemannian sphere. By our previous discussion there exists a linear fractional transformation

$$\mu_2(z) = \frac{\alpha z - \beta}{\bar{\beta} z + \bar{\alpha}}, \qquad \alpha = (1 + |\zeta|^2)^{-1/2},\ \beta = \alpha\zeta, \tag{3.41}$$

which maps these antipodal points ζ and $-\bar{\zeta}^{-1}$ into 0 and ∞, respectively, as we have

$$\mu_2 \circ q \circ \mu_1(0) = \mu_2(\zeta) = \frac{\alpha\zeta - \alpha\zeta}{\bar{\alpha}\,|\zeta|^2 + \bar{\alpha}} = 0, \quad \mu_2 \circ q \circ \mu_1(\infty) = \mu_2(-\bar{\zeta}^{\,-1}) = \frac{-\alpha\bar{\zeta}^{\,-1} - \alpha\zeta}{-\bar{\alpha}\bar{\zeta}\bar{\bar{\zeta}}^{\,-1} + \bar{\alpha}} = \infty.$$

Therefore a z term factors out from the rational function $\mu_2 \circ q \circ \mu_1(z)$. It remains to show that the rational function r given by $zr(z) = \mu_2 \circ q \circ \mu_1(z)$ has rank $n-1$. Here the linear fractional transformations do not change the degree of q. Hence $\mu_2 \circ q \circ \mu_1$ is also of degree n. By factoring out z we have eliminated a pole at ∞ and a root at 0, hence the remaining rational function r is of degree $n-1$. □

Before we proceed let us comment on the definition of the contour integral. If $\varphi(t) : [a,b] \to \hat{\mathbb{C}}$, $a < b$, is a parameterized curve with trace $\varphi = \Gamma \subset \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ then $\int_\Gamma |g(s)|\,ds := \int_a^b |g(\varphi(s))|\,|\varphi'(t)|\,dt$. With the convention $a < b$ this integral is independent of the orientation of the curve. Let us demonstrate this by integrating the great circle $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$,

$$\int_{\hat{\mathbb{R}}} |f(s)|\,ds = \int_{-1}^{1} \frac{|f(\operatorname{artanh} t)|}{1-t^2} dt = \int_{-\infty}^{\infty} |f(t)|\,dt.$$

Here $\operatorname{artanh} : (-1,1) \to \mathbb{R}, x \mapsto \frac{1}{2}\log\frac{1+x}{1-x}$ is the inverse function of tanh.
Wegert and Trefethen [146] call the following theorem "Spijker's lemma on the Riemannian sphere".

Theorem 3.51. *Suppose that q is a complex rational function of degree n. Then*

$$L_n(q) := 2\int_{\mathbb{S}} \frac{|q'(s)|}{1 + |q(s)|^2} ds \le 2\pi n, \qquad \mathbb{S} = \left\{ e^{i\theta} \in \mathbb{C} \,\middle|\, \theta \in [-\pi, \pi] \right\}.$$

Proof. Let $n \ge 1$. By Lemma 3.50 there exist two linear fractional transformations μ_1 and μ_2 such that $\mu_2 \circ q \circ \mu_1(z) = zr(z)$ holds. Then we have $L_n(q) = L_n(\mu_2 \circ q \circ \mu_1)$ since $L_n(\hat{q} \circ \mu_1) = L_n(\hat{q})$ and $L_n(\mu_2 \circ \tilde{q}) = L_n(\tilde{q})$ for any rational functions $\hat{q}, \tilde{q}$ of degree n. To show the first fact $L_n(\hat{q} \circ \mu_1) + L_n(\hat{q})$, note that μ_1 given by $z \mapsto \frac{c-z}{1-\bar{c}z}$ maps the unit sphere $\mathbb{S}$ into itself, and μ_1 is self-inverse with $\mu_1 \circ \mu_1(z) = z$. To this end, note that for $z \in \mathbb{S}$,

$$|\mu_1(z)|^2 = \frac{c-z}{1-\bar{c}z} \cdot \frac{\bar{c}-\bar{z}}{1-c\bar{z}} = \frac{|z|^2 - \bar{c}z - c\bar{z} + |c|^2}{1 - \bar{c}z - c\bar{z} + |c|^2\,|z|^2} = 1,$$

and for $z \in \mathbb{C}$,

$$\mu_1 \circ \mu_1(z) = \left(c - \tfrac{c-z}{1-\bar{c}z}\right)\left(1 - \bar{c}\tfrac{c-z}{1-\bar{c}z}\right)^{-1} = (c - c\bar{c}z - c + z)(1 - \bar{c}z - c\bar{c} + \bar{c}z)^{-1} = \tfrac{1-c\bar{c}}{1-c\bar{c}}z = z.$$

Hence, instead of parameterizing $\mathbb{S}$ via $t \mapsto e^{it}$, $t \in [-\pi,\pi]$ we can use $t \mapsto \mu_1(e^{it})$, $t \in [-\pi,\pi]$.

$$\begin{aligned} L_n(\hat{q} \circ \mu_1) &= 2\int_{\mathbb{S}} \frac{|(\hat{q} \circ \mu_1)'(s)|}{1 + |\hat{q}(\mu_1(s))|} ds = 2\int_{-\pi}^{\pi} \frac{|\hat{q}\,'(\mu_1 \circ \mu_1(e^{it}))\mu_1'(\mu_1(e^{it}))|}{1 + |\hat{q}(\mu_1 \circ \mu_1(e^{it}))|} \left|ie^{it}\mu_1'(e^{it})\right| dt \\ &= 2\int_{-\pi}^{\pi} \frac{|\hat{q}\,'(e^{it})|}{1 + |\hat{q}(e^{it})|} \left|ie^{it}\mu_1'(e^{it})\mu_1'(\mu_1(e^{it}))\right| dt = 2\int_{-\pi}^{\pi} \frac{|\hat{q}\,'(e^{it})|}{1 + |\hat{q}(e^{it})|} \left|ie^{it}\right| dt = L_2(\hat{q}), \end{aligned}$$

since $\mu_1'(z)\mu_1'(\mu_1(z)) = (\mu_1 \circ \mu_1)'(z) = 1$. The second fact $L_n(\mu_2 \circ \tilde{q}) = L_n(\tilde{q})$ can be verified using the following formulas derived from (3.41),

$$\mu'(z) = \frac{1}{(\bar{\beta}z + \bar{\alpha})^2} \quad \text{and} \quad 1 + |\mu(z)|^2 = \frac{1 + |z|^2}{(\bar{\beta}z + \bar{\alpha})(\beta\bar{z} + \alpha)},$$

from which we obtain

$$L_n(\mu_2 \circ \tilde{q}) = 2\int_{\mathbb{S}} \frac{|\mu'(\tilde{q}(s))\,\tilde{q}'(s)|}{1 + |\mu_2(\tilde{q}(s))|^2} ds = 2\int_{\mathbb{S}} \frac{\left|\bar{\beta}\tilde{q}(z) + \bar{\alpha}\right|^2}{\left|\bar{\beta}\tilde{q}(z) + \bar{\alpha}\right|^2} \frac{|\tilde{q}'(s)|}{1 + |\tilde{q}(s)|^2} ds = L_n(\tilde{q}).$$

We therefore have

$$\begin{aligned} L_n(q(s)) = L_n(sr(s)) = 2\int_{\mathbb{S}} \frac{|r(s) + sr'(s)|}{1 + |r(s)|^2} ds &\le 2\int_{\mathbb{S}} \left(|s| \frac{|r'(s)|}{1 + |r(s)|^2} + \frac{|r(s)|}{1 + |r(s)|^2}\right) ds \\ &\le L_{n-1}(r) + 2\int_{\mathbb{S}} \frac{|r(s)|}{1 + |r(s)|^2} ds \le L_{n-1}(r(s)) + 2\pi, \end{aligned}$$

because $|s| = 1$ for $s \in \mathbb{S}$ and $\frac{x}{1+x^2} = (x + \frac{1}{x})^{-1} \le \frac{1}{2}$ for all $x \ge 0$. Now, $L_0 = 0$ and therefore an induction over n proves the theorem. □

The standard formulation of Spijker's Lemma is now a corollary.

Corollary 3.52. *Suppose that r is a complex rational function of degree n. Then*

$$\int_{\mathbb{S}} |r'(s)|\, ds \le 2\pi n \sup_{z \in \mathbb{S}} |r(z)|\,. \tag{3.42}$$

Proof. Consider the polynomial $q(z) = \|r\|_\infty^{-1} r(z)$ where $\|r\|_\infty = \sup_{z \in \mathbb{S}} |r(z)|$. Then

$$\int_{\mathbb{S}} |q'(s)|\, ds \le 2\int_{\mathbb{S}} \frac{|q'(s)|}{1 + |q(s)|^2} ds \quad \text{since } \frac{2}{1 + |q(z)|^2} \ge 1 \text{ as } |q(z)| \le 1 \text{ for all } z \in \mathbb{S}. \tag{3.43}$$

By Theorem 3.51 we have $\int_{\mathbb{S}} |q'(s)|\, ds \le 2\int_{\mathbb{S}} \frac{|q'(s)|}{1+|q(s)|^2}\, ds \le 2\pi n$, and a multiplication with $\|r\|_\infty$ gives (3.42). □

But as the Kreiss Theorem in its classical form ("Ein Satz über Matrizen" in [88]) is only for matrix powers, we need the following reformulation of Spijker's Lemma for the matrix exponential case.

Corollary 3.53. *Let $r(s)$ be a complex rational function of degree $n \geq 1$. For a given $\alpha \in \mathbb{R}$ set $\Gamma_\alpha = \{z \in \hat{\mathbb{C}} \,|\, z = \alpha + i\omega\}$. If $\sup_{z\in\Gamma_\alpha} |r(z)| < \infty$ then*

$$\int_{\Gamma_\alpha} |r'(s)|\, ds \leq 2\pi n \sup_{z\in\Gamma_\alpha} |r(z)| \,. \tag{3.44}$$

Proof. The linear fractional transformation given by $\mu(z) = \frac{(\alpha+1)z-(1-\alpha)}{z+1}$ maps the unit circle $\mathbb{S}$ onto Γ_α. We set $\xi : [-\pi, \pi] \to \mathbb{S}$, $t \mapsto e^{it}$. Define $\gamma : [-\pi, \pi] \to \Gamma_\alpha$, $t \mapsto \mu \circ \xi(t)$ (here $\gamma(\pm\pi) = \infty$). By assumption, $r(s) = \frac{p(s)}{q(s)}$ is a proper rational function of degree n. Then the degree of the numerator is $\deg(p'q - pq') \leq 2n - 2$ as the coefficient of the leading power cancels out. Therefore the degrees of the numerator and denominator of r' differ by at least 2. Hence the integral in the left hand side of (3.44) is well-defined. Now $r' \circ \mu(s) = (r \circ \mu)'(s)\mu'(s)^{-1}$. Setting $s = \gamma(t)$ gives

$$\begin{aligned} L := \int_{\Gamma_\alpha} |r'(s)|\, ds = \int_{-\pi}^{\pi} |r' \circ \gamma(t)|\, |\gamma'(t)|\, dt = \int_{-\pi}^{\pi} |r' \circ (\mu \circ \xi)(t)|\, |(\mu \circ \xi)'(t)|\, dt \\ = \int_{-\pi}^{\pi} \left|(r \circ \mu)'(\xi(t))(\mu' \circ \xi(t))^{-1}\right| |(\mu' \circ \xi(t))\, \xi'(t)|\, dt \\ = \int_{-\pi}^{\pi} |(r \circ \mu)'(\xi(t))|\, |\xi'(t)|\, dt = \int_{\mathbb{S}} |(r \circ \mu)'(s)|\, ds. \end{aligned}$$

Now, we can apply Corollary 3.52 to the rational function $r \circ \mu$ which is also of degree n. Hence

$$L \leq 2\pi n \sup_{z\in\mathbb{S}} |r \circ \mu(z)| = 2\pi n \sup_{z\in\Gamma_\alpha} |r(z)| \,.$$

□

We now have a suitable version of Spijker's Lemma in form of Corollary 3.53 available. Hence let us proceed with the proof of Theorem 3.49.

Proof (of Theorem 3.49). The lower bound in (3.38),

$$k(A, B, C) = \sup_{\mathrm{Re}\, s>0} \mathrm{Re}\, s \left\|C(sI - A)^{-1}B\right\| \leq \sup_{t\geq 0} \left\|Ce^{At}B\right\| = M_0(A, B, C),$$

is a direct consequence of Theorem 3.48 for $k = 1$ and $\beta = 0$. The upper bound in (3.38) is obtained by representing the matrix exponential as the Cauchy integral

$$e^{At} = \frac{1}{2\pi i} \int_\Gamma e^{st}(sI_n - A)^{-1} ds, \tag{3.45}$$

where Γ is any positively oriented simple smooth curve encircling the eigenvalues of A. Let us assume that $M_0(A,B,C)$ is attained at a finite time t_0. Then there exists a pair of vectors $(x,y) \in \mathbb{C}^\ell \times \mathbb{C}^q$ such that

$$M_0(A,B,C) = \max_{t\geq 0} \left\|Ce^{At}B\right\| = \max_{t\geq 0} \left|y^*Ce^{At}Bx\right| = \left|y^*Ce^{At_0}Bx\right|, \quad \|y\|^*_{\mathbb{C}^q} = 1 = \|x\|_{\mathbb{C}^\ell}\,.$$

Inserting (3.45) with $t = t_0$ into this equation gives

$$M_0(A,B,C) {=} \tfrac{1}{2\pi} \left|\int_\Gamma e^{st_0} y^*C(sI-A)^{-1}Bx\right| ds {=} \tfrac{1}{2\pi} \left|\int_\Gamma e^{st_0}\gamma(s)\right| ds {=} \tfrac{1}{2\pi} \sup_{t>0} \left|\int_\Gamma e^{st}\gamma(s)\right| ds.$$

Here $\gamma(s) = y^*C(sI-A)^{-1}Bx$ is a scalar rational function of degree $\leq n$. For a fixed t let the path of integration be given by $\Gamma = \{z \in \mathbb{C} \,|\, \operatorname{Re} z = t^{-1}\}$ which we interpret as a closed curve $\Gamma \cup \{\infty\}$ in $\hat{\mathbb{C}}$. On this contour we have $e^{t \operatorname{Re} s} = e$. Therefore the partial integration $\int_\Gamma e^{st}\gamma(s)ds = -\int_\Gamma \frac{1}{t}e^{st}\gamma'(s)ds$ gives

$$\left|y^*Ce^{At}Bx\right| = \frac{1}{2\pi t}\left|\int_\Gamma e^{st}\gamma'(s)\right| ds \leq \frac{e}{2\pi t}\int_\Gamma |\gamma'(s)|\, ds.$$

Applying Corollary 3.53 we obtain

$$\left|y^*Ce^{At}Bx\right| \leq \tfrac{2\pi en}{2\pi t} \sup_{s\in\Gamma} |\gamma(s)| = \tfrac{en}{t} \sup_{\omega\in\mathbb{R}} \left|y^*C((\tfrac{1}{t}+i\omega)I-A)^{-1}Bx\right|.$$

Maximization over all $t > 0$ yields for $s = t^{-1} + i\omega$

$$\begin{aligned} M_0(A,B,C) &\leq \sup_{t>0} \tfrac{en}{t} \sup_{\omega\in\mathbb{R}} \left\|C((\tfrac{1}{t}+i\omega)I-A)^{-1}B\right\| \\ &= en \sup_{\operatorname{Re} s>0} \operatorname{Re} s \left\|C(sI-A)^{-1}B\right\| = en\, k(A,B,C). \end{aligned}$$

This proves the upper bound in (3.38). □

With the formula presented in Theorem 1.12 it is easy to see that we can express the Kreiss constant via properties of spectral value sets for full block perturbations.

Corollary 3.54. *The Kreiss constant can be expressed in terms of the stability radius,*

$$k(A,B,C) = \sup_{\gamma>0} \gamma\, r(A-\gamma I_n, B, C)^{-1}, \tag{3.46}$$

and in terms of the pseudospectral abscissa,

$$k(A,B,C) = \sup_{\varepsilon>0} \varepsilon^{-1}\alpha_\varepsilon(A,B,C). \tag{3.47}$$

Proof. For full block perturbation structures Theorem 1.12 provides us with the following formula for the spectral value set of A with respect to the level $\varepsilon > 0$,

$$\sigma_\varepsilon(A,B,C) = \sigma(A) \cup \left\{ s \in \mathbb{C} \setminus \sigma(A) \,\middle|\, \left\| C(sI-A)^{-1}B \right\| > \varepsilon^{-1} \right\}.$$

The associated spectral abscissa is given by

$$\alpha_\varepsilon(A,B,C) = \sup \left\{ \operatorname{Re} s \,\middle|\, \left\| C(sI-A)^{-1}B \right\| > \varepsilon^{-1} \right\} = \sup \left\{ \operatorname{Re} s \,\middle|\, \left\| C(sI-A)^{-1}B \right\| = \varepsilon^{-1} \right\}$$

and the stability radius satisfies

$$r(A,B,C) = \left(\sup_{\omega\in\mathbb{R}} \left\| C(i\omega - A)^{-1}B \right\| \right)^{-1}.$$

The Kreiss constant is given by $k(A,B,C) = \sup_{\operatorname{Re} s>0} \operatorname{Re} s \, \|C(sI-A)^{-1}B\|$. We can split $s \in \varrho(A)$ into real and imaginary part, $s = \gamma + i\omega$, with $\gamma, \omega \in \mathbb{R}$. Then

$$k(A,B,C) = \sup_{\gamma>0} \gamma \sup_{\omega\in\mathbb{R}} \left\| C(i\omega I - (A-\gamma I))^{-1}B \right\| = \sup_{\gamma>0} \gamma \, (r(A,B,C))^{-1}$$

which shows (3.46). For (3.47), we consider $s \mapsto \operatorname{Re} s \, \|C(sI-A)^{-1}B\|$ on the contours of $\partial\sigma_\varepsilon(A,B,C) \cap \mathbb{C}_+$. We have

$$\begin{aligned} k(A,B,C) &= \sup_{\varepsilon>0} \sup_{s\in\partial\sigma_\varepsilon(A,B,C)\cap\mathbb{C}_+} \operatorname{Re} s \left\| C(sI-A)^{-1}B \right\| \\ &= \sup_{\varepsilon>0} \varepsilon^{-1} \sup_{\{s\in\mathbb{C}_+ \,|\, \|C(sI-A)^{-1}B\| = \varepsilon^{-1}\}} \operatorname{Re} s = \sup_{\varepsilon>0} \varepsilon^{-1} \alpha_\varepsilon(A,B,C), \end{aligned}$$

so that (3.47) is obtained. □

Thus if A is a stable matrix and for small $\varepsilon > 0$ the spectral value sets $\sigma_\varepsilon(A,B,C)$ move deeply into the right half-plane, then there are some trajectories of the system $\dot{x} = Ax$ with large transient excursions.
Let us collect the set of points where the pseudospectral abscissa is attained.

Definition 3.55. Given $A \in \mathbb{C}^{n\times n}$ and structure matrices $B \in \mathbb{C}^{\ell\times n}$, $C \in \mathbb{C}^{q\times n}$, the set of points

$$\mathcal{F}(A,B,C) := \bigcup_{\varepsilon>0} \{ z \in \partial\sigma_\varepsilon(A,B,C) \,|\, \operatorname{Re} z = \alpha_\varepsilon(A,B,C) \}$$

is called the *front locus* of A with respect to the structure matrices B and C.

Note that the Kreiss constant can be obtained by maximizing $\operatorname{Re} z \; \|C(zI-A)^{-1}B\|$ over all $z \in \mathcal{F}(A,B,C) \cap \mathbb{C}_+$ instead of over the half-plane $\{z \in \mathbb{C} \,|\, \operatorname{Re} z > 0\}$.

3.5.2 Calculating the Front Locus

We have seen in the previous subsection that the Kreiss constant $k(A)$ may be obtained by maximizing the quotient $\alpha_\varepsilon(A,B,C)/\varepsilon$. Let us now consider the unstructured case $B = I$ and $C = I$ and fix the spectral norm $\|\cdot\| = \|\cdot\|_2$. The additional information of computing the front locus $\mathcal{F}(A) = \mathcal{F}(A,I,I)$ comes nearly for free when computing the spectral value sets of A.

Proposition 3.56. *Suppose that $s \in \mathcal{F}(A)$ is a point in the front locus of $A \in \mathbb{C}^{n\times n}$. If $u, v : \mathbb{C} \to \mathbb{C}^n$ are the left and right singular vectors corresponding to the smallest singular value σ_n of $sI - A$ then* $\operatorname{Im} u(s)^* v(s) = 0$. *If we define $\mathcal{F}^* = \{s \in \mathbb{C} \,|\, \operatorname{Im} u(s)^* v(s) \le 0\}$ then $\mathcal{F} \subset \partial\mathcal{F}^* \cap \mathbb{C}_+$.*

Proof. Let us consider the function $s \mapsto \|(sI - A)^{-1}\|$ along lines parallel to the imaginary axis, i.e., with $\operatorname{Re} s$ fixed. Let $u(\omega)$ and $v(\omega)$ be the singular vectors corresponding to the minimal singular value of the function $\omega \mapsto (\alpha + i\omega)I - A$. By Theorem 3.16, $u(\cdot)$ and $v(\cdot)$ are piecewise analytic. For a given real part $\alpha > 0$ we consider the function

$$f_\alpha : \omega \mapsto \left\|((\alpha + i\omega)I - A)^{-1}\right\| = \sigma_n((\alpha + i\omega)I - A) = u(\omega)^*(i\omega I - (A - \alpha I))v(\omega).$$

Let $s_0 = \alpha + i\omega_0 \in \mathcal{F}$. Then there exists $\varepsilon_0 > 0$ such that $s_0 \in \partial\sigma_{\varepsilon_0}(A)$ and $\alpha_{\varepsilon_0}(A) = \operatorname{Re} s_0 = \alpha$. Hence we obtain from Theorem 1.12 that for all $\zeta \in \mathbb{C}$ with $\operatorname{Re}\zeta \ge 0$,

$$\sigma_n((s_0 + \zeta)I - A) \not< \varepsilon_0.$$

Thus $f_\alpha(\cdot)$ attains a local minimum in ω_0. The function f_α is differentiable in local minima, which can be shown analogously to the differentiability of σ_1 in local maxima, see the proof of Proposition 3.29 for details. Now the derivative $f'_\alpha(\omega) := \frac{d}{d\omega} f_\alpha(\omega)$ is given by

$$\begin{aligned} f'_\alpha(\omega) &= \operatorname{Re}\ (u'(\omega)^*((\alpha + i\omega)I - A)v(\omega) + iu(\omega)^*v(\omega) + u(\omega)^*((\alpha + i\omega)I - A)v'(\omega)) \\ &= \operatorname{Re} iu(\omega)^*v(\omega) = \operatorname{Im} u(\omega)^*v(\omega), \end{aligned}$$

because $u'(\omega)^*u(\omega) = 0 = v(\omega)^*v'(\omega)$ since $u(\omega)$, $v(\omega)$ are both of unit length. Thus a necessary condition for local minima of f_α is given by $\operatorname{Im} u^*v = 0$. Clearly, the front locus satisfies $\mathcal{F} \subset \{s \in \mathbb{C}_+ \,|\, \operatorname{Im} u(s)^*v(s) = 0\}$. Now inner points of $\mathcal{F}^*$ for which $\operatorname{Im} u_n^* v_n = 0$ correspond to a saddle-point of f_α as no sign change occurs in the derivative. □

Example 3.57. Figure 3.7 shows the spectral value sets and the set $\mathcal{F}^*$ for the matrix

$$A = \begin{pmatrix} -0.4 & -1 & -4 & \\ 2 & -0.4 & 4 & \\ & & -1.6 & 1 \\ & & & -0.4 \end{pmatrix}.$$

Here $\mathcal{F}^*$ consists of three connected components, hence the gap for $\operatorname{Re} s = 0.8$ is not an artefact of the computational grid. Note that the real axis is part of the set $\mathcal{F}^*$. ■

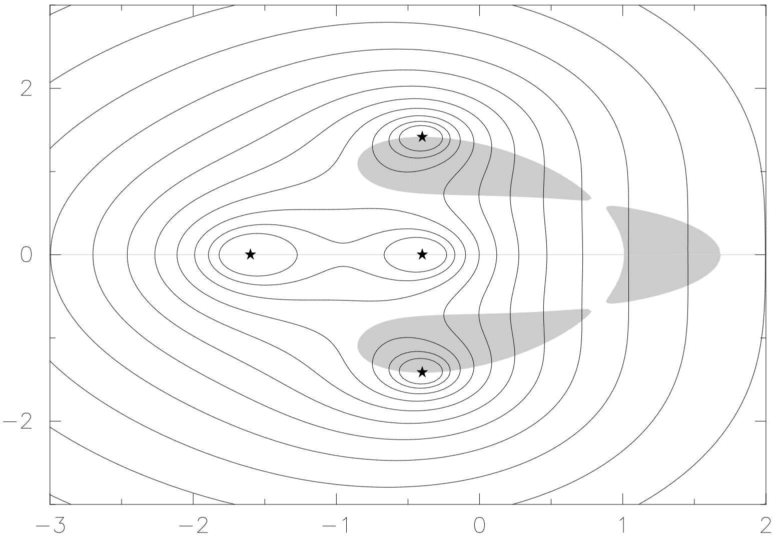

Figure 3.7: Pseudospectra and $\mathcal{F}^*$.

We now present a fast method which calculates the minimal singular vectors of the matrix $sI - A$. The following lemma shows how the singular values can be obtained from the eigenvalues of an Hermitian matrix.

Lemma 3.58 ([91, p. 190])**.** *Let A be a matrix in $\mathbb{C}^{n\times n}$. Then the spectrum of $H = \left(\begin{smallmatrix} 0 & A \\ A^* & 0 \end{smallmatrix}\right) \in \mathbb{C}^{2n\times 2n}$ is given by $\sigma(H) = \{\pm\sigma_k(A) \mid k = 1, \ldots, n\}$ where $\sigma_k(A) \geq 0$ is the kth singular value of A.*

Sophisticated algorithms to deal with such Hamiltonian eigenvalue problems are available in van Loan [139] and Benner, Mehrmann and Xu [112].

Proof. Let $\binom{u}{v} \in \mathbb{C}^{2n}$ be an eigenvector corresponding to an eigenvalue λ of H. Since $H \in \mathcal{H}^n$, its spectrum is real, hence $\lambda \in \mathbb{R}$. Now $H\binom{u}{v} = \lambda\binom{u}{v}$ is equivalent to $A^*u = \lambda v$, $Av = \lambda u$. This implies that $AA^*u = \lambda Av = \lambda^2 Au$ and $A^*Av = \lambda A^*u = \lambda^2 u$ hence $|\lambda|$ is a singular value of A. If $(\lambda, \binom{u}{v})$ is an eigenpair of H then it is easy to verify that $(-\lambda, \binom{u}{-v})$ is also an eigenpair of H. Therefore $\sigma(H) = \{\pm\sigma_k(A)\}$. □

We will present a simple analysis to show that the term $\operatorname{Im} u^*v$ is available with virtually no additional costs when computing the pseudospectra of A. In particular, if A is given in complex Schur form then $B = sI - A$ is an upper triangular matrix for all $s \in \mathbb{C}$. For simplicity, let us assume that $\sigma_n(B) \neq 0$, that is, $s \notin \sigma(A)$, and that $\sigma_n(B) < \sigma_{n-1}(B)$. The inverse power iteration $\left(\begin{smallmatrix} 0 & B \\ B^* & 0 \end{smallmatrix}\right)\binom{u^+}{v^+} = \left\|\binom{u}{v}\right\|_2^{-1}\binom{u}{v}$ can be written as

$$\begin{aligned} &\tilde{u}^{j+1} = B^{-*}v^j, \qquad \tilde{v}^{j+1} = B^{-1}u^j, \qquad \sigma_{j+1} = \left(\sum_{k=1}^{n} (\tilde{u}_k^{j+1})^2 + (\tilde{v}_k^{j+1})^2\right)^{-1/2}, \\ &u^{j+1} = \sigma_{j+1}\tilde{u}^{j+1}, \qquad v^{j+1} = \sigma_{j+1}\tilde{v}^{j+1}. \end{aligned} \tag{3.48}$$

Here $\tilde{u}^{j+1} = B^{-*}v^j$ and $\tilde{v}^{j+1} = B^{-1}u^j$ can be solved by computationally inexpensive forward and backward substitutions[1]. Now, if the initial values u^0 and v^0 are chosen in such way that $\binom{u^0}{v^0}$ is not contained in a non-trivial H-invariant subspace then the sequence $(\sigma_i)_{i\in\mathbb{N}}$ defined in (3.48) satisfies $\sigma_i \to \sigma_n(B)$ for $i \to \infty$. However the "dominant" eigenspace of $H = \left(\begin{smallmatrix} 0 & B \\ B^* & 0 \end{smallmatrix}\right)$ (i.e., the one associated with eigenvalues which has the smallest distance to 0) is not uniquely determined since by Lemma 3.58 the minimal distance to 0 is attained for both $\sigma_n(B)$ and $-\sigma_n(B)$. Therefore the vectors u^j and v^j will not converge although σ_j converges to $\sigma_n(B)$. But if this minimal singular value is of multiplicity one, i.e., $\sigma_n(B) \neq \sigma_{n-1}(B)$, then this sequence of vectors will approach an oscillation between two vectors contained in the subspace spanned by the two eigenvectors of H which are associated with eigenvalues λ_1, λ_2 that satisfy $|\lambda_i| = \sigma_n(A)$, $i = 1, 2$. This cycle is given by $\{\alpha\binom{u}{v} + \beta\binom{u}{-v}, \beta\binom{u}{v} + \alpha\binom{u}{-v}\}$ where $\binom{u}{v}$ is an eigenvector of H with $H\binom{u}{v} = \sigma_n(A)\binom{u}{v}$, and $\alpha, \beta \in \mathbb{C}$ are constants depending on the initial values u^0, v^0. In particular, if $H\binom{u}{v} = \lambda\binom{u}{v}$ then $H\binom{u}{-v} = -\lambda\binom{u}{-v}$ and therefore we get

$$H\begin{pmatrix}(\alpha+\beta)u \\ (\alpha-\beta)v\end{pmatrix} = H\left(\alpha\begin{pmatrix}u \\ v\end{pmatrix} + \beta\begin{pmatrix}u \\ -v\end{pmatrix}\right) = \lambda\left(\alpha\begin{pmatrix}u \\ v\end{pmatrix} - \beta\begin{pmatrix}u \\ -v\end{pmatrix}\right) = \lambda\begin{pmatrix}(\alpha-\beta)u \\ (\alpha+\beta)v\end{pmatrix}.$$

As the sequence (3.48) is renormalized with $\lambda = \sigma_n(A)$ we obtain a cycle between these two elements. To get an approximation of the eigenvector $\binom{u}{v}$ we add two subsequent terms of (3.48), so

$$2\alpha\begin{pmatrix}u \\ v\end{pmatrix} = \begin{pmatrix}(\alpha+\beta)u \\ (\alpha-\beta)v\end{pmatrix} + \begin{pmatrix}(\alpha-\beta)u \\ (\alpha+\beta)v\end{pmatrix} \approx \begin{pmatrix}u^{j+1} \\ v^{j+1}\end{pmatrix} + \begin{pmatrix}u^j \\ v^j\end{pmatrix}. \tag{3.49}$$

Furthermore, we are only interested in the sign of $\operatorname{Im} u^*v$ and therefore a renormalization of the u- and v-components in step (3.49) is not necessary. Collecting these ideas we obtain the following outline of an algorithm.

Algorithm 3.59. We determine the pseudospectra and front locus of $A \in \mathbb{C}^{n\times n}$.

Initialize Replace A by its (complex) Schur form, hence making it upper triangular. Create a grid $G \subset \mathbb{C}$. Allocate storage for grid-sized real matrices P and F.

Main Loop For each grid point $z \in G$ set $A_z = zI - A$. Choose initial values $u^0, v^0 \in \mathbb{C}^n$ and iterate

$$\begin{aligned} \tilde{u}^j &= A_z^{-*}v^{j-1}, \qquad \tilde{v}^j = A_z^{-1}u^{j-1}, \\ \sigma_j &= \left(\left\|\tilde{u}^j\right\|^2 + \left\|\tilde{v}^j\right\|^2\right)^{-1/2}, \\ u^j &= \sigma_j\tilde{u}^j, \qquad v^j = \sigma_j\tilde{v}^j, \end{aligned}$$

until σ_j converges. Set $u = u^j + u^{j-1}$, $v = v^j + v^{j-1}$, and store σ_j and $\operatorname{Im} u^*v$ into P and F, respectively.

End Return the pseudospectra P and F of A.

[1] A solver for general triangular matrices is provided by the LAPACK function family xGETRS, see [1].

The matrix P produced by the above algorithm contains the minimal singular values evaluated for each grid point z, $\|(zI-A)^{-1}\|^{-1}$ while the front locus $\mathcal{F}(A)$ is part of the zero-contour of the height field of the matrix F. If $z = G_{ij}$ is a grid point contained in $\mathcal{F}^*(A)$ then $F_{ij} \leq 0$.

3.6 Notes and References

The notion of (M,β)-stability is somehow a hybrid between the concept of exponential stability and the notion of *practical stability* which has been introduced by LaSalle and Lefschetz [93]. A nonlinear differential system $\dot{x} = f(t,x)$, $f(t_0) = x_0$, $f(t,0) \equiv 0$ is called practically stable (see Lakshmikantham et al. [89]) in $x_* = 0$ for constants $0 < m < M$ if $\|x_0\| < m$ implies $\|x(t;t_0,x_0)\| < M$, $t \geq t_0$. Hence if $\dot{x} = Ax$ is uniformly $(M, \beta = 0)$-stable then it is practically stable for $m = 1$ and M.

Topological properties of the set of (M,β)-stable matrices are studied in Hinrichsen and Pritchard [67]. Generators of type $\mathcal{G}(M,\beta)$ have been discussed in Kato [77]. Classical bounds for the matrix exponential can be found in Moler and van Loan [108], which are mostly based upon an eigenvalue/eigenvector analysis. For an account on how to compute the matrix exponential via a scaling and squaring technique combined with a Padé approximation, see Higham [57] who suggests an algorithm that uses fewer multiplications than MATLAB's `expm` while improving the precision. Most of the computations of matrix exponentials presented in this thesis are obtained from an algorithm presented by Golub and van Loan [48, Algorithm 11.3.1].

The bounds presented in this section mostly concentrate on obtaining estimates for M and β. The bound (3.5) based upon knowledge of the eigenvalues and eigenvectors is mathematical folk tradition. However, as we have seen in Proposition 3.14 the asymptotic behaviour is not governed by $\kappa(V)$, but by $\sup_{\|x\|=1} \left|e_1^\top V^{-1}x\right|$. The bound for Jordan canonical forms in (3.6) is inspired by Higham [56] where analogous bounds are derived for matrix powers.

Although the spectral norm of a matrix is directly related to its SVD, estimates based on the SVD are to the best of the author's knowledge not found in the literature.

Bounds based on quadratic Liapunov functions enjoy a certain popularity, see Veselić [141]. This article also features bounds which are also valid for semigroups on Banach spaces and the underlying idea for the proof of Proposition 3.43. Transient estimates based upon quadratic Liapunov functions have been discussed in Hinrichsen, Plischke and Pritchard [62], where the problem of finding a Liapunov matrix with smallest condition number is also addressed. The notion of quadratic (M,β)-stability has been used in Boyd et al. [22] where optimization problems involving quadratic Liapunov matrices are mentioned. However, as Lemma 3.44 shows, the minimal condition number is always attained at the boundary of a Liapunov cone, which poses numerical problems for the solution. For convergence issues of the inverse power method see Wilkinson [148]. For a recent discussion of the initial growth rate associated with weighted quadratic norms, see Hu and Liu [72].

Bounds based upon the resolvent of A are discussed in Embree and Trefethen [37], see also the articles [136, 137] by Trefethen. An account on the history of the Kreiss Matrix Theorem is found in Wegert and Trefethen [146], see also Spijker [129] and Spijker et al. [130] where issues related to the Kreiss Matrix Theorem are discussed. A version of the Kreiss Matrix Theorem for exponentially stable matrices is found in Aupetit and Drissi [7]. The notion of the front locus has been suggested in [62]. For computational issues involving the pseudospectral abscissa, see Burke, Lewis and Overton [24].

Chapter 4

Examples

This chapter gathers various applications of estimates which have been presented in the last chapters. The organization is as follows. We first derive some explicit formulas for the norm of the matrix exponential. Then we take a closer look at transient Feller norms, and show a formula for an upper exponential estimate of 2×2 blocks which differs from the original by maximally 36%.

After that we compute the quadratic Liapunov matrix associated with a stable 2×2 matrix A, that has the smallest condition number under all solutions of a quadratic Liapunov inequality for A. The geometrical insight gained in this course is used to find joint quadratic Liapunov functions. We close this chapter with a discussion of dissipativity for polytopic norms, that comes in handy for a variety of mathematical applications.

4.1 Explicit Formulas

We will start off with the calculation of the exact transient growth for 2×2 block upper triangular matrices with respect to the spectral norm. These results can be used to judge the quality of the estimates.

Lemma 4.1. *Suppose that $B \in \mathbb{C}^{n\times m}$ and $\alpha, \beta \in \mathbb{C}$. Then the spectral norm of*

$$A = \begin{pmatrix} \alpha I_n & B \\ 0 & \beta I_m \end{pmatrix} \in \mathbb{C}^{(n+m)\times(n+m)} \tag{4.1}$$

is given by

$$\|A\| = \frac{1}{2}\left(\sqrt{(|\alpha| + |\beta|)^2 + \|B\|^2} + \sqrt{(|\alpha| - |\beta|)^2 + \|B\|^2}\right).$$

Remark 4.2. A matrix A with a Schur form (4.1) has a minimal polynomial given by $m_A(s) = s^2 - (\alpha + \beta)s + \alpha\beta$. Hence it satisfies the quadratic matrix polynomial equation $A^2 - (\alpha + \beta)A + (\alpha\beta)I = 0$.

Proof. For any eigenpair $(\lambda, \binom{u}{v})$ of A^*A we have

$$A^*A\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}|\alpha|^2 I & \bar{\alpha}B\\ \alpha B^* & B^*B + |\beta|^2 I\end{pmatrix}\begin{pmatrix}u\\v\end{pmatrix} = \lambda\begin{pmatrix}u\\v\end{pmatrix}. \tag{4.2}$$

If $\alpha = 0$ then $(B^*B + |\beta|^2 I)v = \lambda v$. Hence $\|A\| = \sqrt{|\beta|^2 + \|B\|^2}$ which proves the assertion. If $\alpha \neq 0$ and $u = 0$ or $v = 0$ then we obtain $\lambda = |\beta|^2$ or $\lambda = |\alpha|^2$, respectively. When we assume that both $u, v \neq 0$ and furthermore $\lambda, \alpha \neq 0$ then the following two equations follow from (4.2),

$$Bv = \frac{\lambda - |\alpha|^2}{\bar{\alpha}}u, \qquad B^*u = \frac{\bar{\alpha}}{\lambda}(\lambda - |\beta|^2)v. \tag{4.3}$$

The product of both constants appearing in (4.3) is an eigenvalue of B^*B,

$$\mu^2 := \lambda^{-1}(\lambda - |\alpha|^2)(\lambda - |\beta|^2) \in \sigma(B^*B).$$

Rearranging the term yields two solutions for λ depending on $\mu^2 \in \sigma(B^*B)$ given by

$$\lambda_\pm(\mu^2) = \frac{1}{2}\left(|\alpha|^2 + |\beta|^2 + \mu^2 \pm \sqrt{(|\alpha|^2 + |\beta|^2 + \mu^2)^2 - 4|\alpha\beta|^2}\right).$$

In particular, the maximal eigenvalue of A^*A corresponds to $\lambda_+(\mu^2)$ where $\mu^2 = \|B\|^2$, hence

$$\|A\| = \lambda_+(\|B\|^2)^{1/2} = \frac{1}{2}\left(\sqrt{(|\alpha| + |\beta|)^2 + \|B\|^2} + \sqrt{(|\alpha| - |\beta|)^2 + \|B\|^2}\right).$$

□

As a direct consequence of the unitary invariance of the spectral norm we obtain the following result.

Corollary 4.3. *Given $B \in \mathbb{C}^{m\times n}$, scalars α and β, and unitary matrices $U \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$. Then the spectral norm of $\left(\begin{smallmatrix}\alpha I_m & B\\ 0 & \beta I_n\end{smallmatrix}\right)$ equals the norm of $\left(\begin{smallmatrix}\beta I_m & UBV\\ 0 & \alpha I_n\end{smallmatrix}\right)$.*

Now we consider the matrix exponential for matrices given by (4.1).

Proposition 4.4. *Suppose that $A = \left(\begin{smallmatrix}\alpha I & B\\ 0 & \beta I\end{smallmatrix}\right)$ where $\alpha, \beta \in \mathbb{R}$ are real scalars and $B \in \mathbb{C}^{m\times n}$. Then*

$$\|e^{At}\| = \begin{cases} \frac{1}{2}\left|e^{\alpha t} - e^{\beta t}\right|\left(\sqrt{\coth(\frac{\alpha-\beta}{2}t)^2 + (\frac{\|B\|}{\alpha-\beta})^2} + \sqrt{1 + (\frac{\|B\|}{\alpha-\beta})^2}\right) & \text{if } \alpha \neq \beta,\\ \frac{1}{2}e^{\alpha t}\left(\sqrt{4 + (\|B\|\,t)^2} + \|B\|\,|t|\right) & \text{if } \alpha = \beta. \end{cases} \tag{4.4}$$

Proof. Suppose that $\alpha \neq \beta$. Then $e^{At} = \begin{pmatrix} e^{\alpha t}I & \frac{e^{\alpha t} - e^{\beta t}}{\alpha - \beta}B\\ 0 & e^{\beta t}I\end{pmatrix}$. By Lemma 4.1 the norm of the matrix exponential is given by

$$\|e^{At}\| = \frac{1}{2}\left(\sqrt{(e^{\alpha t} + e^{\beta t})^2 + \gamma^2} + \sqrt{(e^{\alpha t} - e^{\beta t})^2 + \gamma^2}\right),$$

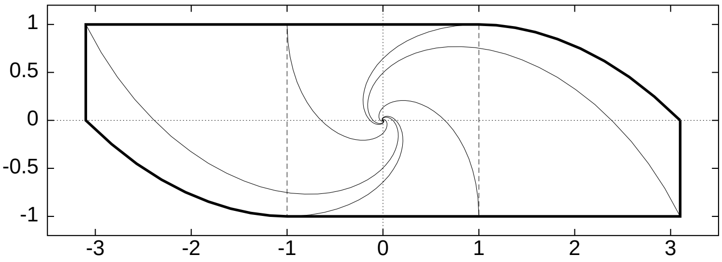

Figure 4.1: Feller norm for a rectangular unit box.

where $\gamma = \frac{e^{\alpha t}-e^{\beta t}}{\alpha-\beta}\,\|B\|$. Factoring out $(e^{\alpha t}-e^{\beta t})^2$ gives

$$= \frac{1}{2}\left|e^{\alpha t}-e^{\beta t}\right|\left(\sqrt{\left(\frac{e^{\alpha t}+e^{\beta t}}{e^{\alpha t}-e^{\beta t}}\right)^2+\left(\frac{\|B\|}{\alpha-\beta}\right)^2}+\sqrt{1+\left(\frac{\|B\|}{\alpha-\beta}\right)^2}\right).$$

Now, as $\coth(x) = \frac{e^{2x}+1}{e^{2x}-1}$ the first part of Proposition 4.4 is proved. In case $\alpha=\beta$ a limiting argument shows that $e^{At} = \left(\begin{smallmatrix} e^{\alpha t}I & te^{\alpha t}B \\ 0 & e^{\alpha t}I \end{smallmatrix}\right)$. Therefore Lemma 4.1 gives

$$\left\|e^{At}\right\| = \frac{1}{2}\sqrt{(2e^{\alpha t})^2+(te^{\alpha t}\,\|B\|)^2}+|t|\,e^{\operatorname{Re}\alpha t}\,\|B\|,$$

which proves the second case in (4.4). □

To find the maximum of $\sup_{t\geq 0}\left\|e^{At}\right\|$ one has to find the critical values of (4.4) which is not pursued here.

4.2 Construction of Transient Norms

In Definition 2.60 we introduced the Feller norm $\|x\|_A = \sup_{t\geq 0}\left\|e^{At}x\right\|$ as a norm for which the transient growth $M_0(A)$ is given by the eccentricity of $\|\cdot\|_A$. We will now show that this norm can be used to derive good estimates for the transient growth in the case of real 2×2 matrices. For a given vector norm ν on $\mathbb{R}^2$ we denote its unit sphere by $\mathbb{S}_\nu = \{x\in\mathbb{R}^2 \mid \nu(x)=1\}$ and its closed unit ball by $\mathbb{B}_\nu = \{x\in\mathbb{R}^2 \mid \nu(x)\leq 1\}$. We have demonstrated in Lemma 2.63 and in (2.48) how to construct transient norms for a stable matrix A.

Example 4.5. Consider the linear system

$$\dot{x} = Ax, \quad \text{where } A = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}. \tag{4.5}$$

It is easy to see that this system is contracting with respect to the maximum norm. For all x on the boundary of the unit square (which is this case a unit box), the vectorfield

$(x, \dot{x} = Ax)$ never points outside this box, as can be seen by checking the signs of the coefficients $\dot{x}_1$ resp. $\dot{x}_2$. The dashed lines in Figure 4.1 show the unit box of the maximum norm, and a few trajectories illustrate that this box is indeed invariant under (4.5). If we deform the norm (by introducing a diagonal weighting matrix) and choose a wider unit box, then with respect to this new norm, system (4.5) is not a contraction anymore as trajectories starting in the vertices now point outside the box. For example consider the rectangle $\mathcal{R} := [-\alpha, \alpha] \times [1, 1]$ as the unit sphere of a suitable norm ν where $\alpha = \sqrt{2}e^{\frac{\pi}{4}} > 1$. To construct the unit ball of a Feller norm we have to identify those points in the unit ball $\mathbb{B}_\nu$ of ν which are invariant under the flow of (4.5). The trajectory $x(t, x^0)$ of (4.5) starting in $x^0 = (\alpha, 0)^\top$ is given by $x(t, x^0) = \alpha e^{-t} (\cos(t), \sin(t))^\top$. This curve remains entirely inside the box $\mathcal{R}$, only touching the border in $(1, 1)^\top$. If we now clip away the area above the curve segment given by $t \in [0, \frac{\pi}{4}]$ and its symmetric part in the lower left corner, the remaining curve is the boundary of a convex and symmetric set $\mathcal{A}$ which contains a neighbourhood of the origin, so that the corresponding Minkowski function $\nu_{\mathcal{A}}(x) = \inf\{\gamma > 0 \,|\, \gamma^{-1}x \in \mathcal{A}\}$ is a norm. Moreover, all points in this set are backwards-stable under $\dot{x} = Ax$. The thick lines in Figure 4.1 mark its unit circle. This norm is the Feller norm ν_A of A associated with ν.

In Proposition 2.68 we introduced another method of constructing Liapunov norms, which we now illustrate in Figure 4.2. Instead of constructing a backwards-stable set, we now create a unit ball which is forward-stable under the flow of $\dot{x} = Ax$. This is done by following all trajectories starting in $\mathcal{R}$ and then taking the convex closure of all these sets. Figure 4.2 shows how the flow acts on the unit square $\mathcal{R}$. The dashed lines denote some snapshots $e^{At_0}\mathcal{R}$ of the unit box for $t_0 \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$. The transient norm ball is then given by $\overline{\text{conv}} \bigcup_{t \geq 0} e^{At}\mathcal{R}$. ■

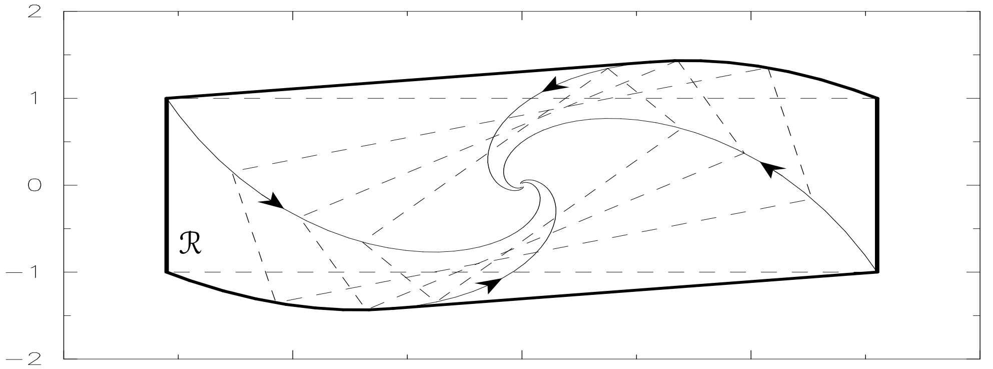

Figure 4.2: Dual transient norm for a rectangular unit box.

4.2.1 Marginally Stable Matrices

We now study Feller norms on $\mathbb{R}^2$ when the underlying norm is the Euclidean norm. As the Euclidean norm is invariant under unitary transformations we only need to consider Schur forms. We discuss the following cases of real Schur forms of stable 2×2 matrices.

1. The matrix A is 0. Then $\left\|e^{At}x\right\| = \|x\|$ for all $t \geq 0$ and all $x \in R^2$, hence this case is of no further interest.

2. The matrix A has two purely imaginary eigenvalues, $A = \left(\begin{smallmatrix} 0 & \beta \\ \alpha & 0 \end{smallmatrix}\right)$ with $\alpha\beta < 0$.

3. The matrix A is marginally stable, hence $A = \left(\begin{smallmatrix} 0 & \alpha \\ 0 & \lambda \end{smallmatrix}\right)$ with $\lambda < 0$.

4. The matrix A is exponentially stable with real spectrum, $A = \left(\begin{smallmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{smallmatrix}\right)$ and $\lambda_1, \lambda_2 < 0$.

5. The matrix A is exponentially stable with a pair of complex conjugate eigenvalues, $A = \left(\begin{smallmatrix} \lambda & \beta \\ \alpha & \lambda \end{smallmatrix}\right)$ where $\lambda < 0$ and $\alpha\beta < 0$.

If $A \in \mathbb{R}^{2\times 2}$ has two purely imaginary eigenvalues the real Schur form of A is given by $A = \left(\begin{smallmatrix} 0 & \beta \\ \alpha & 0 \end{smallmatrix}\right)$ where $\alpha\beta < 0$. Then for $P = \left(\begin{smallmatrix} |\alpha| & 0 \\ 0 & |\beta| \end{smallmatrix}\right)$ the matrix equation $PA + A^\top P$ equals zero, the solutions of $\dot{x} = Ax$ are contained in the level sets of $x \mapsto x^\top Px =: \|x\|_P^2$ which shows that $\|\cdot\|_P$ is invariant under application of the Feller norm generation process.
Moreover, the Feller norm with respect to the Euclidean norm is also a scalar multiple of this P-norm. Assume that wlog. $|\alpha| \leq |\beta|$. Then $x_0 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ corresponds to the larger principal axis of the ellipsoid. The solution $e^{At}x_0$ is entirely contained in the Euclidean unit ball. Hence the Feller norm has the same unit ball as the norm induced from the inner product weighted with P

$$\|x\|_A = \gamma \|x\|_P = \gamma\sqrt{\langle x, Px\rangle},$$

where $\gamma = |\alpha|^{-2}$ denotes a suitable scaling factor such that $1 = \|x_0\|_A = \gamma \|x_0\|_P$. The eccentricity of $\|\cdot\|_A$ is then given by

$$\operatorname{ecc} \|\cdot\|_A = \sqrt{\left|\tfrac{\beta}{\alpha}\right|} \quad \text{for } A = \begin{pmatrix} 0 & \beta \\ \alpha & 0 \end{pmatrix} \text{ and } \alpha\beta < 0, |\beta| \geq |\alpha|.$$

Consider now a real 2×2 marginally stable matrix of the form $A = \left(\begin{smallmatrix} 0 & \alpha \\ 0 & \lambda \end{smallmatrix}\right)$ with $\lambda < 0$. Then the Feller norm induced by A is of the form

$$\|x\|_A = \max\{\|x\|, M\,|\langle v, x\rangle|\}, \tag{4.6}$$

where $v = (\lambda^2 + \alpha^2)^{-1/2}(\lambda, -\alpha)^\top$ and some suitable constant $M > 1$, see Figure 4.3. We will determine the exact value for M by the following geometrical argument. The vector v is the left eigenvector corresponding to 0. Hence, it is orthogonal to the λ-eigenvector of A. Figure 4.3 shows the unit ball of this norm. One can easily see that A is dissipative with respect to $\|\cdot\|_A$. The eccentricity of $\|\cdot\|_A$ is then given by the inverse of the cosine of the angle spanned by the left and right eigenvalue of 0, namely $\operatorname{ecc}\|\cdot\|_A = \left|\left\langle v, \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\right\rangle\right|^{-1}$ which evaluates to

$$M = \frac{\sqrt{\lambda^2 + \alpha^2}}{|\lambda|} \quad \text{for } A = \begin{pmatrix} 0 & \alpha \\ 0 & \lambda \end{pmatrix}, \lambda < 0.$$

We have obtained the result on the transient bound $M_0(A) = \left|\left\langle v, \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\right\rangle\right|^{-1}$ already as part of Corollary 3.20.
In the following example we will derive the transient amplification $M_0(A)$ by analytical means instead of using geometrical considerations.

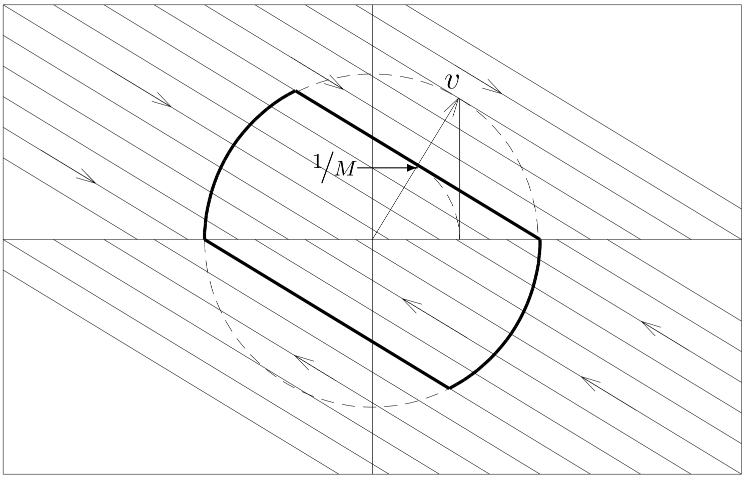

Figure 4.3: The unit ball of the transient norm for a marginally stable 2×2 matrix.

Example 4.6. Consider the matrix $A = \left(\begin{smallmatrix} 0 & k\lambda \\ 0 & -\lambda \end{smallmatrix}\right)$ for $\lambda > 0$ and $k > 0$. By Proposition 4.4 we get the following function for the spectral norm of the matrix exponential,

$$\left\|e^{At}\right\| = \tfrac{1}{2}(1 - e^{-\lambda t}) \left(\sqrt{\coth(\lambda\!/_2\, t)^2 + k^2} + \sqrt{1 + k^2}\right) \xrightarrow{t\to\infty} \sqrt{1+k^2}$$

as $\lim_{x\to\infty} \coth(x) = 1$. Moreover, this function is monotonously increasing. Thus, $M_0(A) = \sqrt{1+k^2} = \sup_{t\geq 0} \left\|e^{At}\right\|$ is the transient amplification for a marginally stable matrix of the given structure. ■

4.2.2 Exponentially Stable Matrices

Let us now study the case where $A \in \mathbb{R}^{2\times 2}$ is an exponentially stable matrix with real spectrum. Then the line segment which appears in the unit ball of the Feller norm in the marginally stable case is now given as part of a trajectory which touches the Euclidean unit circle tangentially in a point x, $\|x\| = 1$. (The other crossing point with the unit circle is traversal). In x the norm of the solution attains a local maximum. Therefore from $\frac{d}{dt}\left\|e^{At}x\right\|^2 = 0$ it follows that $x^\top(A + A^\top)x = 0$ has to hold. Following the trajectory backwards in time, it has to attain its minimum norm in y before leaving the unit ball, see Figure 4.4. For this minimum, we again have $y^\top(A + A^\top)y = 0$.
By replacing the trajectory segment between x and y by a line (dotted in Figure 4.4), we obtain an upper bound on the eccentricity of the transient norm. The points x and y can be computed explicitly, so that we obtain the following bound.

Proposition 4.7. *Suppose that $A \in \mathbb{R}^{2\times 2}$ is given and let $\|\cdot\|_A$ denote the associated transient norm. If A is exponentially stable, but not dissipative then*

$$\operatorname{ecc} \|\cdot\|_A \leq |\langle x, y\rangle|^{-1},$$

where x and y are linear independent unit vectors satisfying

$$x^\top(A+A^\top)x = 0 = y^\top(A+A^\top)y.$$

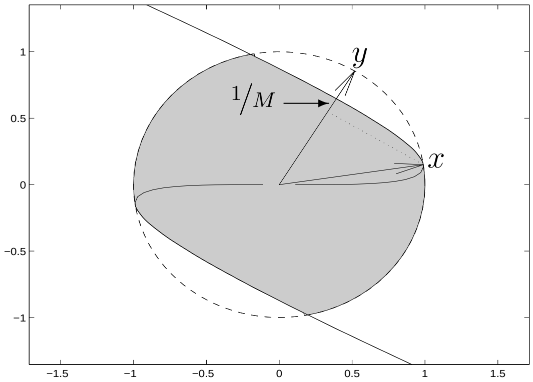

Figure 4.4: The unit ball of the Feller norm for a stable 2×2 matrix.

For an exponentially stable, but not dissipative matrix $A = \left(\begin{smallmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{smallmatrix}\right)$ we obtain by Proposition 4.7 the estimate

$$\operatorname{ecc} \|\cdot\|_A \leq \frac{\sqrt{(\lambda_1-\lambda_2)^2+\alpha^2}}{-(\lambda_1+\lambda_2)},$$

since the vectors x and y are given by normalized multiples of $(\alpha \pm \sqrt{\alpha^2 - 4\lambda_1\lambda_2}, -2\lambda_1)^\top$. Using the fact that A is dissipative if and only if $4\lambda_1\lambda_2 \geq \alpha^2$ we have for any real exponentially stable upper triangular 2×2 matrix that

$$\operatorname{ecc} \|\cdot\|_A \leq \frac{\sqrt{(\lambda_1+\lambda_2)^2 + \max\{0, \alpha^2 - 4\lambda_1\lambda_2\}}}{-(\lambda_1+\lambda_2)} \quad \text{for } A = \begin{pmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{pmatrix}, \lambda_1, \lambda_2 < 0. \tag{4.7}$$

The last case of a stable Schur form belongs to those matrices which have a pair of conjugated eigenvalues located in the left half-plane. Their real Schur form is given by $A = \left(\begin{smallmatrix} \lambda & \beta \\ \alpha & \lambda \end{smallmatrix}\right)$ where $\lambda < 0$ and $\alpha\beta < 0$. A stable matrix $A \in \mathbb{R}^{2\times 2}$ is dissipative if $A + A^\top \preceq 0$ which is equivalent to $\det(A+A^\top) \geq 0$. Hence, the Feller norm will differ from the Euclidean norm if $2\,|\lambda| \leq |\beta+\alpha|$. Carrying out the same calculations as for Proposition 4.7 (which only depend on $A + A^\top$) gives the following upper bound for the transient excursion

$$\operatorname{ecc} \|\cdot\|_A \leq \sqrt{\frac{\max\{4\lambda^2, (\alpha+\beta)^2\}}{4\lambda^2}} \quad \text{for } A = \begin{pmatrix} \lambda & \beta \\ \alpha & \lambda \end{pmatrix}, \lambda < 0,\ \alpha\beta < 0. \tag{4.8}$$

Rewriting the bound (4.8) in terms of determinants and traces of A and $A + A^\top$ we get

$$M_+ := \sqrt{1 - \frac{\det(A+A^\top)}{\operatorname{trace}^2(A)}} \quad \text{if } \det(A+A^\top) < 0, \tag{4.9}$$

which is also valid for the real case, cf. (4.7). Hence stable matrices of the form

$$\begin{pmatrix} a & \sigma b + (1-\tau)c \\ (1-\sigma)b + \tau c & d \end{pmatrix} \quad \text{for all } \tau, \sigma \in \mathbb{R}$$

have the common upper transient bound $M_+ = \sqrt{\frac{(a-d)^2+(b+c)^2}{(a+d)^2}}$ if $4ad \le (b+c)^2$.
Let us now return to matrices with real spectrum. We already know that the bound (4.7) is exact in the limiting cases $\lambda_1 = 0$ and $|\alpha| = 2\sqrt{\lambda_1\lambda_2}$. Therefore it is reasonable to ask for the quality of this bound.

Theorem 4.8. *Given a stable matrix* $A = \left(\begin{smallmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$ *with* $4\lambda_1\lambda_2 \le \alpha^2$. *Then for* $M_0 = \sup_{t\ge 0} \left\| e^{At} \right\|$ *and* $M_+ = \sqrt{(\lambda_1-\lambda_2)^2+\alpha^2}/\left|\lambda_1+\lambda_2\right|$ *the following estimate holds*

$$1 \le \frac{M_+}{M_0} \le \frac{e}{2}.$$

Proof. The norm of the matrix exponential is given by Proposition 4.4

$$\left\| e^{At} \right\| = \begin{cases} \frac{1}{2}\left| e^{\lambda_1 t} - e^{\lambda_2 t} \right| \left(\sqrt{\coth\left(\frac{\lambda_1-\lambda_2}{2}t\right)^2 + \left(\frac{\alpha}{\lambda_1-\lambda_2}\right)^2} + \sqrt{1 + \left(\frac{\alpha}{\lambda_1-\lambda_2}\right)^2} \right), \\ \frac{1}{2}e^{\lambda t}\left(\sqrt{4+(\alpha t)^2} + |\alpha|\, t \right), \qquad \text{if } \lambda = \lambda_1 = \lambda_2. \end{cases}$$

A lower bound for $\left\| e^{At} \right\|$ is given by

$$\left\| e^{At} \right\| \ge \begin{cases} \frac{|\alpha|}{\lambda_1-\lambda_2}\left(e^{\lambda_1 t} - e^{\lambda_2 t} \right), & \text{if } \lambda_1 \ne \lambda_2, \\ |\alpha|\, t e^{\lambda t}, & \text{if } \lambda = \lambda_1 = \lambda_2. \end{cases}$$

It attains a critical value at $t_0 = \frac{1}{\lambda_1-\lambda_2}\log\frac{\lambda_2}{\lambda_1}$, respectively at $t_0 = -\frac{1}{\lambda}$. Let us now concentrate on the case $\lambda_1 \ne \lambda_2$. Then

$$\left\| e^{At_0} \right\| = \frac{1}{2}\left((-\lambda_2)^{\lambda_2}(-\lambda_1)^{-\lambda_1} \right)^{1/(\lambda_1-\lambda_2)} \left(\sqrt{(\lambda_1+\lambda_2)^2+\alpha^2} + \sqrt{(\lambda_1-\lambda_2)^2+\alpha^2} \right).$$

Hence $M_0 \ge \max(1, \left\| e^{At_0} \right\|)$. Now, $M_+/M_0 \le M_+/\left\| e^{At_0} \right\|$ and for this quotient,

$$\begin{aligned} \frac{M_+}{\left\| e^{At_0} \right\|} &= 2\left(\left(1+\frac{\lambda_2}{\lambda_1}\right)\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{\lambda_2}{\lambda_1-\lambda_2}} \left(\sqrt{\frac{(\lambda_1+\lambda_2)^2+\alpha^2}{(\lambda_1-\lambda_2)^2+\alpha^2}} + 1 \right) \right)^{-1} \\ &\le \left(\left(1+\frac{\lambda_2}{\lambda_1}\right)\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{\lambda_2}{\lambda_1-\lambda_2}} \right)^{-1} = \left(\left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{\lambda_2}{\lambda_1-\lambda_2}} + \left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{\lambda_1}{\lambda_1-\lambda_2}} \right)^{-1}. \end{aligned} \tag{4.10}$$

To complete the proof, let us use the following inequality. For $a, b \in \mathbb{R}$, $ab > 0$, $a \ne b$ we have

$$\left(\frac{b}{a}\right)^{\frac{b}{a-b}} + \left(\frac{b}{a}\right)^{\frac{a}{a-b}} \ge \frac{2}{e}. \tag{4.11}$$

To prove this, we assume $b > a > 0$ and set $b = a + \varepsilon$, $\beta = \frac{\varepsilon}{a}$. Then the LHS of (4.11) gives

$$\left(\frac{b}{a}\right)^{\frac{a}{a-b}}\left(\frac{a}{b}+1\right) \geq 2\left(1+\frac{\varepsilon}{a}\right)^{-\frac{a}{\varepsilon}} = 2\left(1+\frac{1}{\beta}\right)^{-\beta} \geq 2\exp(-1).$$

Now using (4.11) in (4.10) shows that $M_+/M_0 \leq e/2$ when $\lambda_1 \neq \lambda_2$.
For the remaining case $\lambda = \lambda_1 = \lambda_2$ we have

$$M_0 \geq \left\|e^{At_0}\right\| = \tfrac{1}{2e}\left(\sqrt{4 + (\tfrac{\alpha}{\lambda})^2} + \left|\tfrac{\alpha}{\lambda}\right|\right).$$

As $M_+ = -\left|\alpha\right|(2\lambda)^{-1}$,

$$\frac{M_+}{M_0} \leq \frac{\left|\alpha\right| e}{\sqrt{(2\lambda)^2 + \alpha^2} + \left|\alpha\right|} \leq \frac{e}{2}.$$

Hence, the bound $\frac{M_+}{M_0} \leq \frac{e}{2}$ holds for all stable upper triangular matrices. □

Hence the bound M_+ of (4.9) satisfies $M \leq M_+ \leq 1.36M$.
We have the following generalization to higher dimensions.

Corollary 4.9. *Suppose that $B \in \mathbb{R}^{n\times m}$ and $\alpha, \beta < 0$. Then for*

$$A = \begin{pmatrix} \alpha I_n & B \\ 0 & \beta I_m \end{pmatrix}: \qquad \sup_{t\geq 0}\left\|e^{At}\right\| \leq \max\left\{-\frac{\sqrt{(\alpha-\beta)^2 + \left\|B\right\|^2}}{\alpha+\beta}, 1\right\}.$$

Proof. The spectral norm of the matrix $\left(\begin{smallmatrix} \alpha & B \\ 0 & \beta I \end{smallmatrix}\right)$ is given by

$$\left\|\begin{pmatrix} \alpha I & B \\ 0 & \beta I \end{pmatrix}\right\| = \frac{1}{2}\left(\sqrt{(\alpha+\beta)^2 + \left\|B\right\|^2} + \sqrt{(\alpha-\beta)^2 + \left\|B\right\|^2}\right)$$

as we have seen in Lemma 4.1. The matrix exponential retains the blockdiagonal structure so everything works out as in the two-dimensional case. □

4.3 Liapunov Matrices of Minimal Condition Number

In this section we continue to study solutions of the Liapunov equation

$$\mathcal{L}_A(P) := PA + A^\top P = -Q \preceq 0, \tag{4.12}$$

where we assume that $A \in \mathbb{R}^{2\times 2}$ is an exponentially stable matrix. As we have seen in Theorem 3.35, the condition number $\kappa(P)$ of a solution $P \in \mathcal{H}^2$ of (4.12) measures the eccentricity of the quadratic Liapunov function $x \mapsto x^\top P x$ compared to $x \mapsto x^\top x$ and therefore gives rise to an upper bound of the matrix exponential. Again, as in Section 3.4 we are interested in a solution P for which the spectral condition number

$$\kappa(P) = \left\|P\right\|\left\|P^{-1}\right\| = \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}, \qquad P = -\mathcal{L}_A^{-1}(Q),$$

attains a minimum under all positive semidefinite $Q \succeq 0$. If the matrix A is dissipative with respect to the spectral norm, $A + A^\top \preceq 0$, then $P = I$ satisfies (4.12). Therefore the optimal condition number in case of a dissipative matrix is $\kappa_{opt} = 1$. So let us now study stable matrices $A \in \mathbb{R}^{2\times 2}$ for which $A + A^\top$ is indefinite.
For a 2×2 regular triangular real matrix A the optimal solution for a real upper triangular matrix $A = \left(\begin{smallmatrix}\lambda_1 & \mu \\ 0 & \lambda_2\end{smallmatrix}\right)$ where $\lambda_i < 0$, $\lambda_1 \neq \lambda_2$ and $\mu \in \mathbb{R}$, may be found by direct computation. Then for $A + A^\top$ being indefinite, Q is given by a rank 1 matrix.

Proposition 4.10. *Let $A = \left(\begin{smallmatrix}\lambda_1 & \mu \\ 0 & \lambda_2\end{smallmatrix}\right) \in \mathbb{R}^2$, be an exponentially stable matrix with $\mu^2 \geq 4\lambda_1\lambda_2$, $\lambda_1 \neq \lambda_2$. Then $\kappa(P)$ is minimal under all solutions $\mathfrak{L}_A(P) = -Q$ of* (4.12) *with $P \succ 0$, $Q \succeq 0$ if and only if Q satisfies $Q = cc^\top$ where c is given by $c = (\lambda_1 - \lambda_2, \mu - \nu)^\top$ (or multiples thereof) with $\nu = \operatorname{sgn}(\mu)\sqrt{\frac{\lambda_2}{\lambda_1}((\lambda_1 - \lambda_2)^2 + \mu^2})$.*

Proof. As a consequence of Proposition 3.43 the optimal solution is found under those Hermitian pairs (P, Q) for which Q is only semidefinite which gives in the 2×2 case a matrix of rank one. Hence we are looking for a right hand side $Q = cc^\top$, $c \in \mathbb{R}^2$, of the Liapunov equation. These matrices can be conveniently parameterized by setting

$$Q(\theta) = \begin{pmatrix}(\lambda_1 + \lambda_2)^2 & (\lambda_1 + \lambda_2)\theta \\ (\lambda_1 + \lambda_2)\theta & \theta^2\end{pmatrix}, \qquad c = (\lambda_1 + \lambda_2, \theta)^\top. \tag{4.13}$$

Then the Liapunov matrix $P(\theta) = \left(\begin{smallmatrix}p_1 & p_3 \\ p_3 & p_2\end{smallmatrix}\right) = -\mathfrak{L}_A^{-1}(Q(\theta))$ is given by the components

$$p_1 = -\frac{(\lambda_1 + \lambda_2)^2}{2\lambda_1}, \quad p_3 = \frac{\lambda_1 + \lambda_2}{2\lambda_1}\mu - \theta, \quad p_2 = -\frac{1}{2\lambda_2}\left(\theta^2 - 2\mu\theta + \frac{\lambda_1 + \lambda_2}{\lambda_1}\mu^2\right).$$

Now, the spectral condition number of a symmetric 2×2 matrix satisfies $\kappa(P) = \frac{\|P\|}{\operatorname{trace} P - \|P\|}$ and we have

$$\|P\| = \frac{p_1 + p_2}{2} + \frac{1}{2}\sqrt{(p_1 - p_2)^2 + 4p_3^2}, \quad \det(P) = p_1p_2 - p_3^2, \quad \operatorname{trace} P = p_1 + p_2.$$

For critical values, $\frac{d}{d\theta}\kappa(P(\theta)) = 0$ has to hold, hence we are searching for solutions of

$$(\operatorname{trace} P(\theta) - \|P\|)\tfrac{d}{d\theta}\|P(\theta)\| = \|P(\theta)\|\tfrac{d}{d\theta}(\operatorname{trace} P(\theta) - \|P\|). \tag{4.14}$$

Now, setting $\|P(\theta)\| = \frac{1}{2}(\operatorname{trace} P(\theta) + \sqrt{\Delta(\theta)})$ with $\Delta(\theta) = (p_1 - p_2)^2 + 4p_3^2$, equation (4.14) simplifies to

$$\sqrt{\Delta(\theta)}\tfrac{d}{d\theta}\operatorname{trace} P(\theta) = \operatorname{trace} P(\theta)\tfrac{d}{d\theta}\sqrt{\Delta(\theta)}.$$

The chain rule for $\frac{d}{d\theta}\sqrt{\Delta(\theta)}$ gives

$$2\Delta(\theta)\tfrac{d}{d\theta}\operatorname{trace} P(\theta) = \operatorname{trace} P(\theta)\tfrac{d}{d\theta}\Delta(\theta). \tag{4.15}$$

The following critical points of (4.15) can be found with the help of a computer algebra system

$$\theta_1 = \frac{\lambda_1 + \lambda_2}{\lambda_1 - \lambda_2}\,\mu, \qquad \theta_{2,3} = \frac{\lambda_1 + \lambda_2}{\lambda_1 - \lambda_2}\left(\mu \pm \sqrt{\tfrac{\lambda_2}{\lambda_1}\left((\lambda_1 - \lambda_2)^2 + \mu^2\right)}\right).$$

The value θ_1 corresponds to a $+\infty$-pole of κ as $c = (\lambda_1 + \lambda_2, \theta_1)^\top$ is a left eigenvector of A, hence (A, c) is not observable and therefore $\mathcal{L}_A^{-1}(-cc^\top)$ is not positive definite. Both other critical values give rise to a local minimum. The global minimum is attained when choosing $\theta_i, i \in \{2, 3\}$ to be of smallest modulus. Then c of (4.13) is given by

$$c = \left(\lambda_1 + \lambda_2, \frac{\lambda_1 + \lambda_2}{\lambda_1 - \lambda_2}(\mu - \nu)\right)^\top, \quad \text{where} \quad \nu = \operatorname{sgn}(\mu)\sqrt{\tfrac{\lambda_2}{\lambda_1}\left((\lambda_1 - \lambda)^2 + \mu^2\right)}.$$

By scaling c by the factor $\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}$ we obtain the required result. □

The following considerations are helpful in understanding the result of Proposition 4.10. Let us only consider a subset of the cone of positive semidefinite symmetric 2×2 matrices $\mathcal{H}_+^2$ given by positive semidefinite matrices H with $\operatorname{trace} H = 2$. Clearly, this set is a basis of the cone. Then each H from this set can be written as $H = \left(\begin{smallmatrix} 1+\alpha & \beta \\ \beta & 1-\alpha \end{smallmatrix}\right)$ for $\alpha^2 + \beta^2 \leq 1$. The semidefinite matrices are given by $\alpha^2 + \beta^2 = 1$, and the positive definite ones by $\alpha^2 + \beta^2 < 1$. In the equivalence class $\mathcal{H}_+^2/\{\text{trace} = 2\}$, addition and inversion follow the rules

$$(\alpha, \beta) + (\gamma, \delta) = \left(\tfrac{\alpha+\gamma}{2}, \tfrac{\beta+\delta}{2}\right) \quad \text{and} \quad (\alpha, \beta)^{-1} = (-\alpha, -\beta),$$

where we identify the pair (α, β) with the matrix $\left(\begin{smallmatrix} 1+\alpha & \beta \\ \beta & 1-\alpha \end{smallmatrix}\right)$. The visualization of the image of the positive definite cone under the inverse Liapunov operator $\mathcal{L}_A^{-1}$ is accomplished easily. We will call the set $\{H \in \mathcal{H}_+^2(\mathbb{R}) \mid \operatorname{trace}(H) = 2,\ HA + A^\top H \preceq 0\}$ the *Liapunov cone* of A. Figure 4.5 shows the Liapunov cones for $A = \left(\begin{smallmatrix} -5 & 36 \\ 0 & -20 \end{smallmatrix}\right)$ and $A = \left(\begin{smallmatrix} -2 & -1 \\ 9 & -2 \end{smallmatrix}\right)$, respectively. Neither cone includes the origin whence $\mu(A + A^\top) > 0$ for both matrices, i.e., they do not generate contractions. The asterisks in Figure 4.5 mark the positions of the Hermitian matrices P of smallest condition. They are oriented towards the center of

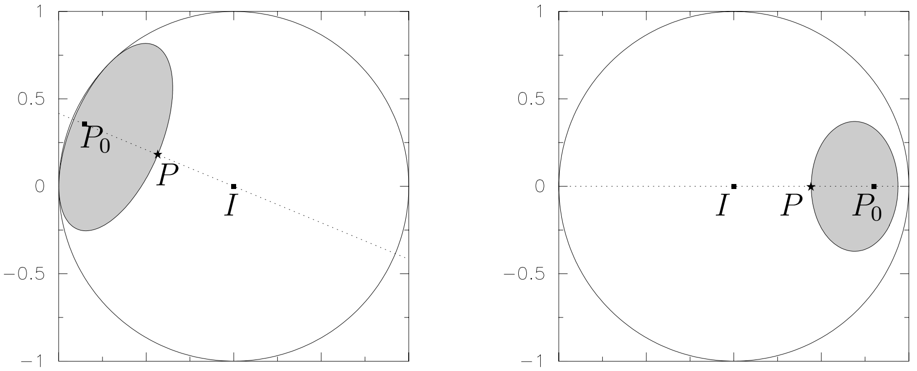

Figure 4.5: Real Liapunov cones.

the Hermitian cone because the spectral condition number κ depends monotonically on the radius $r = \sqrt{\alpha^2 + \beta^2}$. Moreover, the points of the Liapunov cone touching the outer circle of semidefinite matrices correspond to left eigenvectors v_i of A which form symmetric eigenvalues $v_i v_i^\top$ of the inverse Liapunov operator. The matrix A for the right picture of Figure 4.5 contains only complex eigenvalues and therefore the Liapunov cone does not touch the outer boundary given by real semidefinite matrices. By inspection, we see that the sum of the symmetric eigenvalues of $\mathcal{L}_A^{-1}$, i.e., the midpoint between the tangent points (given by symmetric eigenvectors of $\mathcal{L}_A^{-1}$) which is marked by a box, is aligned with the optimally conditioned matrix and the center of the cone. Here the center is identified with the identity matrix I_2. Hence we obtain the following alternative way of obtaining the formula of Proposition 4.10.

Corollary 4.11. *Let $A \in \mathbb{R}^{2\times 2}$ be a stable matrix. If $v_i, i = 1, 2,$ are the left eigenvectors of A corresponding to the eigenvalues λ_i then by setting $P_0 = v_1 v_1^\top + v_2 v_2^\top$ we obtain the quadratic Liapunov matrix of minimal condition by $P = P_0 + \lambda_0 I$ where $\lambda_0 = \min\{\lambda \in \sigma(Q_0, A{+}A^\top) \,|\, \lambda > 0\}$ is the smallest positive eigenvalue of the matrix pencil $Q_0{-}\lambda(A{+}A^\top)$ and $Q_0 = -(P_0 A + A^\top P_0) = -2\mathrm{Re}\,(\lambda_1 v_1 v_1^\top + \bar{\lambda}_2 v_2 v_2^\top)$.*

Example 4.12. Consider the matrix $A = \left(\begin{smallmatrix} -5 & 36 \\ 0 & -20 \end{smallmatrix}\right)$. By Proposition 4.10 the quadratic Liapunov matrix solution of minimal eccentricity satisfies $Q = cc^\top$ where $c = (15, -42)$ as $\nu = 2\sqrt{15^2 + 36^2} = 78$. The RHS Q of the Liapunov equation is given by $Q = \left(\begin{smallmatrix} 625 & -25{\cdot}70 \\ -25{\cdot}70 & 4900 \end{smallmatrix}\right)$ with $\theta = 70$ in (4.13). Therefore $P = \left(\begin{smallmatrix} 62.5 & 20 \\ 20 & 158.5 \end{smallmatrix}\right)$ and its quadratic condition number is optimal and given by $\kappa(P) = \frac{162.5}{58.5} = \frac{25}{9}$. As the associated eccentricity is given by the square root of $\kappa(P)$ we obtain the growth bound $\left\|e^{At}\right\|_2 \leq 5/3$. Some trajectories of the system and an optimal ellipse $\{x \in \mathbb{R}^2 \,|\, \langle x, Px\rangle_2 = \mathit{const}\}$ are depicted in Figure 3.6, cf. Example 3.38.

For the second approach in Corollary 4.11, we set $P_0 = \left(\begin{smallmatrix} 25 & 60 \\ 60 & 313 \end{smallmatrix}\right) = 13^2(v_1 v_1^\top + v_2 v_2^\top)$ where v_i are left eigenvectors of A, $v_1 = \binom{0}{1}, v_2 = \frac{1}{13}\binom{5}{12}$. Then the generalized eigenvalues are given by

$$\sigma(Q_0, A + A^\top) = \sigma\left(\left(\begin{smallmatrix} 250 & 600 \\ 600 & 8200 \end{smallmatrix}\right), \left(\begin{smallmatrix} -10 & 36 \\ 36 & -40 \end{smallmatrix}\right)\right) \approx \{-11.6071, 162.5\}.$$

Now $\lambda_0 = 162.5$, and $P = P_0 + \lambda_0 I = \left(\begin{smallmatrix} 187.5 & 60 \\ 60 & 475.5 \end{smallmatrix}\right)$ which differs from the previously obtained value by the scalar factor 3. Thus this second method leads to the same result as the formula given in Proposition 4.10 with $\kappa(P) \approx 2.77778$. ■

4.3.1 Common Quadratic Liapunov Matrices

The update step for quadratic Liapunov equations described in Proposition 3.43 can also be used to obtain a common Liapunov matrix for two stable 2 × 2 matrices A_0 and A_1. Let us suppose that the Liapunov cones of A_0 and A_1 have a non-empty intersection, hence there exist common Liapunov matrices for A_0 and A_1. Like in Corollary 4.11 let P_i denote the "eigenvector mean" of A_i, $i = 0, 1$ where $P_i = v_1(A_i)v_1^*(A_i) + v_2(A_i)v_2^*(A_i)$ and $v_j(A_i)$ are the normed left eigenvectors of A_i. These matrices are located near the centers of the corresponding Liapunov cones. If a part of the line segment between P_1

and P_2 in $\mathcal{H}^2/\{\text{trace} = 2\}$ is contained in the intersection of the Liapunov cones, we can detect this using the update mechanism of Proposition 3.43. By construction, both $Q_0 := -(P_0A_0 + A_0^\top P_0)$ and $Q_1 := -(P_1A_1 + A_1^\top P_1)$ are positive semidefinite. Now let us set $R_i = P_iA_{1-i} + A_{1-i}^\top P_i$, $i = 0, 1$. If there exists a negative definite R_i then we have found the common Liapunov matrix P_i. If R_i is indefinite we construct an interval of Liapunov matrices

$$\tilde{P}_i = P_i + \lambda_i P_{1-i}, \quad \lambda_i \in [0, \min\{\sigma(Q_i, R_i) \cap \mathbb{R}_+\}], \quad i = 0, 1,$$

using Lemma 3.44. If these intervals overlap (with respect to $\mathcal{H}^2/\{\text{trace} = 2\}$) then all matrices from the intersection of the intervals are common quadratic Liapunov functions for A_0 and A_1. In particular, the intersection of the intervals is non-empty if

$$\lambda_0^*\lambda_1^* \geq 1 \qquad \text{for} \quad \lambda_i^* = \min\{\sigma(Q_i, R_i) \cap \mathbb{R}_+\}, \quad i = 0, 1. \tag{4.16}$$

To this end, note that in $\mathcal{H}_+^2/\{\text{trace} = 2\}$ the Hermitian matrices $P_0^* = P_0 + \lambda_0^* P_1$ and $P_1^* = P_1 + \lambda_1^* P_0$ are contained in the interval $[P_0, P_1]$. Hence we can decide whether the Liapunov cones generated by A_0 and A_1 have a non-empty intersection along this interval by checking if $P_1^* \in [P_0, P_0^*]$. By this is equivalent to $\frac{1}{\lambda_1^*} \leq \lambda_0^*$, hence (4.16) holds. Unfortunately, this is not a necessary condition for the existence of common quadratic Liapunov matrices.

Example 4.13. Consider the matrices $A_0 = \left(\begin{smallmatrix} 0 & 10 \\ -1 & -20 \end{smallmatrix}\right)$ and $A_1 = \left(\begin{smallmatrix} 0 & 2 \\ -16 & -7 \end{smallmatrix}\right)$. We obtain $P_0 = \left(\begin{smallmatrix} 0.7942 & 0.4574 \\ 0.4574 & 1.206 \end{smallmatrix}\right)$ and $P_1 = \left(\begin{smallmatrix} 1.778 & 0.3889 \\ 0.3889 & 0.2222 \end{smallmatrix}\right)$. Then $\lambda_0^* = 1.113$ and $\lambda_1^* = 0.5451$ and $\lambda_0^*\lambda_1^* = 0.6069$ so that (4.16) is not satisfied. However, the left image of Figure 4.6 shows that both Liapunov cones have a non-empty intersection. For the matrix pair A_0 and $A_2 = \left(\begin{smallmatrix} -3 & 5 \\ -1 & -1 \end{smallmatrix}\right)$

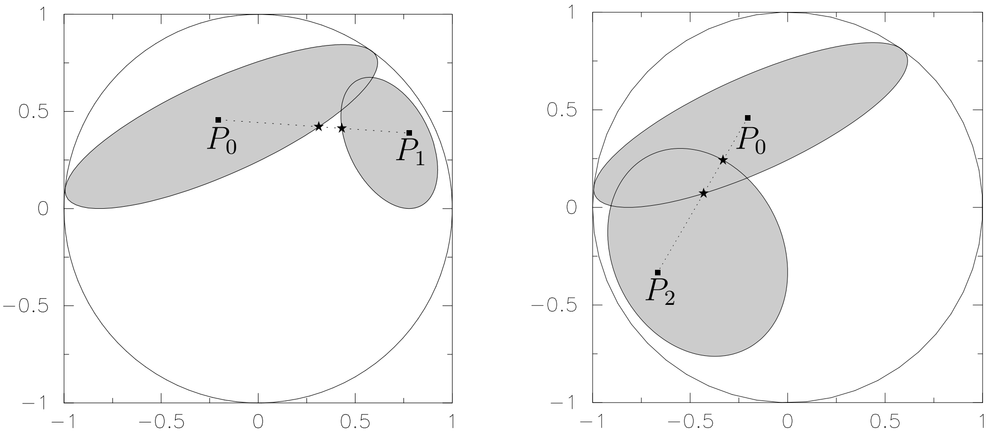

Figure 4.6: Common quadratic Liapunov matrices: Intersecting cones.

we have $P_2 = \frac{1}{3}\left(\begin{smallmatrix} 1 & -1 \\ -1 & 5 \end{smallmatrix}\right)$ and $\lambda_0^* = 0.9521$, $\lambda_2^* = 2.672$, hence $\lambda_1^*\lambda_2^* = 2.544 > 1$. Indeed, the right image of Figure 4.6 shows that the Liapunov cones intersect. Now, every $P_0^* + \lambda P_2^*$ with $\lambda \in [(\lambda_2^*)^{-1}, \lambda_0^*]$ is a common Liapunov matrix for A_0 and A_2. ■

This method of constructing segments of common Liapunov matrices is not restricted to matrices which are used in Corollary 4.11 or to dimension 2. We therefore immediately obtain the following theorem.

Theorem 4.14. *Let $A_0, A_1 \in \mathbb{K}^{n\times n}$ be exponentially stable matrices, and $P_0, P_1 \in \mathcal{H}^n_+(\mathbb{K})$ such that $\mathcal{L}_{A_i}(P_i) \preceq 0$, $i = 0, 1$. We define $R_0 = P_0A_1 + A_1^*P_0$ and $R_1 = P_1A_0 + A_0P_1$. If R_0 or R_1 is negative definite then P_0, respectively P_1, is a common Liapunov matrix of A_1 and A_2. Otherwise consider*

$$\lambda_i^* = \min\{\sigma(Q_i, R_i) \cap \mathbb{R}_+\}, \qquad i = 0, 1.$$

If $\lambda_0^*\lambda_1^* \geq 1$ then all positive linear combinations of P_0 and P_1 which satisfy

$$\theta(P_0 + \lambda P_1), \qquad \theta > 0,\ \lambda \in \left[(\lambda_1^*)^{-1},\ \lambda_0^*\right]$$

are common quadratic Liapunov functions of A_0 and A_1.

4.4 Dissipativity for Polytopic Norms

We close this example section by a discussion of dissipativity for the class of polytopic norms.

Definition 4.15. A point $x \in \mathbb{R}^n$ of a closed convex set $K \subset \mathbb{R}^n$ is called an *extremal point* if for all $a, b \in K\setminus\{x\}$ the point x is not contained in the interval $(a, b) = \{\tau a+(1-\tau)b \,|\, \tau \in (0, 1)\} \subset K$. A norm $\|\cdot\|$ in $\mathbb{K}^n$ is a *polytopic norm* if its unit ball $\mathbb{B} = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ has only a finite set of extremal points.

With every polytopic norm we associate the set $C \subset \mathbb{R}^n$ of extremal points of $\mathbb{B}$. Given a set of points $C \subset \mathbb{R}^n$ such that $\mathbb{B} = \operatorname{conv} C$ is

- balanced, i.e., $x \in \mathbb{B}$ implies $-x \in \mathbb{B}$ (hence $C = -C$),
- absorbing, i.e., for all $x \in \mathbb{R}^n$ there exists $\alpha > 0$ with $\alpha x \in \mathbb{B}$ (hence $\operatorname{span} C = \mathbb{R}^n$),
- its set of extremal points is given by C,

then $\mathbb{B}$ is the unit ball of a polytopic norm which we denote by $\|\cdot\|_C$. For a polytopic norm $\|\cdot\|_C$ the dual norm $\|\cdot\|_C^*$ is also polytopic, as $\|y\|_C^* = \max_{x\in C}\{|\langle x, y\rangle_2|\}$. In particular, the set C^* of extremal points of the dual norm is constructed from normals to the faces of $\mathbb{B}$. Hence $\|\cdot\|_C^* = \|\cdot\|_{C^*}$.
Now, for polytopic norms dissipativity needs only to be tested for pairs of extremal points.

Lemma 4.16. *Suppose $\|\cdot\|_C$ is a polytopic norm with vertex set C. If for all dual pairs (x_i, y_j) with $x_i \in C$ and $y_j \in C^*$ the inequality $y_j^\top Ax_i < 0$ holds, then A is strictly dissipative.*

Proof. Suppose that $(x, y_1), (x, y_2)$ are dual pairs of $\|\cdot\|_C$ with $x \in C$ and $y_1, y_2 \in C^*$. Then $y_0 = \lambda y_1 + (1-\lambda) y_2$ is also a dual vector of x and $y_0^\top Ax = \lambda y_1^\top Ax + (1-\lambda) y_2^\top Ax < 0$. By induction over the set of dual vectors, the solutions of $\dot{x} = Ax$ are strictly decaying for every initial value in C. Now consider a face of $\{x \in \mathbb{R}^n \mid \|x\| = 1\}$ given by its normal vector $y \in C^*$. Then all adjacent corners $x_i \in C$ form dual pairs (x_i, y). Any convex combination $x = \sum_i \alpha_i x_i, \sum_i \alpha_i = 1$ also defines a dual pair (x, y) which satisfies $y^\top Ax < 0$. As all possible dual pairs have this structure, A is dissipative. □

Especially, the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are polytopic, so that the result of Lemma 4.16 is also applicable to them. As a consequence from Theorem 2.74, for an exponentially stable matrix $A \in \mathbb{K}^{n\times n}$ one always finds a polytopic norm which is also a strict Liapunov norm. Now every unit ball $\mathbb{B}$ of a norm can be approximated via a polytopic norm ball by choosing a set of points C on the unit sphere $\partial\mathbb{B}$ which respects the above-motioned requirements. This gives an inner approximation $\mathbb{B}_C \subset \mathbb{B}$. The dual polytopic unit ball then becomes an outer approximation of the original dual norm, $\mathbb{B}_C^* \supset \mathbb{B}^*$.

Given a matrix $A \in \mathbb{R}^{n\times n}$ one would like to conclude from the dissipativity of A with respect to the polytopic norm that A is also dissipative with respect to the original norm, if only the approximation of these two norms is good enough. This problem is still unsolved.

4.5 Notes and References

Explicit formulas for the matrix exponential of 2 × 2 matrices can be found in Engel and Nagel [38, Example I.2.7 (iii)]. However, computing a closed formula for the norm is not carried out in that work.

Exponential bounds based on the Feller norm are to the best of the author's knowledge currently not available in the literature. The problem of determining a quadratic Liapunov norm of minimal eccentricity has been addressed in Khusainov, Komarov and Yun'kova [82, 85] and Sarybekov [123]. Obolenskii [111] introduces a different condition number $\kappa'(A) = \frac{\operatorname{trace}(A)}{n\sqrt[n]{\det(A)}}$, and shows the existence and uniqueness of an optimal solution which respect to this new condition number.

The visualization of 2×2 Liapunov cones has been used in [123] and Cohen and Lewkovicz [27]. The construction of common quadratic Liapunov matrices is an active area of research, see Ando [2] and Mason and Shorten [106].

Chapter 5

Positive Systems Techniques

A dynamical system is said to be positive if the positive orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \,|\, x_i \geq 0\}$ is invariant under its flow. This invariance is crucial for the property that the state space of positive systems and of related system-theoretic concepts (like Liapunov functions) can be restricted to the positive orthant. Positive systems are often encountered in applications when positivity constraints are given, i.e., modelling populations and concentrations.

A linear system $\dot{x} = Ax, A \in \mathbb{R}^{n\times n}$, is positive if and only if the off-diagonal entries of A are all nonnegative [40], such matrices are called *Metzler matrices*. The matrix exponential of a Metzler matrix is a nonnegative matrix. One can expect that Metzler matrices exhibit the worst transient behaviour of all stable matrices, as no cancellation of terms can occur in the formation of the matrix exponential. In this chapter we will shed some more light on the transient behaviour of Metzler matrices and their use to derive bounds for arbitrary matrices. We first study the properties of Metzler matrices, and derive transient bounds for linear positive systems. To this end, we introduce the concept of a Liapunov vector. Each Liapunov vector induces a Liapunov norm. We then answer the question how to optimally choose the Liapunov vector in order to minimize the eccentricity of the induced Liapunov norm. The next section is devoted to the study of common Liapunov vectors for a set of Metzler matrices. And finally, we show that the bounds for positive systems may also be applied to general systems.

5.1 Properties of Metzler Matrices

In this chapter we will use the following notions. A matrix $A \in \mathbb{R}^{n\times n}$ is said to be *nonnegative*, $A \geq 0$, if all of its entries are nonnegative. If all of its entries are positive, it is called *strictly positive*. Sometimes we speak of *positive* matrices, which are nonnegative and nonzero. The set of all nonnegative matrices is denoted by $\mathbb{R}^{n\times n}_+$. For $A, B \in \mathbb{R}^{n\times n}$ we write $A \geq B$ if $A - B \geq 0$. The modulus $|A| \in \mathbb{R}^{n\times n}$ of $A \in \mathbb{K}^{n\times n}$ is the componentwise modulus, $|A|_{ij} = |a_ij|$. Let us recall that $\rho(A) = \max\{|\lambda| \,|\, \lambda \in \sigma(A)\}$ denotes the spectral radius while $\alpha(A) = \max\{\operatorname{Re}\lambda \,|\, \lambda \in \sigma(A)\}$ denotes the spectral abscissa. The spectral radius satisfies the following monotonicity property, see Horn and Johnson [71],

$$\text{for all } A \in \mathbb{K}^{n\times n}, B \in \mathbb{R}_+^{n\times n} : \quad |A| \leq B \;\Rightarrow\; \rho(A) \leq \rho(|A|) \leq \rho(B). \tag{5.1}$$

A matrix $M \in \mathbb{R}^{n\times n}$ is called a *Metzler matrix* if there exists a scalar shift $\nu \in \mathbb{R}$ such that $\nu I + M \geq 0$, i.e., all off-diagonal entries are nonnegative. As a consequence, results from the Perron-Frobenius theory of positive matrices are applicable to Metzler matrices. The set of all Metzler matrices is denoted by $\mathbb{R}_{\mathrm{M}}^{n\times n}$.

Proposition 5.1 ([68]). *Suppose that $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ is a Metzler matrix. Then*

1. *$\alpha(A)$ is an eigenvalue of A and there exists a nonnegative eigenvector $x \geq 0, x \neq 0$, (called* Perron vector*) such that $Ax = \alpha(A)x$. If $A \geq 0$ then $\alpha(A) = \rho(A) \geq 0$.*
2. *If $\lambda \neq \alpha(A)$ is any other eigenvalue of A then* $\operatorname{Re}\lambda < \alpha(A)$.
3. *Given $\beta \in \mathbb{R}$ there exists a nonzero vector $x \geq 0$ such that $Ax \geq \beta x$ if and only if $\alpha(A) \geq \beta$.*
4. *$(tI - A)^{-1}$ exists and is nonnegative if and only if $t > \alpha(A)$. Moreover,*

$$\alpha(A) < t_1 \leq t_2 \implies 0 \leq (t_2 I - A)^{-1} \leq (t_2 I - A)^{-1}.$$

5. *The matrix exponential $\mathrm{e}^{At} \in \mathbb{R}_+^{n\times n}$ is nonnegative for all $t \geq 0$.*

A matrix $A \in \mathbb{R}^{n\times n}$ is called *resolvent positive* if $(tI - A)^{-1}$ exists and is nonnegative for all $t > \alpha(A)$. The last item of Proposition 5.1 shows that every Metzler matrix is resolvent positive. In fact, $A \in \mathbb{R}^{n\times n}$ is a Metzler matrix if and only if it is resolvent positive [43]. If we additionally assume in Proposition 5.1 that A is an irreducible Metzler matrix then we obtain some strict inequalities. Here A is called *reducible*, if there exists a permutation matrix P such that A is transformed into upper block-triangular form, $P^{-1}AP = \left(\begin{smallmatrix} A_1 & A_2 \\ 0 & A_3 \end{smallmatrix}\right)$. If A is not reducible, then A is called irreducible.

Corollary 5.2. *Suppose that $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ is an irreducible Metzler matrix. Then*

1. *The Perron vector $x > 0$ is strictly positive.*
2. *$(tI - A)^{-1}$ exists and is strictly positive if and only if $t > \alpha(A)$.*

The relation of a positive system $\dot{x} = Ax$, $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ on $\mathbb{R}^n$ to its restriction on the positive orthant $\mathbb{R}_+^n$ is of key importance for this chapter. Given two initial vectors x_0 and x_1 in $\mathbb{R}^n$ with $x_0 \leq x_1$, the associated solutions of the differential equation $\dot{x} = Ax$, $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$, satisfy the *monotonicity property* $x(t, x_0) \leq x(t, x_1)$ for all $t \geq 0$. In particular, $-|x_0| \leq x_0 \leq |x_0|$ holds so that

$$x(t, -|x_0|) \leq x(t, x_0) \leq x(t, |x_0|), \qquad t \geq 0.$$

Thus $|x(t, x_0)| \leq x(t, |x_0|)$ for all $x_0 \in \mathbb{R}^n$, so that we can restrict the state space of positive systems to the positive orthant $\mathbb{R}_+^n$ when looking for transient estimates.

5.2 Transient Bounds for Metzler Matrices

In this section we investigate how to obtain bounds for the transient effects of positive systems. Let us first gather some ideas based on the following monotonicity property for positive systems.

Lemma 5.3. *If $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ is a Metzler matrix and $B \in \mathbb{R}^{n\times n}_{+}$ is a nonnegative matrix then*

$$e^{At} \leq e^{(A+B)t} \qquad \textit{for } t \geq 0.$$

Proof. As B is nonnegative and A is Metzler there exists a shift $\alpha \in \mathbb{R}$ such that $0 \leq A + \alpha I \leq A + \alpha I + B$. Then all powers also satisfy $(A + \alpha I)^k \leq (A + \alpha I + B)^k$, $k \in \mathbb{N}$, hence it also holds for the matrix exponential that $\exp(A + \alpha I)t \leq \exp(A + \alpha I + B)t$, $t \geq 0$. Dividing by $e^{\alpha t}$ gives the required result. □

Now, if $\|\cdot\|$ is a monotone vector norm on $\mathbb{R}^n$ then for Metzler matrices A and B with $A \leq B$, we have $0 \leq e^{At} \leq e^{Bt}$ for all $t \geq 0$ and the induced operator norm satisfies $\|e^{At}\| \leq \|e^{Bt}\|$, see Lemma 1.9. If we find an easily obtainable transient estimate for e^{Bt} then this bound is also valid for e^{At}. Such a transient bound is relatively easy to obtain for a *Toeplitz matrix* $B = (b_{ij}) \in \mathbb{R}^{n\times n}$ which is constant along all diagonals, i.e., $b_{ij} = b_{j-i}$ for $i, j = 1, \ldots, n$. We will demonstrate this in the following example.

Example 5.4. Let B denote the $n\times n$ *Ostrowski matrix* associated with the eigenvalue $\lambda \in \mathbb{R}$,

$$B = \begin{pmatrix} \lambda & 1 & \ldots & 1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & \lambda \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

As B is a triangular Toeplitz matrix, its matrix exponential $T(t) = e^{Bt}, t \geq 0$ is also a triangular Toeplitz matrix. Moreover, it is also nonnegative by Proposition 5.1. We now need a cheap method of estimating $\|T(t)\|_2$. To this end, recall Lemma 3.58 from which we conclude that $\|T(t)\| = \alpha \begin{pmatrix} 0 & T(t) \\ T(t)^\top & 0 \end{pmatrix}$. An estimate of this eigenvalue can be obtained via Gershgorin's Theorem 2.45, which implies $\|T(t)\|_2 \leq \max\{\|T(t)\|_1, \|T(t)\|_\infty\}$. Since $T(t)$ is Toeplitz we have $\|T(t)\|_1 = \|T(t)\|_\infty$ so that the spectral norm is bounded by $\|T(t)\|_\infty$. Moreover, T is not only Toeplitz, but also nonnegative and upper triangular, hence $\|T(t)\|_\infty$ is the sum of the first row of $T(t)$. An explicit calculation of the matrix exponential $T(t) = e^{Bt}$ shows that the entries in the first row are constructed from binomial coefficients, and so the transient behaviour of B is bounded by $\|e^{Bt}\|_2 \leq e^{\lambda t}p(t), t \geq 0$ where the polynomial p is given by

$$p(t) = \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{t^k}{k!}.$$

By Lemma 5.3 the transient behaviour of B is an upper bound for all triangular Metzler matrices A with $A \leq B$, $\|e^{At}\|_2 \leq \|e^{Bt}\|_2 \leq e^{\lambda t}p(t)$, $t \geq 0$. ■

We now show that for Metzler matrices the determination of the transient behaviour with respect to the operator norms $\|e^{At}\|_1$ and $\|e^{At}\|_\infty$ reduces to solving just one initial value problem. Therefore the initial growth rates μ_1 and μ_∞ are easily obtained by simple matrix computations.

Lemma 5.5. *Given $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$. Then*

$$\mu_1(A) = \max_j (\mathbf{1}^\top A)_j, \qquad \mu_\infty(A) = \max_i (A\mathbf{1})_i,$$

where $\mathbf{1} = (1,\dots,1)^\top \in \mathbb{R}^n$ is a vector of ones. Moreover, for the matrix exponential we have $\|e^{At}\|_1 = \|\mathbf{1}^\top e^{At}\|_1$ and $\|e^{At}\|_\infty = \|e^{At}\mathbf{1}\|_\infty$.

Proof. Direct manipulation of the formulas presented in Theorem 2.41 yields

$$\mu_1(A) = \max_j \left(\operatorname{Re} a_{jj} + \sum_{i\neq j} |a_{ij}|\right) = \max_j \left(a_{jj} + \sum_{i\neq j} a_{ij}\right) = \max_j \sum_i a_{ij} = \max_j (\mathbf{1}^\top A)_j,$$

and analogously $\mu_\infty(A) = \max_i (A\mathbf{1})_i$. The matrix exponential e^{At} of $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ is a nonnegative matrix for $t \geq 0$. Hence $\|e^{At}\|_\infty = \|e^{At}\mathbf{1}\|_\infty$ and $\|e^{At}\|_1 = \|\mathbf{1}^\top e^{At}\|_1$ for $t \geq 0$. Choosing an initial value $x_0 = \mathbf{1}$ we therefore obtain the ∞-norm of the matrix exponential by considering the norm of the solution $x(t,\mathbf{1}) = e^{At}\mathbf{1}$. □

Let us now derive estimates on the transient growth based on Corollary 2.57. To take advantage of the positivity of the system, all vector norms under consideration must be monotone. Let us therefore introduce positive diagonal weights for the standard norms $\|\cdot\|_i, i \in \{1,2,\infty\}$. If $W = \operatorname{diag}(w_i)$ with $w \in \mathbb{R}^n$, $w > 0$, is such a positive diagonal weight and if $\|\cdot\|$ is a monotone vector norm then $\|W\cdot\|$ is also a monotone vector norm, and by Proposition 2.58 its eccentricity is given by $\operatorname{ecc}(\|W\cdot\|, \|\cdot\|) = \kappa(W) = \frac{\max_i w_i}{\min_i w_i}$. To obtain a transient estimate from Corollary 2.57, we need to know the initial growth rate associated with a weighted norm. The formula has already been derived in Proposition 2.58.
Candidates for diagonal weights are given by Perron vectors.

Theorem 5.6. *Suppose $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ is a stable Metzler matrix with Perron vector $x > 0$ and left Perron vector $y > 0$, $Ax = \alpha(A)x$, $y^\top A = \alpha(A)y^\top$. Then*

$$\|e^{At}\|_1 \leq \kappa(y)\ e^{\alpha(A)t}, \quad \|e^{At}\|_2 \leq \left(\kappa((\tfrac{y_i}{x_i})_i)\right)^{1/2} e^{\alpha(A)t}, \quad \|e^{At}\|_\infty \leq \kappa(x)\ e^{\alpha(A)t},$$

where $\kappa(z) = (\max_i z_i)(\min_i z_i)^{-1}$ is the condition number of a strictly positive vector $z > 0$.

Proof. Given a Metzler matrix A where the left and right Perron vectors y and x are strictly positive. Setting $W = \operatorname{diag}(y_i)$ gives $\mathbf{1}^\top WAW^{-1} = y^\top AW^{-1} = \alpha(A)y^\top \operatorname{diag}(y_i^{-1}) = \alpha(A)\mathbf{1}^\top$, hence by Proposition 2.58 the weighted initial growth rate satisfies $\mu_{1,W}(A) = \mu_1(WAW^{-1}) = \alpha(A)$. The condition number of W is given by $\kappa(W) = \kappa(\operatorname{diag}(y)) = \kappa(y)$. Hence Corollary 2.57 gives the estimate $\|e^{At}\|_1 \leq \kappa(y)e^{\alpha(A)t}, t \geq 0$. Analogously, $W^{-1} =$

$\operatorname{diag}(x_i)$ gives $\mu_{\infty,W}(A) = \alpha(A)$ with condition number $\kappa((x_i^{-1})_i) = \kappa(x)$. For the spectral norm, set $D = \operatorname{diag}(\frac{y_i}{x_i})$. Then $DA + A^\top D - 2\alpha(A)D$ is a symmetric Metzler matrix. We claim that it has the same Perron vector x as A. This may be seen by

$$(DA + A^\top D - 2\alpha(A)D)x = (\alpha(A)I + A^\top - 2\alpha(A)I)Dx = (A^\top - \alpha(A)I)y = 0.$$

The Perron vector is an eigenvector associated with the spectral abscissa, hence $DA + A^\top D - 2\alpha(A)D$ is negative semidefinite. Therefore we have the following inequality with respect to the Hermitian order relation,

$$DA + A^\top D \preceq 2\alpha(A)D.$$

Corollary 3.32 and Theorem 3.35 then give the transient estimate for the spectral case. □

The choice of Perron vectors as weights provides an estimate for the optimal decay rate $\alpha(A)$. This approach is impossible if the Perron vectors contain 0 entries. But weights which yield a transient estimate can be chosen from a much larger set.

Proposition 5.7. *Given a Metzler matrix $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$.*

(i) If A is exponentially stable then for every vector $b \in \mathbb{R}_+^n$ there exists a vector $w \in \mathbb{R}_+^n$ such that $Aw = -b$.

(ii) If there exists $w > 0$ with $Aw \leq 0$ ($Aw < 0$) then A is (exponentially) stable.

(iii) If the vector w satisfies the conditions of (ii) then the norm $\|Wx\|_\infty$ with $W^{-1} = \operatorname{diag}(w)$ is a Liapunov norm for $\dot{x} = Ax$. Its eccentricity is given by $\kappa(w)$, the corresponding initial growth rate by $\mu_{\infty,W}(A) = \max_j \frac{(Aw)_j}{w_j}$.

Proof. Let us assume that A is exponentially stable. Then Proposition 5.1 shows that $-A^{-1} \in \mathbb{R}_+^{n\times n}$. Hence $w = -A^{-1}b$ is a nonnegative vector, which shows *(i)*. If $w > 0$ is a strictly positive vector with $b = -Aw \geq 0$ then $W = \operatorname{diag}(w_i^{-1})$ gives $WAW^{-1}\mathbf{1} = WAw = -Wb \leq 0$. By Proposition 2.58 and Lemma 5.5 the weighted initial growth rate satisfies

$$\begin{aligned}\mu_{\infty,W}(A) = \mu_\infty(WAW^{-1}) &= \max_j (WAW^{-1}\mathbf{1})_j = \max_j (\operatorname{diag}(w_i^{-1})Aw)_j \\ &= \max_j \left(-\operatorname{diag}(w_i^{-1})b\right)_j = -\min_j \tfrac{b_j}{w_j} \leq 0.\end{aligned}$$

Hence *(iii)* is proved. Now *(iii)* implies the (exponential) stability in *(ii)*, as the initial growth rate is non-positive for $b \geq 0$, and it is negative for $b > 0$. □

A dual result of Proposition 5.7 holds for $\|\cdot\|_1$. We list it here for completeness.

Corollary 5.8. *Given a Metzler matrix $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$.*

(i) If A is exponentially stable then there exists for every $b \in \mathbb{R}_+^n$ a vector $w \in \mathbb{R}_+^n$ such that $w^\top A = -b^\top$.

(ii) *If there exists $w > 0$ with $w^\top A \leq 0$ ($w^\top A < 0$) then A is (exponentially) stable.*

(iii) *If the vector w satisfies the conditions of (ii) then norm $\|Wx\|_1$ with $W = \operatorname{diag}(w)$ is a Liapunov norm for $\dot{x} = Ax$. Its eccentricity is given by $\kappa(w)$, the corresponding initial growth rate by $\mu_{1,W}(A) = \max_j \frac{(A^\top w)_j}{w_j}$.*

Proof. We only show *(iii)* as statements *(i)* and *(ii)* follow analogously to the proofs in Proposition 5.7. As $w > 0$ is a strictly positive vector, W is invertible and $\mathbf{1}^\top WAW^{-1} = w^\top AW^{-1} = -bW^{-1} \leq 0$. Hence the initial growth rate with respect to $\|W\cdot\|_1$ is given by

$$\begin{aligned}\mu_{1,W}(A) = \mu_1(WAW^{-1}) &= \max_i(\mathbf{1}^\top WAW^{-1})_i = \max_i(w^\top AW^{-1})_i \\ &= \max_i(-b^\top W^{-1})_i = -\min_i \tfrac{b_i}{w_i}.\end{aligned}$$

Thus $\|W\cdot\|_1$ is a Liapunov norm for A if $b \geq 0$, and a strict Liapunov norm for $b > 0$. □

Note that we have the following simple formulas if $x \in \mathbb{R}^n_+$, because for positive x and and positive diagonal weight $W = \operatorname{diag}(w_i)$ we have

$$\left\|W^{-1}x\right\|_\infty = \max_i \tfrac{x_i}{w_i}, \qquad \|Wx\|_1 = \sum_i (w_i x_i).$$

Proposition 5.7 and Corollary 5.8 motivate the following definition.

Definition 5.9. For a given Metzler matrix $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ the strictly positive vector $w \in \mathbb{R}^n_+$ is called a right (or left) *Liapunov vector* of A if $Aw \leq 0$ or $w^\top A \leq 0$, respectively. If the strict inequality holds, $Aw < 0$ or $w^\top A < 0$, then w is called a *strict* Liapunov vector.

If there exists a left Liapunov vector v of a given matrix $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ then $\mu_\nu(A) \leq 0$ for the vector norm $\nu(x) = \|\operatorname{diag}(w)x\|_1$. If v is a strict Liapunov vector of A then A generates a uniform contraction semigroup.

Lemma 5.10. *Suppose that A is an invertible Metzler matrix. There exists $z \in \mathbb{R}^n_+$ with $A^{-1}z < 0$ if and only there exists a right Liapunov vector of A.*

Proof. This becomes obvious by considering the right Liapunov vector $w = -A^{-1}z$ of A. □

For the spectral norm we obtain the following result which extends Theorem 5.6.

Proposition 5.11. *Suppose that A is a Metzler matrix. For all strictly positive vectors $v, w > 0$ such that $A^\top v \leq 0$ and $Aw \leq 0$ the diagonal matrix $P = \operatorname{diag}({}^{v_i}/_{w_i})$ is a quadratic Liapunov matrix for A which satisfies $PA + A^\top P \preceq 0$.*

For the proof of this proposition we need the following lemma.

Lemma 5.12. *Suppose that $R \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ is a symmetric Metzler matrix. If there exists a right Liapunov vector $v > 0$ of R then R is negative semidefinite, $R \preceq 0$.*

Proof. If there exists a right Liapunov vector $v > 0$ of R then R is a stable matrix by Proposition 5.7 *(ii)*. As R is symmetric, it is negative semidefinite. □

Proof (of Proposition 5.11). First note that for $P = \operatorname{diag}(v_i/w_i)$ the matrix $R = PA + A^\top P$ is a symmetric Metzler matrix, and

$$Rw = (PA + A^\top P)w = PAw + A^\top v \leq 0.$$

Hence w satisfies the condition of Lemma 5.12 for $R = PA + A^\top P$. Therefore $R \preceq 0$ and P is a quadratic Liapunov matrix for A. □

In [40] it has been shown that

Proposition 5.13. *A Metzler matrix $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ is stable if and only if there exists a diagonal quadratic Liapunov function, $P = \operatorname{diag}(p_i) \succ 0$ with $PA + A^\top P \preceq 0$.*

Proof. The existence of a diagonal quadratic Liapunov matrix P follows from the existence of left and right Liapunov vectors by Proposition 5.7 and Corollary 5.8. Proposition 5.11 shows how to construct the matrix P from these vectors. The converse implication follows from Liapunov's direct stability theorem. □

5.3 Optimal Liapunov Vectors

The last section showed that there is a broad range of Liapunov vectors available for positive systems. We will now show how to obtain a Liapunov vector for which the condition number is minimal. This is of interest for bounding the norm of the matrix exponential. To this end, note that if $w \in \mathbb{R}_+^n$ is a Liapunov vector of a Metzler matrix A with $Aw \leq 0$ then its condition number $\kappa(w) = \max_i w_i / \min_i w_i$ gives an estimate of the transient growth via Corollary 2.57,

$$\left\| e^{At} \right\|_\infty \leq \kappa(w) e^{\mu_{\infty,w}(A)t} \leq \kappa(w),$$

where $\mu_{\infty,w}$ is the initial growth rate with respect to the vector norm $\|\operatorname{diag}(w_i)^{-1}\cdot\|_\infty$. This initial growth rate then satisfies $\mu_{\infty,w}(A) \leq 0$.

Varying the weights w we try to minimize the condition number such that we obtain an optimal estimate of the transient bound. For this, we pose the following optimization problem.

Problem 5.14. *For a given exponentially stable Metzler matrix $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ find a vector $\hat{x} \in \mathbb{R}_+^n$ which is a minimizing argument of*

$$\hat{\gamma} = \min_{x\geq 0, x\neq 0} \left[\max_i (-A^{-1}x)_i \right] \left[\min_i (-A^{-1}x)_i \right]^{-1}. \tag{5.2}$$

As (5.2) is invariant under multiplication with positive scalars, x in (5.2) may be chosen from a compact basis of the cone $\mathbb{R}_+^n$. If $\hat{x}$ is a positive vector which minimizes (5.2) then the optimal weight $\hat{w} = -A^{-1}\hat{x}$ is a Liapunov vector for A, and the optimal value $\hat{\gamma}$ is the condition number $\kappa(\hat{w})$ of this Liapunov vector. Let us now characterize the optimal values of Problem 5.14.

Proposition 5.15. *Suppose that $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ is an exponentially stable Metzler matrix. For a given weight $w \in \mathbb{R}_{+}^{n}$ with $\min_i w_i = 1$ we define the index sets*

$$J(w) = \{i \in \{1,\ldots,n\} \,|\, w_i = 1\}\,, \qquad H(w) = \{h \in \{1,\ldots,n\} \,|\, (Aw)_h = 0\}\,. \tag{5.3}$$

The strictly positive vector w is an optimal weight of Problem 5.14 satisfying $\hat{\gamma} = \max_i w_i$ if and only if $H(w) \cup J(w) = \{1,\ldots,n\}$. Moreover, such an optimal weight w always exists. It is uniquely determined if $J(w) = H(w)^{\mathrm{C}}$.

Proof. Let us first show that the feasible set of Problem 5.14 is non-empty. By Proposition 5.7 there exists a right Liapunov vector w^0 of A. Hence the set $\left\{x \in \mathbb{R}_{+}^{n} \,\middle|\, x \neq 0\right\}$ contains the point $x^0 = -Aw^0$ and therefore the problem

$$\hat{\gamma} = \min_{Aw\leq 0, w>0} \frac{\max_i w_i}{\min_i w_i} = \min_{x\in\mathbb{R}_{+}^{n}, x\neq 0} \frac{\max_i(-A^{-1}x)_i}{\min_i(-A^{-1}x)_i} \tag{5.4}$$

is feasible. If A is diagonally dominant then $w = \mathbf{1}$ satisfies Problem 5.14 with $\hat{\gamma} = 1$, $J(w) = \{1,\ldots,n\}$. Hence $J(w) = \{1,\ldots,n\}$ so that $H(w) \cup J(w) = \{1,\ldots,n\}$. Let us now suppose that $w \neq \mathbf{1}$ is a positive vector with $\min_i w_i = 1$ which corresponds to the optimal solution $\hat{\gamma} = \max w_i$ of Problem 5.14. Then w is also an optimal feasible solution of the linear programming (LP) problem,

$$\text{minimize} \quad w_{i_0} \quad \text{subject to} \quad w_i \geq 1,\ (Aw)_i \leq 0, \quad i = 1,\ldots,n,$$

for some suitable index i_0. Writing $y = w - \mathbf{1}$ and introducing slack variables z we rewrite this linear programming problem into standard form,

$$\text{minimize } [e_{i_0}^{\top}\, 0] \begin{bmatrix} y \\ z \end{bmatrix} + 1 \text{ subject to } \begin{bmatrix} y \\ z \end{bmatrix} \geq 0,\ \begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = -A\mathbf{1}.$$

If the solution $\left[\begin{smallmatrix} y \\ z \end{smallmatrix}\right]$ is optimal then it satisfies the Kuhn-Tucker conditions. For LP problems these conditions are called complementary slackness and provide a necessary and sufficient condition for optimal solutions, see [102, Section 4.4]. In this case, the optimal positive vectors y and z are orthogonal, that is, for each $i \in \{1,\ldots,n\}$ either $y_i \geq 0$ and $z_i = 0$ or $y_i = 0$ and $z_i > 0$. In terms of w this means that w is an optimal solution if and only if $H(w) \cup J(w) = \{1,\ldots,n\}$.

Let us now show that the optimal solution w of (5.4) is uniquely determined under the additional condition that $J(w) = H(w)^{\mathrm{C}}$. To see this, let us assume that w^1 and w^2 are two different optimal weights with $\min_i w_i^j = 1$, $\max_i w^j = \hat{\gamma}$, and $J(w^j) = H(w^j)^{\mathrm{C}}$, $j = 1, 2$. and associated index sets J_j If $J(w^1) \neq J(w^2)$ then $w' = \frac{1}{2}(w^1 + w^2)$ satisfies $Aw' \leq 0$ and $w' \geq \mathbf{1}$. Especially, $(Aw')_i = 0$ for $i \in H(w') = H(w^1) \cap H(w^2)$ and $w'_i = 1$ for $J(w') = J(w^1) \cap J(w^2)$, whence $H(w') \cup J(w') \neq \{1,\ldots,n\}$. Thus w' is not optimal and there exists a feasible search direction which decreases the condition number of w'. But as $\kappa(w') \leq \hat{\gamma}$, this contradicts the optimality of $\hat{\gamma}$. Therefore $J_1 = J_2$ has to hold. From (5.3) and $J = H^{\mathrm{C}}$ we get n linear independent equations which are simultaneously satisfied by w^1 and w^2. But this implies $w^1 = w^2$, i.e., the optimal weight is unique. □

The following algorithm solves Problem 5.14.

Algorithm 5.16. Let $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ be an exponentially stable Metzler matrix. Let $S = -A^{-1}$ where $S_{(J,J)} \in \mathbb{R}_{+}^{m\times m}$ denotes the submatrix obtained from S by keeping columns and rows with indices in the ordered set $J = \{j_1, \dots, j_m \,|\, j_1 < \dots < j_m\}$ and $J_{(K)} = \{j_k \in J \,|\, k \in K\}$ denotes the ordered index set obtained from J by keeping the elements indexed by K. Analogously, the vector x_J consists of the elements of x indexed by J. Then the following algorithm calculates an optimal weight $w = Sx$ for Problem 5.14 when $S = -A^{-1}$.

```
Init  Set $J = \{1, \dots, n\}$.
Loop  Solve $S_{(J,J)}y = \mathbf{1}$ for $y$.
      If $y \geq 0$ then set $x_J = y$, $x_{i \notin J} = 0$, and return $w = Sx$.
      Otherwise set $K = \{i \,|\, y_i < 0\}$ and $J = J_{(K)}$.
```

The algorithm terminates in a finite number of steps, namely if $J = \{j\}$ then $S_{(J,J)} = s_{jj} > 0$ as $S \in \mathbb{R}_{+}^{n\times n}$ and $x = s_{jj}^{-1}e_j$. The first iteration of the algorithm is skipped by starting with the index set $J = \{i \in \{1, \dots, n\} \,|\, (A\mathbf{1})_i < 0\}$ since in the first step $y_i = (-A\mathbf{1})_i$. Algorithm 5.16 produces an optimal value for Problem 5.14.

Corollary 5.17. *The weight $\hat{w} = S\hat{x}$ calculated by Algorithm 5.16 is an optimal solution of Problem 5.14 with $\hat{\gamma} = \kappa(\hat{w})$.*

Proof. By construction a weight $\hat{w} = S\hat{x}$ computed by Algorithm 5.16 satisfies $(A\hat{w})_i = -\hat{x}_i = 0$ for $i \notin J = \{i \,|\, \hat{w}_i = 1\}$. Hence it is an optimal weight by Proposition 5.15 and the optimal condition number is given by $\hat{\gamma} = \kappa(\hat{w}) = \max_i \hat{w}_i$. □

Example 5.18. Consider the system $\dot{x} = Ax$, $A = \left(\begin{smallmatrix} -5 & 36 \\ 2 & -20 \end{smallmatrix}\right)$. An optimal right Liapunov vector is given by $w^r = \binom{7.2}{1}$ and an optimal left Liapunov vector is given by $w^\ell = \binom{1}{1.8}$. Figure 5.1 shows the boxes $\|x\|_\infty = 1$ and $\|x\|_1 = 2$ in $\mathbb{R}_+^2$ shaded in gray. Some trajectories show that these are not invariant under the flow of $\dot{x} = Ax$. Note that we only have to check trajectories with initial values in the vertices of these boxes by Lemma 5.5. Now, the boxes induced by the optimal weights are both invariant under the flow. The trajectories enter the optimal boxes tangentially, so that in both cases the associated weighted initial growth is 0.
For the 1-norm we see that the transient amplification $M_0 = 1.5$ is bounded by the eccentricity of the norm which is the condition number of the left Liapunov vector, $\hat{\gamma} = 1.8$.
In contrast, the estimate provided by Theorem 5.6 gives $\left\|e^{At}\right\|_1 \leq 1.91e^{-1.18t}$. For the ∞-norm, the transient amplification is $M_0 = 2$ which is bounded by the condition number of the right Liapunov vector, $\kappa = 7.2$. Based on the right Perron vector, Theorem 5.6 gives $\left\|e^{At}\right\|_\infty \leq 9.41e^{-1.18t}$. ■

Remark 5.19. Problem 5.14 has the following geometric interpretation. If $\hat{\gamma}$ is defined by (5.2) then $\log \hat{\gamma}$ is the distance of the polyhedron $\{x \in \mathbb{R}^n | Ax \leq 0\} \subset \mathbb{R}_+^n$ to the diagonal $\mathbb{R}\mathbf{1}$ is if this distance is measured with respect to Hilbert's projective metric,

$$d(x, y) = -\log\left(\min_i \tfrac{x_i}{y_i} \min_i \tfrac{y_i}{x_i}\right), \qquad x, y \in \mathring{\mathbb{R}}_+^n. \tag{5.5}$$

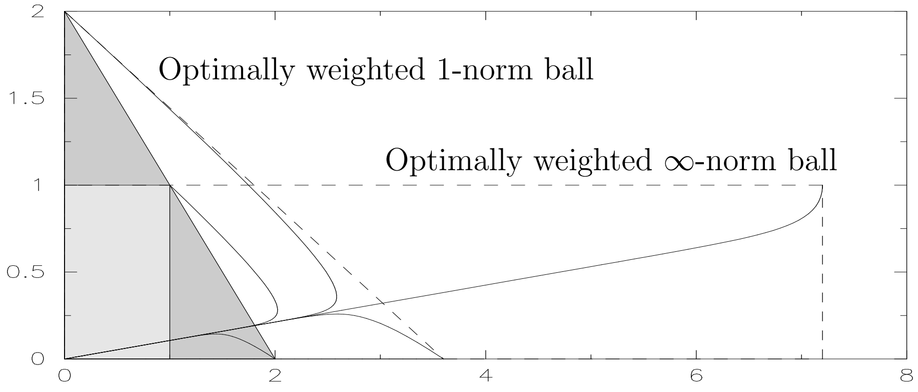

Figure 5.1: Liapunov norms induced by optimal weights.

Algorithm 5.16 selects those faces of the polyhedron that have the shortest distance to the diagonal $\mathbb{R}\mathbf{1}$. Passing to a subset of indices is a projection on a lower dimensional subface for which the procedure of the algorithm is repeated.

The projective metric (5.5) is also related to the transient behaviour of the spectral norm.

Theorem 5.20. *Suppose that $A \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ is an exponentially stable Metzler matrix. Consider the sets*

$$\mathcal{W}^\ell = \{z \in \mathbb{R}^n_+ \mid \|z\|_2 = 1, A^\top z < 0\}, \qquad \mathcal{W}^r = \{z \in \mathbb{R}^n_+ \mid \|z\|_2 = 1, Az < 0\}$$

of normed left and right Liapunov vectors. The minimal projective distance of the points in these sets is given by

$$d(\mathcal{W}^\ell, \mathcal{W}^r) = \inf\left\{-\log\left(\min_i \frac{x_i}{y_i} \cdot \min_i \frac{y_i}{x_i}\right) \,\Big|\, x \in \mathcal{W}^\ell, y \in \mathcal{W}^r\right\}.$$

This quantity provides an upper bound to the spectral transient excursion through

$$\left\|e^{At}\right\|_2 \le e^{1/2\, d(\mathcal{W}^\ell,\mathcal{W}^r)} \qquad \textit{for all} \quad t \ge 0. \tag{5.6}$$

Proof. If $x \in \mathcal{W}^\ell$ and $y \in \mathcal{W}^r$ then the matrix $P = \operatorname{diag}(x_i/y_i)_i$ is a quadratic Liapunov matrix for A, see Proposition 5.11. Hence Theorem 3.35 implies that

$$\left\|e^{At}\right\|_2 \le \inf_{x\in\mathcal{W}^\ell, y\in\mathcal{W}^r} \sqrt{\kappa\left(\operatorname{diag}(\tfrac{x_i}{y_i})_i\right)}, \quad t \ge 0.$$

The condition number under the square root is given by

$$\kappa(\operatorname{diag}(\tfrac{x_i}{y_i})_i) = \max_i(\tfrac{x_i}{y_i})_i / \min_i(\tfrac{x_i}{y_i})_i = \left(\min_i(\tfrac{y_i}{x_i})_i \cdot \min_i(\tfrac{x_i}{y_i})_i\right)^{-1}, \tag{5.7}$$

so that the infimum of (5.7) over all Liapunov vectors is given by $e^{d(\mathcal{W}^\ell,\mathcal{W}^r)}$. Taking square roots gives (5.6). □

The projective distance $d(\mathcal{W}^\ell, \mathcal{W}^r)$ gives the minimal condition number of a diagonal quadratic Liapunov matrix. These diagonal matrices are the only ones for which the associated elliptic norms are monotone, as for a monotone norm the induced operator norm has to satisfy $\|W^{-1}DW\|_2 = \|D\|_2 = \max_i |d_i|$ for all diagonal matrices $D = \operatorname{diag}(d_i)$, see Lemma 1.9, which is only possible if W itself is diagonal.
We do not provide an algorithm to compute the distance $d(\mathcal{W}^\ell, \mathcal{W}^r)$ but let us consider the following special case.

Corollary 5.21. *If $\mathcal{W}^\ell \cap \mathcal{W}^r \neq \emptyset$ then A generates a spectral contraction.*

Proof. Clearly, if $x \in \mathcal{W}^\ell \cap \mathcal{W}^r$ then $P = I = \operatorname{diag}(x_i/x_i)$ is a quadratic Liapunov matrix for A. Hence A is already dissipative with respect to the spectral norm. □

If the cones generated by the positive linear combinations of the columns of A and $A^\top$ have non-empty intersection then there exist strictly positive vectors x, y, z such that $x = Ay = A^\top z$, or equivalently, as A is invertible, there exists $z > 0$ with $A^{-1}A^\top z > 0$. In the next section we will generalize this fact about a common Liapunov vector when we replace A and $A^\top$ by A_1 and A_2 and look for a common Liapunov function.

5.4 Common Liapunov Vectors

In this section we derive necessary and sufficient conditions for the existence of common Liapunov vectors for a set of positive systems.

Theorem 5.22. *Given a set of square matrices $A_i \in \mathbb{R}^{n\times n}, i \in \{1, \dots, k\}$, there exists a vector $w \in \mathbb{R}^n_+$ with $w^\top A_i < 0$ for all $i \in \{1, \dots, k\}$ if and only if $[A_1 \dots A_k]y \ngeq 0$ holds for all vectors $y \in \mathbb{R}^{nk}_+$, $y \neq 0$.*

Proof. The proof follows directly from a separation principle for two convex cones, see [133, Theorem 3.3.4]. Consider the polytopic convex cone, $\operatorname{cone}(A) = \{Ax \,|\, x \geq 0\} \subset \mathbb{R}^{nk}$, generated from the columns of

$$A = \begin{bmatrix} A_1^\top \\ \vdots \\ A_k^\top \end{bmatrix},$$

and the cone given by the (strictly) negative orthant $\mathring{\mathbb{R}}^{nk}_- = \{y \in \mathbb{R}^{nk} \,|\, y < 0\}$. Then either $\operatorname{cone}(A) \cap \mathring{\mathbb{R}}^{nk}_- \neq \emptyset$ or there exists a separating hyperplane induced by a vector $y \in \mathbb{R}^{nk}$, such that

$$\forall z \in \operatorname{cone}(A): \; y^\top z \geq 0, \qquad \forall b \in \mathring{\mathbb{R}}^{nk}_-: \; y^\top b < 0. \tag{5.8}$$

Now, if $\operatorname{cone}(A) \cap \mathring{\mathbb{R}}^{nk}_-$ is non-empty then there exists $w \in \mathbb{R}^n_+$ such that $Aw \in \mathring{\mathbb{R}}^{nk}_-$. Hence $w^\top A_i < 0$ for all $i = 1, \dots, k$. On the other hand, if $y \in \mathbb{R}^{nk}_+$ satisfies $[A_1 \dots A_k]y \geq 0$ then (5.8) holds, hence the cones are separated by a hyperplane induced by $y \in \mathbb{R}^{nk}_+$, $y \neq 0$. □

Here we do not assume that the matrices are of Metzler type. To turn the following results into stability characterizations, the matrices must be Metzler to ensure that the strictly positive vector $w > 0$ with $A^\top w < 0$ give rise to a strict Liapunov function.
For sets of Metzler matrices we introduce the following notion.

Definition 5.23. Let $\mathcal{A} \subset \mathbb{R}^{n\times n}_{\mathrm{M}}$ be a set of Metzler matrices. The strictly positive vector $w > 0$ is called a *common (right) Liapunov vector* for $\mathcal{A}$ if for all $A \in \mathcal{A}$, $Aw \leq 0$ holds. The terms *common left Liapunov vector*, *common strict Liapunov vector* are defined in accordance with Definition 5.9.

The results of Proposition 5.7 and of Corollary 5.8 also hold for common Liapunov vectors. Hence these common Liapunov vectors define joint Liapunov norms for sets of Metzler matrices, see Subsection 2.4.2. Let us now consider two Metzler matrices $A_1, A_2 \in \mathbb{R}^{n\times n}$.

Proposition 5.24. *Given $A_1, A_2 \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ where A_1 is exponentially stable, then there exists a common strict left Liapunov vector for the pair (A_1, A_2) if and only if there exists $z > 0$ with $z^\top A_1^{-1} A_2 > 0$.*

Proof. To prove the assertion, note that if $x \in \mathbb{R}^n_+$ is a common strict left Liapunov vector of (A_1, A_2) then

$$x^\top [A_1\ A_2] = x^\top A_1 [I\ A_1^{-1} A_2] < 0. \tag{5.9}$$

Now, setting $z = -A_1^\top x$ we obtain from (5.9) the inequalities $z > 0$ and $z^\top A_1^{-1} A_2 > 0$. On the other hand, if the positive vector z satisfies $z^\top A_1^{-1} A_2 > 0$ then $x = -A_1^{-\top} z$ defines a common Liapunov vector. □

This proposition does not cover the case when there only exists a weak common Liapunov vector as the following example shows.

Example 5.25. Consider the matrices $A_1 = \left(\begin{smallmatrix} -10 & 5 \\ 5 & -3 \end{smallmatrix}\right)$ and $A_2 = \left(\begin{smallmatrix} -10 & 4 \\ 6 & -3 \end{smallmatrix}\right)$. Then for $w = (3, 5)^\top$ we have $w^\top A_1 = (-5, 0)$ and $w^\top A_2 = (0, -3)$. Now $A_1^{-1} A_2 = \left(\begin{smallmatrix} 0 & 3/5 \\ -2 & 2 \end{smallmatrix}\right)$ has a column of non-positive values, hence the condition of Proposition 5.24 cannot be satisfied by any positive vector. ■

Arguing as is the example, we can draw the following conclusion.

Corollary 5.26. *Let $A_1, A_2 \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ where A_1 is exponentially stable. If $A_1^{-1} A_2$ contains a column of negative entries then there does not exist a common strict left Liapunov vector.*

Remark 5.27. For an arbitrary matrix $A \in \mathbb{R}^{n\times n}$, the existence of a strictly positive vector $w > 0$ with $Aw \leq 0$ does not guarantee its stability. We can only conclude that the trajectories restricted to the positive orthant are bounded. This implies that there does not exists a positive eigenvector of A which is associated with an eigenvalue of positive real part. Figure 5.2 shows some trajectories of $A = \left(\begin{smallmatrix} 1 & -1 \\ -1 & -1 \end{smallmatrix}\right)$ for which $v = \binom{1}{2}$ satisfies $Av < 0$. Here the trajectories enter the triangle depicted in the figure through the segment $\{x > 0 \,|\, v^\top x = \text{const}\}$. Unfortunately, this triangle is not invariant under the flow of A. Now, we introduce the Metzler matrix $A_1 = \left(\begin{smallmatrix} -3 & 0 \\ 1 & -1 \end{smallmatrix}\right)$ for which v is a left Liapunov vector.

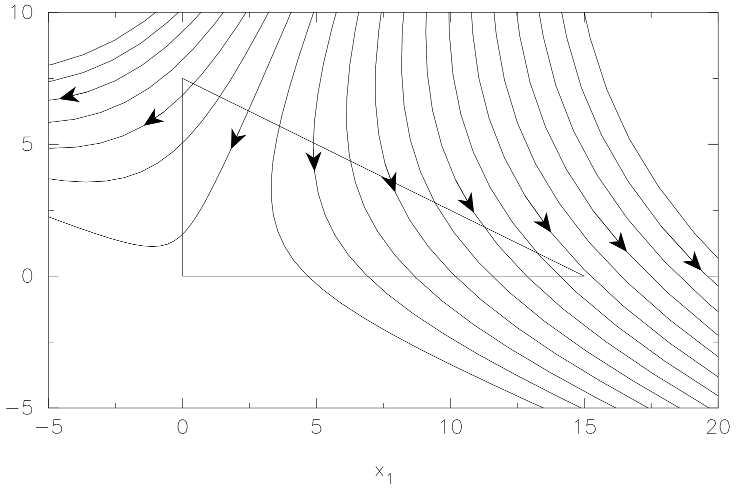

Figure 5.2: Trajectories of a non-positive system with a Liapunov vector.

For all $\alpha \in \mathbb{R}$ the matrix $A + \alpha A_1, \alpha \geq 0$ is not of Metzler type. However, for $\alpha > \alpha^* = 1$ the matrix $A + \alpha A_1$ is exponentially stable. Therefore we can think of A_1 as a Metzler direction towards stability. A related result for quadratic Liapunov matrices and rank-one update matrices is presented in Shorten et al. [125].
Proposition 5.24 gives a stability criterion only if the matrices A_1, A_2 are both of Metzler type. In this case, the Liapunov vector w defines a linear Liapunov function given by $x \mapsto \|\text{diag}(w)x\|_1$.

The existence of a common Liapunov vector allows us to conclude that a whole set of matrices consists of exponentially stable matrices.

Proposition 5.28. *Suppose that $A_1, A_2 \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ are Metzler matrices and that there exists $z \in \mathbb{R}^n_+$ which satisfies $z^\top A_1^{-1} A_2 > 0$ and $z^\top A_1^{-1} < 0$. Then the* matrix interval

$$[[A_1, A_2]] := \{\tau A_1 + (1-\tau)A_2 \,|\, \tau \in [0,1]\} \subset \mathbb{R}^{n\times n}_{\mathrm{M}} \tag{5.10}$$

consists of exponentially stable matrices which all satisfy the same transient bound,

$$A \in [[A_1, A_2]] : \qquad \left\|e^{At}\right\|_1 \leq \kappa(A_1^{-\top} z),\ t \geq 0.$$

Proof. The vector $w = -A_1^{-\top} z$ is a common strict left Liapunov vector of A_1 and A_2, i.e., $w^\top A_i < 0$, $i = 1, 2$. But then w is also a Liapunov vector for all convex combinations of A_1 and A_2. Hence $w^\top A < 0$ for all $A \in [[A_1, A_2]]$. By Corollaries 2.57 and 5.8 all matrices A from this matrix interval satisfy the growth estimate $\left\|e^{At}\right\|_1 \leq \kappa(w)e^{\mu_{1,w}(A)t} \leq \kappa(w)$ for $t \geq 0$ as w induces a Liapunov norm for the whole matrix interval, $\mu_{1,w}(A) \leq 0$, $A \in [[A_1, A_2]]$. □

We can generalize Proposition 5.24 to multiple matrices. If there is no common Liapunov vector for a set of matrices then there is clearly no Liapunov vector for a larger set.

Corollary 5.29. *There exists a common strict left Liapunov vector for the Metzler matrices $A_1, A_2, \dots, A_k$ where A_1 is exponentially stable, if and only if there exists $z \in \mathbb{R}^n_+$ such that for all $\ell = 1, \dots, k : z^\top A_1^{-1} A_\ell > 0$.*

Proof. The condition of the corollary can be rewritten as $z^\top [I, A_1^{-1}A_2, \dots, A_1^{-1}A_k] > 0$. Hence z is strictly positive. As A_1 is an exponentially stable Metzler matrix, $A_1^{-1} \leq 0$. Setting $y = -A_1^{-\top} z > 0$ we obtain $y^\top [A_1, \dots A_k] < 0$, hence y is a Liapunov vector for all matrices $A_1, \dots, A_k$. Conversely, if $y > 0$ is a common strict left Liapunov vector for $A_1, \dots, A_k$, we set $z = -A_1 y > 0$ and obtain the required condition $z^\top A_1^{-1} A_\ell = -y^\top A_\ell > 0$ for all $\ell = 1, \dots, k$. □

Let us now consider the relation between common quadratic and linear Liapunov functions. By Proposition 5.13, we only have to consider diagonal quadratic Liapunov matrices. Combining Proposition 5.11 and Proposition 5.24 we obtain the following result which can be viewed as a corollary to Theorem 2.64.

Corollary 5.30. *Suppose that $A_1, A_2 \in \mathbb{R}^{n \times n}_{\mathrm{M}}$ are Metzler matrices and that A_1 is exponentially stable. If there exist positive vectors $z_1, z_2 > 0$ which satisfy $z_1^\top A_1^{-1} A_2 < 0$ and $A_2 A_1^{-1} z_2 < 0$ then there exists a diagonal common quadratic Liapunov matrix for A_1 and A_2 given by $P = \mathrm{diag}({}^{w_1}/_{w_2})$ where $w_1 = -A_1^\top z_1 > 0$ and $w_2 = -A_1^{-1} z_2 > 0$.*

Example 5.31. There are pairs of Metzler matrices which do not have a common linear Liapunov function, but a quadratic one. Consider A_1 of Example 5.25 and $A_3 = \left(\begin{smallmatrix} -10 & 2 \\ 8 & -3 \end{smallmatrix}\right)$. Then Corollary 5.26 shows that there does not exist a left Liapunov vector because $A_1^{-1} A_3 = \left(\begin{smallmatrix} -2 & 1.8 \\ -6 & 4 \end{smallmatrix}\right)$ has a column of negative entries. However, $P = \left(\begin{smallmatrix} 5 & 0 \\ 0 & 3 \end{smallmatrix}\right)$ is a positive definite matrix with $PA_i + A_i^\top P \prec 0$ for $i = 1, 3$. ■

Now, let us study the converse question, if the existence of a Liapunov vector implies the existence of a common diagonal quadratic Liapunov matrix. Unfortunately, this is not true as the following example shows.

Example 5.32. Let us consider the matrices $A_1 = \left(\begin{smallmatrix} -5 & 39 \\ 0 & -5 \end{smallmatrix}\right)$ and $A_2 = \left(\begin{smallmatrix} -1 & 6 \\ 2 & -20 \end{smallmatrix}\right)$. These two Metzler matrices have a common right Liapunov vector $\binom{8}{1}$, but no common left Liapunov vector as $A_1^{-1} A_2$ has a column of negative entries, see Corollary 5.26. Hence we cannot construct a common diagonal quadratic Liapunov matrix based upon Proposition 5.11. Using the visual method developed in Subsection 4.3.1 we see that the Liapunov cones associated with A_1 and A_2 contain a common subset $\{\left(\begin{smallmatrix} 1+\alpha & \beta \\ \beta & 1-\alpha \end{smallmatrix}\right) \mid \alpha^2 + \beta^2 \leq 1\}$, for example, an element is given by $\alpha = -0.95$ and $\beta = -0.2$, but there is no element in this intersection which corresponds to $\beta = 0$. Hence there exists no common diagonal quadratic Liapunov matrix for the matrices A_1 and A_2. ■

One can also ask for the existence of common full-block quadratic Liapunov matrices, but – as already noted – weighting the spectral norm with such non-diagonal matrices destroys the monotonicity of the norm which is undesirable.

For non-autonomous positive linear systems, the following result is a direct application of Theorem 2.70 and Proposition 5.7.

Theorem 5.33. *Consider the time-dependent linear differential equation* $\dot{x}(t) = A(t)x(t)$ *where* $A : \mathbb{R}_+ \to \mathbb{R}_{\mathrm{M}}^{n\times n}$ *is locally integrable. Given a strictly positive vector* $w \in \mathbb{R}_+^n$ *we set* $b(t) = -A(t)w$. *Then the solutions* $x(t, t_0, x_0)$ *satisfy the following growth bound,*

$$\|x(t,t_0,x_0)\|_\infty \le \kappa(w) e^{-\int_{t_0}^t \min_i \left(\frac{b(s)}{w}\right) ds} \|x_0\|_\infty, \qquad t \ge t_0.$$

If $b(t)$ *is nonnegative almost everywhere, the vector* w *is a common Liapunov vector and*

$$\|x(t,t_0,x_0)\|_\infty \le \kappa(w) \|x_0\|_\infty, \qquad t \ge t_0.$$

5.5 The Metzler Part of a Matrix

We now want to apply the results obtained for Metzler matrices to arbitrary matrices. Let us associate a Metzler matrix with every matrix $A = (a_{ij}) \in \mathbb{K}^{n\times n}$, called the *Metzler part* of A which is given by

$$M(A) = \mathrm{Re}\ \mathrm{Diag}(A) + |A - \mathrm{Diag}(A)| = (m_{ij}), \qquad m_{ij} = \begin{cases} \mathrm{Re}\, a_{ij}, & i = j, \\ |a_{ij}|, & i \ne j, \end{cases} \tag{5.11}$$

where $\mathrm{Diag}(A) = \mathrm{diag}(a_{11}, \cdots, a_{nn})$. The following simple lemma is of basic importance.

Lemma 5.34. *Let* $A \in \mathbb{K}^{n\times n}$, *then*

(i) The function $r \mapsto \rho(A + rI_n) - r$ *is monotonically decreasing on* $\mathbb{R}_+$ *and*

$$\alpha(A) = \lim_{r\to\infty} (\rho(A + rI_n) - r). \tag{5.12}$$

(ii) The map $r \mapsto M_r(A) := |A + rI_n| - rI_n$ *is componentwise decreasing on* $\mathbb{R}_+$ *and*

$$M(A) = \lim_{r\to\infty} |A + rI_n| - rI_n. \tag{5.13}$$

Proof. (i). For every $\lambda \in \mathbb{C}$ we have

$$0 \le r_1 \le r_2 \quad \Rightarrow \quad |\lambda + r_2| - r_2 = |\lambda + r_1 + (r_2 - r_1)| - r_2 \le |\lambda + r_1| - r_1. \tag{5.14}$$

Using $|r + \lambda| = \left((r+\lambda)(r+\bar{\lambda})\right)^{1/2}$ and $\sqrt{1+2z} = 1 + z + O(z^2)$ the limit is given by

$$\lim_{r\to\infty} (|\lambda + r| - r) = \lim_{r\to\infty} \left(r\sqrt{1 + 2\tfrac{\mathrm{Re}\,\lambda}{r} + \tfrac{|\lambda|^2}{r^2}} \right) - r = \mathrm{Re}\,\lambda, \quad \lambda \in \mathbb{C}. \tag{5.15}$$

Now by definition $\rho(A + rI_n) - r = \max\{|\lambda + r| - r \mid \lambda \in \sigma(A)\}$ and so the monotonicity property of $r \mapsto \rho(A + rI_n) - r$ follows directly from (5.14), while (5.12) follows from (5.15). *(ii).* Applying (5.14) and (5.15) to the diagonal entries of $M_r(A) := |A + rI_n| - rI_n$ we get $(|a_{ii} + r| - r) \to \mathrm{Re}\, a_{ii}$ monotonically as $r \to \infty$ whereas the off-diagonal entries $|a_{ij}|, i \ne j$, of $M_r(A)$ remain constant. Hence we obtain (5.13). □

As a consequence we obtain the following monotonicity property for the spectral abscissa which is a counterpart to (5.1).

$$\forall A \in \mathbb{K}^{n\times n},\ B \in \mathbb{R}^{n\times n}_{\mathrm{M}}: \quad M(A) \le B \ \Rightarrow\ \alpha(A) \le \alpha(M(A)) \le \alpha(B). \tag{5.16}$$

To this end, note that the spectral abscissa depends continuously on the matrix. By the previous lemma we have $\alpha(A) = \lim_{r\to\infty}(\rho(A + rI_n) - r)$ and $\alpha(M(A)) = \lim_{r\to\infty}\alpha(|A + rI_n| - rI_n)$. For all $r > 0$, equation (5.1) shows

$$\rho(A + rI_n) - r \le \rho(|A + rI_n|) - r.$$

As $|A + rI_n| \ge 0$, the spectral radius equals the spectral abscissa, $\rho(|A + rI_n|) - r = \alpha(|A + rI_n|) - r = \alpha(|A + rI_n| - rI_n)$. Passing to the limit $r \to \infty$ proves $\alpha(A) \le \alpha(M(A))$. The second inequality of (5.16) follows directly from (5.1) since we have for any Metzler matrix $B \in \mathbb{R}^{n\times n}_{\mathrm{M}}$

$$\alpha(B) = \alpha(B + rI_n) - r = \rho(B + rI_n) - r, \quad r \in \{t \ge 0 \,|\, B + tI_n \ge 0\}. \tag{5.17}$$

If $A \in \mathbb{R}^{n\times n}$ is real then it is easy to see that $\|A\|_1 = \|M(A)\|_1$ and $\|A\|_\infty = \|M(A)\|_\infty$, moreover the Gershgorin disks of A and $M(A)$ coincide, $\mathcal{G}(A) = \mathcal{G}(M(A))$, see Theorem 2.45. For a matrix $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ the radii of the Gershgorin disks $R_i = \sum_{j\ne i}|a_{ij}|$ coincide with the radii of $M(A)$, while the centers of the disks may differ only by a purely imaginary number. Corollary 2.48 shows that $\mu_\infty(A) = \max_{s\in\mathcal{G}(A)} \operatorname{Re} s$ such that $\mu_\infty(A) = \mu_\infty(M(A))$. If $\mathcal{G}(A) \subset \mathbb{C}_-$ then the matrix A is strictly diagonally dominant and its Metzler part $M(A)$ is also exponentially stable. We therefore have shown the following result which shows that the definition of $M(A)$ is reasonable.

Proposition 5.35. *Let A be a matrix in $\mathbb{K}^{n\times n}$. Then its initial growth with respect to ∞-norm satisfies $\mu_\infty(A) < 0$ if and only if the Gershgorin set of A is contained in the left half-plane, $\mathcal{G}(A) \subset \mathbb{C}_-$.*

In other words, if the Metzler part of A is strictly diagonally dominant, then A itself is already exponentially stable. The next results further exploit this idea. We consider the initial growth rates associated with monotone vector norms.

Lemma 5.36. *Given $A \in \mathbb{K}^{n\times n}$ and a monotone vector norm $\|\cdot\|$ on $\mathbb{K}^n$.Then the associated initial growth rate satisfies $\mu(A) \le \mu(M(A))$.*

Proof. Setting $r = t^{-1}$ in (2.25) and using Lemma 1.9 gives

$$\begin{aligned}\mu(A) &= \lim_{r\to\infty}(\|A + rI\| - r) \le \lim_{r\to\infty}(\|\,|A + rI|\,\| - r) = \lim_{r\to\infty}(\|\,|A + rI| - rI + rI\| - r)\\ &= \lim_{r\to\infty}(\|M(A) + rI\| - r) = \mu(M(A)).\end{aligned}$$

Hence the initial growth rate of A is bounded from above by the initial growth rate of the Metzler part $M(A)$. □

Note that all p-norms are monotone. Therefore $\mu_2(A) \leq \mu_2(M(A))$. Moreover, for $p = 1, \infty$ we even have equality,

$$\mu_1(A) = \mu_1(M(A)), \quad \mu_\infty(A) = \mu_\infty(M(A)), \qquad A \in \mathbb{K}^{n \times n},$$

which can be directly verified using the formulas of Theorem 2.41.
With every diagonally dominant matrix A we can associate the following diagonally dominant sets.

Proposition 5.37. *Given $A \in \mathbb{K}^{n\times n}$. If $\mathcal{G}(A) \subset \mathbb{C}_-$ then the sets*

$$\mathcal{A}_1 := \left\{ B \in \mathbb{K}^{n\times n} \,\middle|\, \textit{there exists a permutation}\, \pi \textit{ with}\, b_{\pi(i)\pi(i)} = a_{ii} \textit{ and} R_{\pi(i)}(B) \leq R_i(A) \right\},$$
$$\mathcal{A}_2 := \left\{ B \in \mathbb{K}^{n\times n} \,\middle|\, M(B) \leq M(A) \right\}$$

consist of exponentially stable matrices.

Proof. For every $B \in \mathcal{A}_1$ the associated Gershgorin set satisfies $\mathcal{G}(B) \subset \mathcal{G}(A)$, whence by assumption $\mathcal{G}(B) \subset \mathbb{C}_-$. Theorem 2.45 now implies that B is exponentially stable. If $B = (b_{ij}) \in \mathcal{A}_2$ then $\operatorname{Re} b_{ii} \leq \operatorname{Re} a_{ii} < 0$ and $R_i(B') \leq R_i(A)$ for $i = 1, \ldots, n$. Hence $\mu_\infty(B') \leq \mu_\infty(A) < 0$ which shows the exponential stability of B. □

5.6 Transient Bounds for General Matrices

In this section we will first study the relation between the matrix exponential of an arbitrary matrix A and the matrix exponential of its Metzler part $M(A)$. We have already seen that the initial growth rates of A and $M(A)$ coincide for the 1- and ∞-norms. The rest of this section deals with perturbation results for arbitrary matrices based upon the Metzler part. The matrix exponential of the Metzler part provides an upper bound for the matrix exponential of the original matrix A. This fact is established in the following theorem.

Theorem 5.38. *For every $A \in \mathbb{K}^{n\times n}$ and all $t \geq 0$, $\left|e^{At}\right| \leq e^{M(A)t}$ holds elementwise. Moreover,*

$$\left(e^{M(A)t}\right)_{ij} = \inf_{r\in\mathbb{R}} \left(e^{(|A+rI|-rI)t}\right)_{ij}, \qquad t \geq 0, \quad i,j = 1, \ldots, n.$$

Proof. For all $t \geq 0$ and $r \in \mathbb{R}$ we obtain

$$e^{rt}\left|e^{At}\right| = \left|e^{(A+rI)t}\right| \leq \sum_{k=0}^{\infty} \frac{|(A+rI)t|^k}{k!} = e^{|A+rI|t}.$$

The continuity of the matrix exponential and Lemma 5.34 yield the result

$$\left|e^{At}\right| \leq \lim_{r\to\infty} e^{(|A+rI|-rI)t} = e^{M(A)t},\ t \geq 0.$$

Moreover, as the limit in (5.13) is monotone, $\left(e^{M(A)t}\right)_{ij} = \inf_{r\in\mathbb{R}} \left(e^{(|A+rI|-rI)t}\right)_{ij}$ holds componentwise. □

Corollary 5.39. *Given $A \in \mathbb{K}^{n\times n}$. Then the following inequality holds elementwise*

$$\left|(sI-A)^{-1}\right| \leq (\operatorname{Re} s - M(A))^{-1}, \qquad \operatorname{Re} s > \alpha(M(A)) \geq \alpha(A).$$

Proof. The matrix $A - sI$ is exponentially stable for $\operatorname{Re} s > \alpha(M(A)) \geq \alpha(A)$. Hence the integral representation of the resolvent (Corollary 2.9) is well defined and we obtain

$$\left|(sI-A)^{-1}\right| \leq \int_0^\infty \left|e^{(A-sI)t}\right| dt \leq \int_0^\infty e^{(M(A)-\operatorname{Re} sI)t} dt = (\operatorname{Re} sI - M(A))^{-1},$$

the absolute value of the resolvent is bounded componentwise by the resolvent of the Metzler part. □

Hence Metzler matrices A are *exponential positive* and *resolvent positive* in the sense that the matrix exponential e^{At} and the resolvent $(sI - A)^{-1}$ are nonnegative functions for $t > 0, s > 0$. For an operator norm induced by a monotone vector norm the following inequalities hold for any matrix $A \in \mathbb{K}^{n\times n}$, we obtain from Corollary 5.39 and Lemma 1.9

$$\begin{aligned} \left\|e^{At}\right\| &\leq \left\|e^{M(A)t}\right\|, & t &\geq 0, \\ \left\|(sI-A)^{-1}\right\| &\leq \left\|(\operatorname{Re} sI - M(A))^{-1}\right\|, & \operatorname{Re} s &> \alpha(M(A)). \end{aligned} \tag{5.18}$$

In particular, the first equation implies that if $M(A)$ is (M, β)-stable then A is also (M, β)-stable. The following corollary is direct consequence of Theorem 5.38 and of Proposition 5.7.

Corollary 5.40. *Given $A \in \mathbb{K}^{n\times n}$. If there exists a (strict) Liapunov vector for $M(A)$ then A is (exponentially) stable.*

For a set of arbitrary matrices, we can extend Corollary 5.40 to a generalization of Corollary 5.29.

Corollary 5.41. *Given a finite set of matrices $A_1, A_2, \ldots, A_k \in \mathbb{K}^{n\times n}$ such that the Metzler part $M(A_1)$ is exponentially stable. The differential inclusion*

$$\dot{x} \in \operatorname{conv}\{A_i \mid i = 1, \ldots, k\}x \tag{5.19}$$

is exponentially stable if there exists $z > 0$ with $z^\top M(A_1)^{-1} M(A_i) > 0$ for $i = 2, \ldots, k$.

Proof. If $z^\top M(A_1)^{-1} M(A_i) > 0$ holds for all $i = 1, \ldots, k$ then Corollary 5.29 implies that there exists a common Liapunov vector $w > 0$. Therefore w gives a common Liapunov norm $x \mapsto \|\operatorname{diag}(w)x\|_1$ for all $M(A_i)$. Theorem 5.38 shows that this norm is also a common Liapunov norm for the original matrices A_i, $i = 1, \ldots, k$. By Corollary 2.71 the differential inclusion (5.19) is asymptotically stable. □

For a practical use of the results obtained so far, the Metzler part of A should be stable if A is stable.

Remark 5.42. Theorem 5.38 and Lemma 5.3 open the gate to some perturbation results. Interestingly, adding purely imaginary values to the diagonal elements of a Metzler matrix A cannot worsen its transient behaviour,

$$\left|e^{(A+i\Lambda)t}\right| \le e^{M(A+i\Lambda)t} = e^{At}, \text{ where } \Lambda = \operatorname{diag}(\lambda_i),\ \lambda_i \in \mathbb{R}.$$

Moreover, if the Metzler part $M(A)$ of $A \in \mathbb{K}^{n\times n}$ is (M,β)-stable then A itself is (M,β)-stable. By Lemma 5.36 the initial growth rates of A and $M(A)$ satisfy $\mu(A) \le \mu(M(A))$.

We have seen in Proposition 5.37 and in Theorem 5.38 that a Metzler matrix B provides spectral and exponential bounds for all matrices A with $M(A) \le B$. We want to make this statement more precise by introducing suitable perturbation structures.

Suppose that $P \in \mathbb{R}_+^{n\times n}$ is a given nonnegative matrix. Then we define the index set

$$I(P) = \left\{(i,j) \in \{1,\dots,n\}^2 \,\middle|\, p_{ij} > 0\right\},$$

and introduce the following sets of complex perturbation matrices

$$\mathbf{\Delta}_{I(P)} = \left\{\Delta \in \mathbb{C}^{n\times n} \,\middle|\, \Delta_{ij} = 0 \text{ for all } (i,j) \notin I(P)\right\}, \tag{5.20}$$

$$\mathbf{\Delta}_P = \mathbb{C}P, \tag{5.21}$$

both with associated norm $\|\Delta\|_P := \max_{(i,j)\in I(P)} p_{ij}^{-1} |\Delta_{ij}|$. Clearly, $\mathbf{\Delta}_P \subset \mathbf{\Delta}_{I(P)}$. These perturbation structures heavily depend on the coordinate system. The norm has the nice property that for all $\delta > 0$,

$$\|\Delta\|_P < \delta \iff |\Delta| < \delta P. \tag{5.22}$$

For a given stable Metzler matrix $B \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ and a given level $\delta \ge 0$ let us consider the set of all matrices A in $\mathbb{C}^{n\times n}$ which can be written as $A = B + \Delta$ where Δ is a matrix of one of the perturbation structures $(\mathbf{\Delta}_{I(P)}, \|\cdot\|_P)$, $(\mathbf{\Delta}_P, \|\cdot\|_P)$ with $\|\Delta\|_P \le \delta$. We can interpret all these matrices A as perturbations of the Metzler matrix $B \in \mathbb{R}_{\mathrm{M}}^{n\times n}$,

$$B \rightsquigarrow B + \Delta, \qquad \Delta \in \mathbf{\Delta}_{I(P)} \text{ or } \Delta \in \mathbf{\Delta}_P, \text{ and } \|\Delta\|_P < \delta. \tag{5.23}$$

Before we derive explicit formulas for the spectral value sets and the stability radius for these perturbation structures let us recall the following lemma, see [70, Corollary 8.1.29].

Lemma 5.43. *Given $P \in \mathbb{R}_+^{n\times n}$ and a strictly positive vector $x \in \mathbb{R}_+^n$, if $\alpha, \beta \ge 0$ satisfy $\alpha x \le Px \le \beta x$ then $\alpha \le \rho(P) \le \beta$.*

The following result provides a detailed perturbation analysis for the situation of Lemma 5.3.

Theorem 5.44. *Suppose that $P \in \mathbb{R}_+^{n\times n}$ is a given nonnegative matrix. Then the spectral value sets of a Metzler matrix $B \in \mathbb{R}_{\mathrm{M}}^{n\times n}$ corresponding to the levels $\delta \ge 0$ with respect to the perturbation structures $(\mathbf{\Delta}_{I(P)}, \|\cdot\|_P)$ and $(\mathbf{\Delta}_P, \|\cdot\|_P)$ satisfy*

$$\sigma_\delta\left(B \,|\, \mathbf{\Delta}_P\right) = \sigma(B) \cup \left\{s \in \varrho(B) \,\middle|\, \rho(P(\,sI - B\,)^{-1}) > \delta^{-1}\right\}, \tag{5.24}$$

$$\sigma_\delta\left(B \,|\, \mathbf{\Delta}_{I(P)}\right) \subset \sigma(B) \cup \left\{s \in \varrho(B) \,\middle|\, \rho(P\left|(sI - B)^{-1}\right|) > \delta^{-1}\right\}. \tag{5.25}$$

Equality in (5.25) *holds if B is diagonal.*
If B is exponentially stable then the associated stability radii satisfy

$$r\left(B \,\middle|\, \mathbf{\Delta}_{I(P)}\right) = r\left(B \,|\, \mathbf{\Delta}_P\right) = \rho(-PB^{-1})^{-1}. \tag{5.26}$$

Proof. Since $\mathbf{\Delta}_P \subset \mathbf{\Delta}_{I(P)}$ the associated spectral value sets satisfy

$$\sigma_\delta(B \,|\, \mathbf{\Delta}_P) \subset \sigma_\delta(B \,|\, \mathbf{\Delta}_{I(P)}), \qquad \delta \geq 0. \tag{5.27}$$

Let us first derive the formula for the stability radius, hence B is an exponentially stable Metzler matrix. If $\sigma_\delta(B \,|\, \mathbf{\Delta}_{I(P)}) \subset \mathbb{C}_-$ then $M(B + \Delta) \leq B + |\Delta| \leq B + \delta P$ for all $\Delta \in \mathbf{\Delta}_{I(P)}$, $\|\Delta\|_P \leq \delta$, and $B + \delta P$ is exponentially stable. Hence there exists a Liapunov vector $v > 0$ such that $(B + \delta P)v < 0$. Therefore

$$(B + \delta P)v = (I + \delta PB^{-1})(Bv) < 0.$$

As $P \in \mathbb{R}_+^{n\times n}$, v is also a Liapunov vector for B. Setting $w := -Bv > 0$ gives

$$(I - \delta(-PB^{-1}))w > 0, \text{ i.e., } w > \delta(-PB^{-1})w.$$

Now $-PB^{-1} \in \mathbb{R}_+^{n\times n}$ and Lemma 5.43 shows that $1 > \delta\rho(-PB^{-1})$. Thus

$$r\left(B \,\middle|\, \mathbf{\Delta}_{I(P)}\right) = \sup\left\{\delta > 0 \,\middle|\, \exists\, v \in \mathbb{R}_+^n, (B + \delta P)v < 0\right\} \leq \rho(-PB^{-1})^{-1}. \tag{5.28}$$

Let us now introduce $\delta_0 = \rho(-PB^{-1})^{-1}$ and $\Delta_0 = \delta_0 P = \frac{P}{\rho(-PB^{-1})} \in \mathbf{\Delta}_P \subset \mathbf{\Delta}_{I(P)}$. The matrix $-PB^{-1}$ is nonnegative, and its Perron vector w satisfies $-PB^{-1}w = \rho(-PB^{-1})w$. Multiplying $B + \Delta_0$ with $z = B^{-1}w$ gives

$$(B + \Delta_0)z = (B + \delta_0 P)B^{-1}w = w - \delta_0\rho(-PB^{-1})w = 0.$$

Hence $0 \in \sigma_{\delta_0}(B \,|\, \mathbf{\Delta}_P)$, therefore $r(B \,|\, \mathbf{\Delta}_P) \geq \rho(-PB^{-1})^{-1}$. Together with (5.28) and (5.27) this gives the formula for the spectral radius (5.26).
Let us now derive (5.24). If $s \in \varrho(B)$ with $\rho(P(sI - B)^{-1}) > \delta^{-1}$ then there exists an eigenvector $v \in \mathbb{C}^n$ of $P(sI - B)^{-1}$ such that $P(sI - B)^{-1}v = \lambda v$ with $|\lambda| > \delta^{-1} > 0$. Now setting $w = (sI - B)^{-1}v$ gives

$$Pw = \lambda(sI - B)w \quad \text{or} \quad (B + \tfrac{1}{\lambda}P)w = sw.$$

Hence w is an eigenvector of the perturbed matrix $B + \frac{1}{\lambda}P$ corresponding to the eigenvalue $s \in \mathbb{C}$. Now, $\Delta = \frac{1}{\lambda}P \in \mathbf{\Delta}_P$ has norm $\|\Delta\|_P < \delta$ from which $s \in \sigma_\delta(B \,|\, \mathbf{\Delta}_P)$ follows. On the other hand, if $s \in \sigma_\delta(B \,|\, \mathbf{\Delta}_P) \setminus \sigma(B)$ then there exist $\Delta \in \mathbb{C}^{n\times n}$ and $v \in \mathbb{C}^n$ such that $(B - sI + \Delta)v = 0$ and $\eta\Delta = P$ for some $\eta \in \mathbb{C}$ with $|\eta|^{-1} < \delta$. Now

$$(B - sI + \Delta)v = \left(I - \Delta(sI - B)^{-1}\right)(B - sI)v = 0, \tag{5.29}$$

hence $I - \Delta(sI - B)^{-1}$ is not invertible, and therefore $\rho(\Delta(sI - B)^{-1}) \geq 1$. We conclude from $\eta\Delta = P$ that $\rho(\Delta(sI - B)^{-1}) = |\eta|^{-1}\rho(P(sI - B)^{-1}) \geq 1$, and therefore $\rho(P(sI - B)^{-1}) > \delta^{-1}$ for all $s \in \sigma_\delta(B \,|\, \mathbf{\Delta}_P)$. Thus (5.24) holds.

To show (5.25), note that (5.29) holds for an $s \in \sigma_\delta(A \,|\, \mathbf{\Delta}_{I(P)})$. Hence there exists $\Delta \in \mathbf{\Delta}_{I(P)}$ with $|\Delta| < \delta P$ such that $\rho(\Delta(sI-B)^{-1}) \geq 1$. Taking advantage of the monotonicity of the spectral radius, we have

$$1 \leq \rho\left(\Delta(sI-B)^{-1}\right) \leq \rho\left(|\Delta|\left|(sI-B)^{-1}\right|\right) < \delta\rho\left(P\left|(sI-B)^{-1}\right|\right),$$

which shows $\rho(P\,|(sI-B)^{-1}|) > \delta^{-1}$ for all $s \in \sigma_\delta(B \,|\, \mathbf{\Delta}_{I(P)})$. Therefore the inclusion (5.25) holds. Additionally, if B is diagonal then the missing inclusion "$\supset$" in (5.25) follows from the construction of a suitable perturbation matrix $\Delta \in \mathbf{\Delta}_{I(P)}$. To this end, if $\rho := \rho(P\,|(sI-B)^{-1}|) > \delta^{-1}$ holds for a given $s \in \varrho(B)$ then there exists a vector $v \in \mathbb{R}^n_+$ such that $P\,|(sI-B)^{-1}|\,v = \rho v$ with $\rho^{-1} < \delta$. Let us introduce $R = (sI-B)^{-1}$ and the vectors $w = Rv$, $\tilde{w} = |R|\,v$. Then the matrix

$$\Delta = \left(p_{ij}\frac{\tilde{w}_j}{w_j}\right)_{ij} = \left(p_{ij}\frac{|(s-b_{jj})^{-1}|\,v_j}{(s-b_{jj})^{-1}v_j}\right)_{ij} \in \mathbf{\Delta}_{I(P)}$$

satisfies $\|\Delta\|_P = 1$ and $\Delta Rv = P\,|R|\,v$. Now $B + \Delta/\rho$ has an eigenvector corresponding to $s \in \mathbb{C}$ given by $x = Rv = (sI-B)^{-1}v$,

$$\left(B + \tfrac{1}{\rho}\Delta\right)x = Bx + \tfrac{1}{\rho}\Delta Rv = Bx + \tfrac{1}{\rho}P\left|(sI-B)^{-1}\right|v = Bx + v = (B + sI - B)x = sx.$$

Therefore $s \in \sigma_\delta(B \,|\, \mathbf{\Delta}_{I(P)})$. Hence equality holds in (5.25) if B is diagonal. □

Example 5.45. Consider the stable Metzler matrix $B \in \mathbb{R}^{3\times 3}_{\mathrm{M}}$ and the nonnegative matrix $P \in \mathbb{R}^{3\times 3}_+$ given by

$$B = \begin{pmatrix} -8 & 10 & 0 \\ 1 & -8 & 6 \\ 0 & 2 & -10 \end{pmatrix}, \qquad P = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}.$$

Figure 5.3 shows the spectral value sets of B corresponding to $\mathbf{\Delta}_P$ (solid lines) and an upper bound of the spectral value sets corresponding to $\mathbf{\Delta}_{I(P)}$ (dashed lines and gray-shaded areas) for the levels $\delta \in \{\frac{1}{10}, \frac{1}{3}, \frac{2}{3}, 1\}$. Both contours differ substantially around $s = -8.87$ while the difference is not apparent for small $\delta > 0$ near the other eigenvalues of B. The stability radius of B with respect to both perturbation structures induced by P is given by $r = \rho(PB^{-1}) = 1.02$. And indeed, the contour level for $\delta = 1$ is still contained in $\mathbb{C}_-$. ■

All of these perturbed matrices satisfy a common transient bound.

Proposition 5.46. *Let $B \in \mathbb{R}^{n\times n}_M$, $P \in \mathbb{R}^{n\times n}_+$ and $\delta > 0$. If the vector $w > 0$ is a Liapunov vector of $B + \delta P$ with $v = -(B+\delta P)w \geq 0$ then for all $\Delta \in \mathbf{\Delta}_{I(P)}$ with $\|\Delta\|_P < \delta$,*

$$\left\|e^{(B+\Delta)t}\right\|_\infty \leq \kappa(w)e^{-t\min_i(\frac{v_i}{w_i})}, \quad t \geq 0.$$

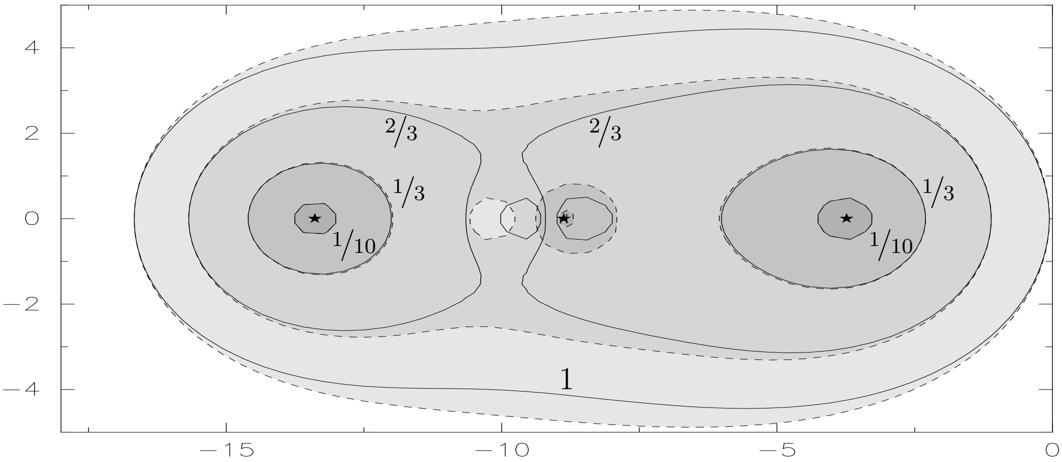

Figure 5.3: SVS associated with $\mathbf{\Delta}_P$ and upper bounds for SVS associated with $\mathbf{\Delta}_{I(P)}$.

Proof. As $|\Delta| < \delta P$, we obtain using Theorem 5.38

$$\left|e^{(B+\Delta)t}\right| \leq e^{M(B+\Delta)t} = e^{(B+M(\Delta))t} \leq e^{(B+|\Delta|)t} \leq e^{(B+\delta P)t}.$$

Now $\|\cdot\|_\infty$ is monotone, hence Proposition 5.7 implies that

$$\left\|e^{(B+\Delta)t}\right\|_\infty \leq \left\|e^{(B+\delta P)t}\right\|_\infty \leq \kappa(w)e^{-t\min_i(\frac{v_i}{w_i})}, \quad t \geq 0,$$

holds for the Liapunov vector w which proves the proposition. □

5.7 Notes and References

Positive systems arise naturally in applications like economics, biology, chemistry, and numerical analysis. Their study has been an active field of research for many decades, including works like Varga [140], Berman and Plemmmons [17], Krause and Nesemann [86], and Farina and Rinaldi [40]. The study of transient effects, however, has been neglected in the literature.

For results on Metzler matrices see Fiedler and Ptak [42], Luenberger [101], and Horn and Johnson [71]. Proofs of Gershgorin's Disk Theorem can be found in standard references like Horn and Johnson [70] or Faddeev and Faddeeva [39]. For a more functional analytic approach than these direct proofs see Bhatia [18].

If w is a left Liapunov vector of $A \in \mathbb{R}^{n\times n}_{\mathrm{M}}$ then the function $x \mapsto w^\top x$ is also called a *copositive* Liapunov function of A, see Mason and Shorten [106].

Vector-valued Liapunov functions for the stability analysis have been used in Bellman [14] Willems [149], and Kiendl et al. [83]. In Polanski [116] a polytopic vector norm (polyhedral Liapunov function) is optimized using a linear programming approach. This can be viewed as an extension of finding a weight with optimal eccentricity.

The special role of the 1- and ∞-norms for positive systems was noted by Vidyasagar [142]. The convexity of μ_∞ is used in Liu and Molchanov [95] to derive a common Liapunov function for nonlinear systems

$$\dot{x}(t) = A(t)x(t) + BN(Cx(t), t)$$

where $A(t) \in \operatorname{conv}\{A_1, \ldots, A_q\} \subset \mathbb{R}^{n \times n}$ and the nonlinearity N satisfies a sector condition. An investigation of the properties of the stability radius for positive systems can be found in articles of Hinrichsen and Son [69] and Fischer, Hinrichsen and Son [43]. Hinrichsen, Karow and Pritchard [61] study perturbation structures which resemble (5.20). The results obtained therein are derived via μ-analysis and not directly as in our result of Theorem 5.44.

Chapter 6

Differential Delay Systems

This chapter will be devoted to the study of linear differential delay systems of the form

$$\Sigma : \qquad \dot{x}(t) = A_0 x(t) + \sum_{k=1}^{m} A_k x(t - h_k), \qquad t \geq 0, \tag{6.1}$$

where $A_k \in \mathbb{C}^{n \times n}$ and $0 < h_1 < h_2 < \ldots h_m = H$ are given positive *delays*. For $t = 0$, (6.1) only fixes the one-sided differential $\dot{x}(0+) = \lim_{h \searrow 0} \frac{1}{h}(x(h) - x(0))$, which has to satisfy $\dot{x}(0+) = A_0 x(0) + \sum_{k=1}^{m} A_k x(-h_k)$. To specify an initial value problem which has a unique solution, an initial function with values on the interval $[-H, 0]$ has to be prescribed.
We will demonstrate some problems in the following example.

Example 6.1. We consider the "hot shower problem", see Kolmanovskii and Myshkis [84],

$$\dot{x}(t) = -\alpha x(t - h), \qquad \alpha > 0, h > 0, \quad t > 0, \tag{6.2}$$

which can be seen as a simple feedback controller where the current feedback is based on an old state of the system. With a "human in the loop", see Figure 6.1, this corresponds to the problem of stabilizing the output of a hot shower using a mixer tab: If the water is too hot, the mixer is turned to cool and vice versa. But the water currently leaving the shower is not influenced by this decision. Depending on the length of the pipes only the temperature of water arriving sooner or later at the shower is controlled.
To solve (6.2) we have to prescribe a initial value function on the interval $[-h, 0]$. Let us use a linear ramp from $\varphi(-h) = -1$ to $\varphi(0) = 1$. Figure 6.2 shows two solutions, one with $\alpha = 1$, $h = 1$ and the other with $\alpha = 1$, $h = 2$. From these pictures we can expect that the first system is stable, while the latter is not. Asides from the stability question, we want to find bounds on the transient behaviour of such a delay system. ■

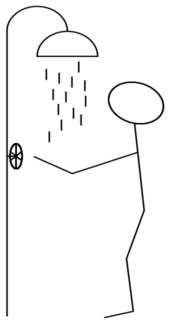

Figure 6.1: Taking a hot shower.

The properties of differential delay systems have been studied in, e.g., Bellman and Cooke [15], Hale and Verduyn Lunel [51], and Curtain and Zwart [29].

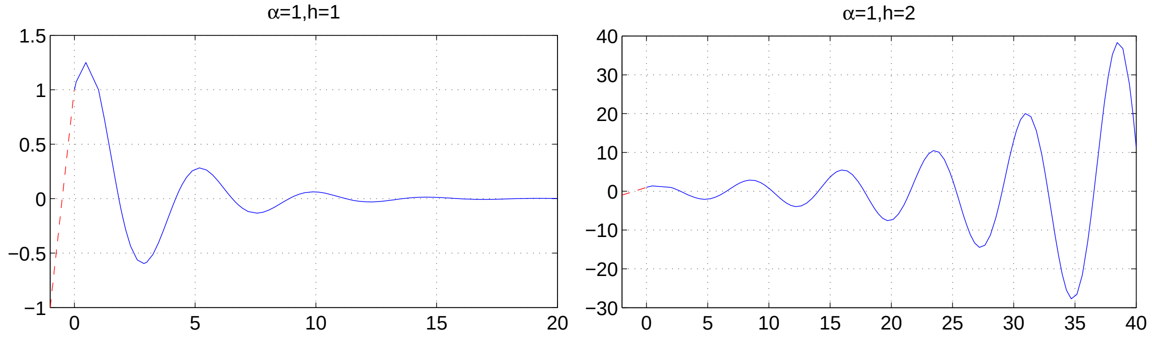

Figure 6.2: Stability or instability of the hot shower problem.

In the following we will derive transient estimates for solutions of (6.1) on the basis of Liapunov functionals which now operate on solution segments. Before we approach the construction of such functionals let us formulate a precise notion of solutions for (6.1) with respect to a suitable initial value problem. We introduce fundamental matrices and show how the solutions of (6.1) can be represented with their help. We show that the solutions of the delay equation are a semigroup on some suitable Hilbert space.

6.1 Functional Analytic Approach

We study the following initial value problem associated with (6.1),

$$\begin{aligned} \dot{x}(t) &= A_0 x(t) + \sum_{k=1}^{m} A_k x(t-h_k), && t \geq t_0, \\ x(t_0) &= x_0, \\ x(t) &= \varphi(t-t_0), && t_0 - H \leq t < t_0, \end{aligned} \tag{6.3}$$

where $x_0 \in \mathbb{C}^n$ and $\varphi \in L^2([-H,0],\mathbb{C}^n)$. The following proposition shows existence and uniqueness for such an initial value problem of the delay system.

Proposition 6.2 ([29, Theorem 2.4.1])**.** *For every $x_0 \in \mathbb{C}^n$ and $\varphi \in L^2([-H,0],\mathbb{C}^n)$ there exists a unique function $x(\cdot)$ which is absolutely continuous on bounded intervals of* $[t_0,\infty)$ *and satisfies the differential equation in* (6.3) *almost everywhere. This function is called the* solution *of the initial value problem* (6.3) *with respect to the initial data x_0 and φ and is denoted by $x(\cdot;t_0,x_0,\varphi)$. It satisfies*

$$x(t;t_0,x_0,\varphi) = e^{A_0(t-t_0)}x_0 + \sum_{k=1}^{m} \int_{t_0}^{t} e^{A_0(t-s)} A_k x(s-h_k)ds, \qquad t \geq t_0. \tag{6.4}$$

Notice that the system (6.1) is time-invariant so that we fix $t_0 = 0$, if not noted otherwise. To keep the notation short we introduce $z = (x_0,\varphi) \in \mathbb{C}^n \times L^2([-H,0],\mathbb{C}^n)$ and set $x(t,z) := x(t;0,x_0,\varphi)$.

Let us denote the space of continuous vector functions on $[-H,0]$ by $C = C([-H,0],\mathbb{C}^n)$ which is endowed with the sup-norm, $\|\varphi\|_\infty = \sup_{\theta\in[-H,0]}\|\varphi(\theta)\|$. To include both the initial value x_0 and the initial function φ mentioned in Proposition 6.2 into a suitable space, we define $M^2 = M^2([-H,0],\mathbb{C}^n) = \mathbb{C}^n \times L^2([-H,0],\mathbb{C}^n)$ to be the space of pairs of vectors and L^2-integrable functions on $[-H,0]$. This space becomes a Hilbert space using the inner product of the direct sum, see p. 12,

$$\left\langle \begin{pmatrix} x \\ f \end{pmatrix}, \begin{pmatrix} y \\ g \end{pmatrix} \right\rangle_{M^2} := \langle x,y\rangle_2 + \langle f,g\rangle_{L^2([-H,0],\mathbb{C}^n)} = \langle x,y\rangle_2 + \int_{-H}^{0} \langle f(\theta), g(\theta)\rangle_2 d\theta. \tag{6.5}$$

In the following we discuss the solutions of (6.3) with respect to initial values $z = (x_0,\varphi) \in M^2$ and with respect to continuous initial values where $\varphi \in C$ and $x_0 = \varphi(0)$. For stability issues we note the following definition.

Definition 6.3. The delay equation (6.1) is called *exponentially stable* if there exist constants $M \geq 1$ and $\beta < 0$ such that for all continuous initial conditions $\varphi \in C$ we have

$$\|x(t;0,\varphi(0),\varphi)\|_2 \leq Me^{\beta t}\|\varphi\|_\infty, \qquad t \geq 0. \tag{6.6}$$

The exponential stability of a delay equation (6.1) can be verified by considering the associated characteristic equation.

Definition 6.4. The function $\chi : \mathbb{C} \to \mathbb{C}$ given by $\chi(s) = \det(sI - A_0 - \sum_{k=1}^m A_k e^{-sh_k})$ is called the *characteristic function* of (6.1), and the equation $\chi(s) = 0$ is called the characteristic equation of (6.1).

The complex value s is a solution of the characteristic equation $\chi(s) = 0$ if and only if there exists a non-trivial vector $x_0 \in \mathbb{C}^n$ such that $(sI - A_0 - \sum_{k=1}^m A_k e^{-sh_k})x_0 = 0$. In this case a non-trivial solution of (6.3) is given by $e^{st}x_0$, $t \geq 0$, which corresponds to the initial segment $e^{st}x_0$, $t \in [-H,0]$. Here $x_0 \neq 0$ is called an *eigenvector* of the system Σ in (6.1) . The special solution $e^{st}x_0$ is called an *eigenmotion* of the delay equation (6.1).

Proposition 6.5 ([131]). *The delay equation* (6.1) *is exponentially stable if and only if* $\{s \in \mathbb{C} \,|\, \operatorname{Re} s \geq 0,\ \chi(s) = 0\} = \emptyset$.

Let us now define an equivalent of the matrix exponential for the delay equation (6.1). Consider the following initial value problem for a matrix delay equation,

$$\begin{aligned} &\dot{K}(t) = A_0K(t) + \sum_{k=1}^m A_kK(t-h_k), \qquad t \geq 0, \\ &K(0) = I_n, \quad K(t) = 0_n \qquad \text{for } t < 0. \end{aligned} \tag{6.7}$$

Here the derivative of K in 0 is to be understood as the one-sided derivative, $\dot{K}(0) = \lim_{t\searrow 0}\dot{K}(t)$. If K solves (6.8) then it is easy to see that the columns of K are solutions of (6.3) corresponding to an initial value $z_i = (e_i, 0) \in M^2$, $K(t)e_i = x(t,z_i)$, $t \geq 0$, where $e_i \in \mathbb{C}^n$ is the i-th unit vector, $i = 1,\ldots,n$. Hence by Proposition 6.2 this solution K exists on $\mathbb{R}_+$ and is uniquely determined.

Definition 6.6. The matrix function $K : \mathbb{R} \to \mathbb{C}^{n\times n}$ which satisfies the initial value problem (6.7) is called the *fundamental matrix* of (6.1).

The following properties hold for the fundamental matrix.

Lemma 6.7 ([81])**.** *The fundamental matrix K of* (6.1) *is a continuous matrix function for $t > 0$. Moreover, it is exponentially bounded. In addition to* (6.7) *it also satisfies the following initial value problem where the A_k and K terms are exchanged,*

$$\begin{aligned}\dot{K}(t) &= K(t)A_0 + \sum_{k=1}^{m} K(t-h_k)A_k, \qquad t \geq 0,\\ K(0) &= I_n, \qquad K(t) = 0_n, \quad t < 0.\end{aligned}$$

We can represent any solution of (6.1) in closed form using the fundamental matrix. One can easily verify the following result using (6.1) and (6.7).

Corollary 6.8 ([15, Theorem 6.4])**.** *The solution $x(\cdot, z)$ of* (6.3) *with $z = (x_0, \varphi) \in M^2$ is given by*

$$x(t,z) = K(t)x_0 + \sum_{k=1}^{m} \int_{-h_k}^{0} K(t-h_k-\theta)A_k\varphi(\theta)d\theta, \qquad t \geq 0. \tag{6.8}$$

By Proposition 6.2 there exists a uniquely determined solution of (6.3) for every initial value $z = (x_0, \varphi) \in M^2$.

Definition 6.9. Let $x(\cdot, z)$ be the solution of (6.3) with initial value $z = (x_0, \varphi) \in M^2$. Then the corresponding *solution segment* for $t > 0$ is given by the function

$$x_t(z) \in L^2([-H,0],\mathbb{C}^n), \quad (x_t(z))(\tau) = \begin{cases} x(t+\tau, z), & t+\tau \geq 0\\ \varphi(t+\tau), & t+\tau < 0,\end{cases} \qquad \tau \in [-H, 0].$$

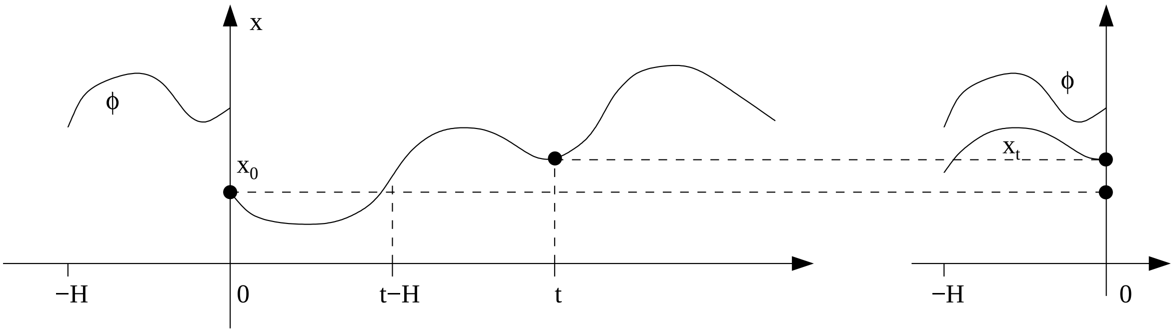

Figure 6.3: Initial segment and solution segment.

When it is clear from the context we drop the dependence on the initial segment. Figure 6.3 illustrates the definitions in the scalar case. We immediately obtain from the definition the following "smoothing" property for the solutions of (6.3).

Lemma 6.10. *If $x(t,z)$ is a solution of (6.3) for the initial value $z = (x_0, \varphi) \in M^2$ then the solution segment $x_t(z)(\cdot) : \tau \mapsto x(t+\tau, z)$ is continuous on $[-H, 0]$ for $t \geq H$. If the initial segment $\varphi \in C$ is already continuous and $\varphi(0) = x_0$ the solution segment $x_t \in C$ is a continuous function for all $t \geq 0$.*

More precisely, if $\varphi \in L^2$ and $x_0 \in \mathbb{C}^n$ then $t \mapsto x(t; 0, x_0, \varphi)$, $t \geq 0$, is by definition an absolute continuous function; if φ is of class C^k on $[-H, 0]$, $k \in \mathbb{N}$, then $t \mapsto x(t; 0, \varphi(0), \varphi)$ is of class C^{k+1} on $\mathbb{R}_+$ which follows from formula (6.4).
Each continuous segment $\varphi \in C$ has an M^2-equivalent given by $\hat{\varphi} = (\varphi(0),\ \varphi)$. On the other hand, given an initial segment $z \in M^2$, the solution segment $x_t(z)$ is continuous for $t \geq H$, see Lemma 6.10. Hence we have a map $M^2 \to C$ given by $z \mapsto x_H(z)$. We may use this continuous segment as a new initial function.

Lemma 6.11. *For a given initial value $z \in M^2$ the segment $\psi = x_H(z)$ is continuous on $[-H, 0]$. The associated solution $x(\cdot, \hat{\psi})$ of (6.1) satisfies $x_{t+H}(z) = x_t(\hat{\psi})$.*

Let us now show how continuous initial segments fit into an M^2-framework.

Proposition 6.12. *The map $C([-H, 0], \mathbb{C}^n) \to M^2([-H, 0], \mathbb{C}^n)$, $\varphi \mapsto \hat{\varphi} := (\varphi(0), \varphi)$ defines a continuous dense embedding from $C([-H, 0], \mathbb{C}^n)$ into $M^2([-H, 0], \mathbb{C}^n)$.*

Proof. As $\|\hat{\varphi}\|^2_{M^2} = \|\varphi(0)\|^2_2 + \|\varphi\|^2_{L^2} \leq (1+H)\, \|\varphi\|^2_\infty$ for all $\varphi \in C$ we see that this embedding is continuous. To show that $\varphi \mapsto \hat{\varphi}$ is dense, we construct for a given $(x,\ f) \in M^2$ a sequence of continuous segments $\varphi_n \in C = C([-H, 0], \mathbb{C}^n)$ with $\varphi_n \xrightarrow{L^2} f$ and $\varphi_n(0) = x$. As $C([-H, 0], \mathbb{C}^n)$ is dense in $L^2([-H, 0], \mathbb{C}^n)$ there exists a sequence of continuous segments $f_n \in C$ with $f_n \xrightarrow{L^2} f$. Moreover, let us define the sequences of continuous functions,

$$x_n(t) = \begin{cases} 0, & t \in [-H, -\frac{1}{n}], \\ (1+nt)x, & t \in [-\frac{1}{n}, 0], \end{cases} \quad \text{and} \quad g_n(t) = \begin{cases} 1, & t \in [-H, -\frac{1}{n}], \\ -nt, & t \in [-\frac{1}{n}, 0]. \end{cases}$$

Then $\varphi_n = g_n f_n + (1 - g_n) x_n$ is a sequence of continuous functions with $\varphi_n \xrightarrow{L^2} f$ and $\varphi_n(0) = x$ for all $n \in \mathbb{N}$. Hence the continuous segments are dense in M^2. □

We can associate a strongly continuous semigroup $(T(t))_{t \in \mathbb{R}_+}$ on M^2 with the solutions of (6.1), see [29, Theorem 2.4.4]. This *solution semigroup* is given by

$$T(t) : M^2 \to M^2 : z = \begin{pmatrix} x_0 \\ \varphi \end{pmatrix} \mapsto \hat{x}_t(z) = \begin{pmatrix} x_t(z)(0) \\ x_t(z)(\cdot) \end{pmatrix} = \begin{pmatrix} x(t,z) \\ x_t(z)(\cdot) \end{pmatrix}, \qquad t \geq 0. \tag{6.9}$$

Theorem 6.13 ([29, Theorem 2.4.6])**.** *The generator of the semigroup T of (6.9) is given by*

$$A \begin{pmatrix} x \\ f \end{pmatrix} = \begin{pmatrix} A_0 x + \sum_{k=1}^m A_k f(-h_k) \\ \frac{df}{dt} \end{pmatrix}, \qquad \begin{pmatrix} x \\ f \end{pmatrix} \in D(A), \tag{6.10}$$

with domain

$$D(A) = \left\{ \binom{x}{f} \in M^2([-H, 0], \mathbb{C}^n) \,\middle|\, f \textit{ is abs. cont.}, \tfrac{df}{dt} \in L^2([-H, 0], \mathbb{C}^n) \textit{ and } f(0) = x \right\}.$$

The spectrum of A consists only of eigenvalues, it is a discrete subset of $\mathbb{C}$, and it is given by the solutions of the characteristic equation,

$$\sigma(A) = \{s \in \mathbb{C} \,|\, \chi(s) = 0\}\,. \tag{6.11}$$

The multiplicity of every eigenvalue of A is finite. For every $\alpha \in \mathbb{R}$ there are only finitely many eigenvalues of A in $\{s \in \mathbb{C} \,|\, \operatorname{Re} s > \alpha\}$.

For the interpretation of the semigroup operation $t \mapsto T(t)z = x_t(z)$ on M^2 as a solution of an abstract Cauchy problem compare with Lemma 2.3 and Proposition 2.11.
When the initial segment is already continuous the setup of the abstract Cauchy problem reduces to the solution semigroup $S(t) : C \to C : \varphi \mapsto x_t(\varphi)$. Its generator is given by

$$A_C : C \to C : \varphi \mapsto \tfrac{d}{dt}\varphi, \qquad \varphi \in D(A_C),$$

with domain

$$D(A_C) = \left\{ \varphi \in C^1([-H, 0], \mathbb{C}^n) \;\middle|\; \tfrac{d}{dt}\varphi(0) = A_0\varphi(0) + \sum_{k=1}^{m} A_k \varphi(-h_k) \right\},$$

see [38, Example II.3.29].
The following proposition shows equivalent conditions for the exponential stability of (6.1).

Proposition 6.14. *The following statements are equivalent.*

(i) The delay equation (6.1) *is exponentially stable.*

(ii) For all $s \in \mathbb{C}$, $\operatorname{Re} s \geq 0$, the characteristic function of (6.1) *satisfies $\chi(s) \neq 0$.*

(iii) The C-solution semigroup $(S(t))_{t \in \mathbb{R}_+}$ is exponentially stable.

(iv) The M^2-solution semigroup $(T(t))_{t \in \mathbb{R}_+}$ is exponentially stable.

(v) There exist constants $M \geq 1$ and $\beta < 0$ such that $\|K(t)\|_2 \leq Me^{\beta t}$ for all $t \geq 0$.

Proof. The equivalence of *(i)* and *(ii)* is due to Proposition 6.5. Now, [29, Theorem 5.1.7] shows that *(ii)* and *(iv)* are equivalent. The implication *(v)* $\Longrightarrow$ *(iii)* follows directly from formula (6.8). If *(iii)* is satisfied then $\|x(t, \hat{\varphi})\|_2 \leq \|x_t(\hat{\varphi})\|_\infty = \|S(t)\varphi\|_\infty \leq Me^{\beta t}\|\varphi\|_\infty$ holds for all continuous $\varphi \in C$. Thus *(iii)* implies *(i)*.
To round up the proof we now show *(i)* $\Longrightarrow$ *(v)*. For this, we assume that (6.1) is exponentially stable. There exists a sequence of continuous segments $(\varphi_k) \subset C$ for a given $v \in \mathbb{R}^n$ such that $\lim_{k\to\infty} \varphi_k(0) = v$, $\lim_{k\to\infty} \varphi_k(t) = 0$ for $t \in [-H, 0)$, and $\|\varphi_K\|_\infty = \|v\|_2$. Then for all $k = 1, 2, \ldots$ we have $\|x(t, \hat{\varphi}_k)\|_2 \leq Me^{\beta t}\|\varphi_k\|_\infty = Me^{\beta t}\|v\|_2$, $t \geq 0$. In the limit $k \to \infty$ we obtain $\|K(t)v\|_2 \leq Me^{\beta t}\|v\|$, $t \geq 0$, from Lebesgue's dominated convergence theorem. □

6.2 Liapunov Functionals

Quadratic Liapunov functions provide means of analysing the behaviour of linear delay-free ordinary differential equations, cf. Section 3.4. For delay systems, Liapunov functions depend on solution segments. We have seen on Proposition 6.14 that the M^2-solution semigroup T of (6.1) is exponentially stable if and only if the delay system (6.1) is exponentially stable.

6.2.1 Liapunov Equations in Hilbert Spaces

In the following we want to check this stability property using Liapunov techniques. For this, let us recall the notion of an abstract Liapunov equation, see Curtain and Zwart [29, Theorem 5.1.3, Exercise 5.3].

Theorem 6.15. *Given a generator A of a strongly continuous semigroup $(T(t))_{t\in\mathbb{R}_+}$ on a Hilbert space X, then T is exponentially stable if and only if there exist a coercive self-adjoint linear operator $P \in \mathfrak{L}(X)$ and $\varepsilon > 0$ such that*

$$\langle Ax, Px\rangle + \langle Px, Ax\rangle < -\varepsilon\langle x, x\rangle \qquad \textit{for all}\, x \in D(A) \setminus \{0\}. \tag{6.12}$$

Moreover, if T is exponentially stable then for every coercive self-adjoint linear operator $Q \in \mathfrak{L}(X)$ the coercive self-adjoint linear operator $P \in \mathfrak{L}(X)$ given by

$$P = \int_0^\infty T(t)^* Q T(t) dt \tag{6.13}$$

satisfies (6.12).

Here, (6.13) is the solution of the *Liapunov equation*

$$\langle Ax, Px\rangle + \langle Px, Ax\rangle = -\langle x, Qx\rangle \qquad \text{for all } x \in D(A) \setminus \{0\}. \tag{6.14}$$

for the operator A. The following proof draws heavily from the machinery developed in Chapter 2.

Proof. We only show that (6.12) implies exponential stability of T, as it is easy to see that if $Q \in \mathfrak{L}(X)$ is coercive and T is exponentially stable then P defined by (6.13) is a bounded coercive operator which satisfies (6.14) and therefore also (6.12).
Let P be a coercive bounded operator. The norm $\|x\|_P = \sqrt{\langle x, Px\rangle}$ is a norm on X for which there exist $\alpha, \beta > 0$ such that

$$\alpha\langle x, x\rangle < \|x\|_P^2 \leq \beta\langle x, x\rangle \qquad \text{for all} \quad x \in X \setminus \{0\}. \tag{6.15}$$

With respect to this norm, the initial growth rate of A is given by

$$\mu_P(A) = \tfrac{1}{2} \sup_{x\in D(A), x\neq 0} \frac{\langle Ax, Px\rangle + \langle Px, Ax\rangle}{\langle x, Px\rangle} = \tfrac{1}{2} \sup_{x\in D(A), x\neq 0} \frac{-\langle x, Qx\rangle}{\langle x, Px\rangle} \leq -\frac{\varepsilon}{2\beta} < 0,$$

see Definition 2.29. Hence A is strictly dissipative with respect $\|\cdot\|_P$, thus generates a uniform contraction semigroup T on $(X, \|\cdot\|_P)$. Now, by (6.15) the operator norms $\|\cdot\|$ and $\|\cdot\|_P$ are equivalent on $\mathcal{L}(X)$. We conclude that A generates an exponentially stable semigroup on $(X, \|\cdot\|)$, see also Corollary 2.57. □

As a general assumption for the rest of this chapter we consider only those delay equations (6.1) for which the matrices $A_k \in \mathbb{R}^{n\times n}$, $k = 0, 1, \ldots, m$ in (6.1) are all *real*. We will derive an explicit formula for the solution of a Liapunov equation for the generator A of the solution semigroup T of (6.9). Let us assume that this semigroup is exponentially stable. Hence there exist $M \geq 1$ and $\beta < 0$ such that the M^2-operator norm satisfies $\|T(t)\|_{M^2} \leq Me^{\beta t}$.

Definition 6.16. Suppose that the solution semigroup associated with (6.1) is exponentially stable. For a given positive definite matrix $W \in \mathcal{H}^n_+(\mathbb{R})$ we define the *delay Liapunov function* of (6.1) by

$$U : \mathbb{R} \to \mathbb{R}^{n\times n}, \qquad U(t) = \int_0^\infty K(\tau)^\top W K(t+\tau) d\tau, \tag{6.16}$$

where $K(\cdot)$ is the fundamental matrix of (6.1), see Definition 6.6.

This integral is well-defined if T is exponentially stable as the fundamental matrix K is decaying exponentially for $|t| \to \infty$, see Proposition 6.14. Hence the integral in (6.16) is bounded,

$$\begin{aligned}\left\|\int_0^\infty K(\tau)^\top W K(t+\tau) d\tau\right\| &\leq \int_0^\infty \|W\| \, \|K(\tau)\| \, \|K(t+\tau)\| \, d\tau \\ &\leq e^{\beta t} \|W\| \int_0^\infty (Me^{\beta\tau})^2 d\tau = \|W\| \, M^2 \frac{e^{\beta t}}{-2\beta}.\end{aligned}$$

The name "delay Liapunov function" owes to the fact that U takes over the role of a classical quadratic Liapunov matrix for delay-free systems. In particular, if (6.1) is a differential equation without delays, i.e., of the form $\dot{x} = A_0 x$ then the fundamental matrix is just the matrix exponential, $K(t) = e^{A_0 t}, t \geq 0$, and (6.16) reduces to $U(t) = U(0)e^{A_0 t}$ where $P := U(0)$ satisfies

$$P = \int_0^\infty e^{A_0^\top \tau} W e^{A_0 \tau} d\tau.$$

This is the classical explicit formula of the solution of the quadratic Liapunov equation $PA_0 + A_0^\top P = -W$ where $W \in \mathcal{H}^n(\mathbb{R})$ is a positive definite matrix.
We now collect some properties of U.

Lemma 6.17. *Suppose that T is exponentially stable. Then the matrix function $U(t)$ defined by* (6.16) *is continuous, decaying exponentially, and satisfies the* symmetry condition *$U(t) = U(-t)^\top$ for all $t \in \mathbb{R}$.*

Proof. We have already seen that U is exponentially bounded as $t \to \infty$ with a negative growth rate β. For the continuity of U on $\mathbb{R}_+$, note that we have for all $t > 0$ and all $\varepsilon \in (0, t)$ that $U(t+\varepsilon) - U(t) = \int_0^\infty K(\tau)^\top W\left(K(t+\varepsilon+\tau) - K(t+\tau)\right) d\tau \to 0$ as $\varepsilon \to 0$. The *symmetry condition* can be shown by applying the integral transformation $\theta = \tau - t$ to (6.16),

$$U(-t) = \int_0^\infty K(\tau)^\top W K(\tau - t) d\tau = \int_0^\infty K(\theta + t)^\top W K(\theta) d\theta = U(t)^\top. \tag{6.17}$$

This symmetry property implies that $U(t)$ is continuous for $t \leq 0$, hence for all $t \in \mathbb{R}$. □

Let us now introduce the M^2-operator Q for which we will construct an explicit solution P of the Liapunov equation (6.14) associated with the M^2-generator A. For given symmetric weights $W_0, W_H \in \mathcal{H}^n(\mathbb{R})$ we set $W = W_0 + H\,W_H$ in (6.16) and define the operator $Q : M^2 \to M^2$ via

$$Q\begin{pmatrix} x \\ f \end{pmatrix} = \begin{pmatrix} W_0 x \\ W_H f \end{pmatrix}. \tag{6.18}$$

Lemma 6.18. *If the weight matrices $W_0, W_H \succ 0$ are both positive definite then Q defined by* (6.18) *is a bounded self-adjoint coercive linear operator.*

Proof. We have

$$\left\langle \tbinom{x}{f}, Q\tbinom{x}{f} \right\rangle_{M^2} = \langle x, W_0 x\rangle_2 + \langle f, W_h f\rangle_{L^2} = \langle W_0 x, x\rangle_2 + \langle W_h f, f\rangle_{L^2} = \left\langle Q\tbinom{x}{f}, \tbinom{x}{f} \right\rangle_{M^2}.$$

Hence Q is symmetric, and

$$\min\{\lambda_{\min}(W_0), \lambda_{\min}(W_H)\} \left\| \tbinom{x}{f} \right\|_{M^2}^2 \leq \left\langle \tbinom{x}{f}, Q\tbinom{x}{f} \right\rangle_{M^2} \leq \max\{\lambda_{\max}(W_0), \lambda_{\max}(W_H)\} \left\| \tbinom{x}{f} \right\|_{M^2}^2$$

shows that Q is bounded and coercive. □

The candidate $P : M^2 \to M^2$ for the solution of the Liapunov equation 6.14 is partitioned as follows

$$P\begin{pmatrix} x \\ f \end{pmatrix} = \begin{pmatrix} U(0)x + P_1 f \\ P_1^* x + P_2 f \end{pmatrix}. \tag{6.19}$$

Let us now discuss its components. If $1_k = 1_{[-h_k, 0]}$ denotes the characteristic function of the interval $[-h_k, 0]$ then the linear operators $P_1 : L^2 \to \mathbb{R}^n$ and $P_2 : L^2 \to L^2$ are defined by

$$\begin{aligned} P_1 f &= \int_{-H}^0 \sum_{j=1}^m 1_j(\theta) U(-h_j - \theta) A_j f(\theta) d\theta, \\ (P_2 f)(t) &= \sum_{k=1}^m 1_k(t) A_k^\top \int_{-H}^0 \sum_{j=1}^m U(t - \theta + h_k - h_j) A_j 1_j(\theta) f(\theta) d\theta + (H + t) W_H f(t). \end{aligned} \tag{6.20}$$

It is not difficult to prove that these operators are bounded. With this definitions we now check that the operator P is a self-adjoint bounded linear operator with respect to the M^2-inner product.

Lemma 6.19. *The matrix $U(0) \in \mathbb{R}^{n\times n}$ is a symmetric matrix, P_2 is a bounded self-adjoint linear operator on L^2. P_1 is a bounded linear operator, its adjoint $P_1^* : \mathbb{R}^n \to L^2$ is given by*

$$(P_1^* x)(t) = \sum_{k=1}^{m} 1_k(t) A_k^\top U(t+h_k)x. \tag{6.21}$$

Hence P is a bounded self-adjoint linear operator on M^2.

Proof. For $t = 0$, (6.17) takes the form $U(0) = U(0)^\top$, thus $U(0)$ is symmetric. Let us denote the right hand side of (6.21) by $\tilde{P}_1$. For every $x \in \mathbb{C}^n$ and every $f \in L^2$ this operator $\tilde{P}_1$ satisfies

$$\begin{aligned}\left\langle f, \tilde{P}_1 x\right\rangle_{L^2} &= \int_{-H}^{0} \left(\sum_{k=1}^{m} 1_k(\tau) A_k^\top U(\tau+h_k)x\right)^* f(\tau)d\tau \\ &= x^* \int_{-H}^{0} \sum_{k=1}^{m} 1_k(\tau) U(-\tau - h_k) A_k f(\tau) d\tau = \langle P_1 f, x\rangle_2,\end{aligned}$$

where we used that $U(-t) = U(t)^\top = U(t)^*$. Hence the adjoint of P_1 is $P_1^* = \tilde{P}_1$. The domains of P_1 and P_1^* are given by $D(P_1) = L^2$ and $D(P_1^*) = \mathbb{R}^n$. With the same symmetry argument for U we can prove that P_2 is a symmetric operator on L^2,

$$\begin{aligned}\langle P_2 f, g\rangle_{L^2} &= \int_{-H}^{0} g(t)^* \sum_{k=1}^{m} 1_k(t) A_k^\top \int_{-H}^{0} \sum_{j=1}^{m} U(t-\theta+h_k-h_j) A_j 1_j(\theta) f(\theta) d\theta dt \\ &\quad + \int_{-H}^{0} g(t)^* (H+t) W_H f(t) dt \\ &= \int_{-H}^{0} \left(\sum_{j=1}^{m} 1_j(\theta) A_j^\top \int_{-H}^{0} \sum_{k=1}^{m} U(\theta - t + h_j - h_k) A_k 1_k(t) g(t) dt\right)^* f(\theta) d\theta \\ &\quad + \int_{-H}^{0} \left((H+t) W_H g(t)\right)^* f(t) dt \\ &= \langle f, P_2 g\rangle_{L^2},\end{aligned}$$

where we changed the order of summation and integration. Hence P is a symmetric operator on M^2. The boundedness of P follows from the boundedness of its components. □

We now use the integral representation (6.16) of U to show that for a given Q of the form (6.18), the operator P defined in (6.19) solves the Liapunov equation (6.13) associated with the generator A in M^2.

Theorem 6.20. *If T of* (6.9) *is exponentially stable and P and Q are given by* (6.19) *and* (6.18) *where W_0 and W_H are Hermitian matrices, then*

$$\int_0^\infty \langle \hat{x}_t(z), Q\hat{x}_t(z)\rangle_{M^2} dt = \langle z, Pz\rangle_{M^2}, \qquad z = (x_0,\ \varphi) \in M^2. \tag{6.22}$$

Proof. The integral of the inner product $\langle \hat{x}_t, Q\hat{x}_t\rangle_{M^2}$ is given by

$$\int_0^\infty \left(x(t)^* W_0 x(t) + \int_{-H}^0 x(t+\theta)^* W_H x_t(t+\theta) d\theta \right) dt. \tag{6.23}$$

Let us study its first term, $\int_0^\infty x(t)^* W_0 x(t)\, dt$. Using (6.8) and then sorting for different quadratic and mixed terms we get

$$\begin{aligned}
\int_0^\infty x(t)^* W_0 x(t)\, dt &= \int_0^\infty \left(K(t)x_0 + \sum_{k=1}^m \int_{-h_k}^0 K(t-h_k-\theta_1) A_k \varphi(\theta_1) d\theta_1 \right)^* \\
&\quad \cdot W_0 \left(K(t)x_0 + \sum_{j=1}^m \int_{-h_j}^0 K(t-h_j-\theta_2) A_j \varphi(\theta_2) d\theta_2 \right) dt \\
&= \int_0^\infty x_0^* K(t)^\top W_0 K(t) x_0 dt + 2\mathrm{Re} \int_0^\infty x_0^* K(t)^\top W_0 \sum_{k=1}^m \int_{-h_k}^0 K(t-h_k-\theta) A_k \varphi(\theta) d\theta\, dt \\
&\quad + \int_0^\infty \left(\sum_{k=1}^m \int_{-h_k}^0 K(t-h_k-\theta_1) A_k \varphi(\theta_1) d\theta_1 \right)^* \\
&\quad \cdot W_0 \left(\sum_{j=1}^m \int_{-h_j}^0 K(t-h_j-\theta_2) A_j \varphi(\theta_2) d\theta_2 \right) dt \qquad (6.24) \\
&= x_0^* U^0(0) x_0 + 2\mathrm{Re}\, x_0^* \sum_{k=1}^m \int_{-h_k}^0 U^0(-h_k-\theta) A_k \varphi(\theta)\, d\theta \\
&\quad + \sum_{k=1}^m \int_{-h_k}^0 \varphi(\theta_1)^* A_k^\top \sum_{j=1}^m \int_{-h_j}^0 U^0(h_k+\theta_1-h_j-\theta_2) A_j \varphi(\theta_2)\, d\theta_2\, d\theta_1,
\end{aligned}$$

where $U^0(\tau) = \int_0^\infty K(t)^\top W_0 K(t+\tau)\, dt$. Here we used a parameter transformation to get

$$\int_0^\infty K(t-\tau_1)^\top W_0 K(t-\tau_2)\, dt = \int_0^\infty K(t)^\top W_0 K(t+\tau_1-\tau_2)\, dt = U^0(\tau_1-\tau_2) \tag{6.25}$$

for $\tau_1, \tau_2 \in \mathbb{R}$. Let us now discuss the second term of (6.23), $\int_0^\infty \int_{-H}^0 x(t+\tau)^* W_H x(t+\tau) d\tau dt$. When changing the order of integration one has to take into account that $x(t) = \varphi(t)$ if $t \in [-H, 0)$, so that

$$\int_0^\infty \int_{-H}^0 x(t+\theta)^* W_H x(t+\theta) d\theta dt = \int_{-H}^0 \left(\int_\theta^0 \varphi(t)^* W_H \varphi(t) dt + \int_0^\infty x(t)^* W_H x(t) dt \right) d\theta. \tag{6.26}$$

Again, changing the order of integration in the first term of the right hand side of (6.26) gives

$$\begin{aligned}
\int_{-H}^0 \int_\theta^0 \varphi(t)^* W_H \varphi(t)\, dt d\theta &= \int_{-H}^0 \int_{-H}^0 1_{[\theta,0]}(t) \varphi(t)^* W_H \varphi(t)\, dt d\theta \\
&= \int_{-H}^0 \varphi(t)^* W_H \varphi(t) \left(\int_{-H}^0 1_{[-H,t]}(\theta) d\theta \right) dt = \int_{-H}^0 (H+t) \varphi(t)^* W_H \varphi(t)\, dt. \qquad (6.27)
\end{aligned}$$

The second term in (6.26) is independent of θ, and combined with (6.27) we arrive at the following expression for (6.26)

$$\int_{-H}^{0}(H+t)\varphi(t)^*W_H\varphi(t)\,dt + H\int_{0}^{\infty} x(t)^*W_H x(t)\,dt.$$

However, an analogous term to $\int_0^\infty x(t)^*W_Hx(t)dt$ has already been treated in (6.24). Hence replacing $U^0(\tau)$ in (6.24) with $U^H(\tau) := \int_0^\infty K(t)^\top W_H K(t+\tau)\,dt$ and using this in (6.26), we obtain for (6.26)

$$\begin{aligned}\int_0^\infty\int_{-H}^0 x(t+\tau)^*W_Hx(t+\tau)\,d\tau\,dt &= \int_{-H}^0 (H+\tau)\varphi(\tau)^*W_H\varphi(\tau)d\tau + H\Big(x_0^*U^H(0)x_0\\ &\quad + 2\mathrm{Re}\,x_0^*\sum_{k=1}^m\int_{-h_k}^0 U^H(-h_k-\theta)A_k\varphi(\theta)\,d\theta\\ &\quad + \sum_{k=1}^m\int_{-h_k}^0\varphi(\theta_1)^*A_k^\top\sum_{j=1}^m\int_{-h_j}^0 U^H(h_k-h_j+\theta_1-\theta_2)A_j\varphi(\theta_2)d\theta_2d\theta_1\Big).\end{aligned} \tag{6.28}$$

Returning to (6.23), we get by summing (6.24) and (6.28)

$$\begin{aligned}\int_0^\infty\langle\hat{x}_t,Q\hat{x}_t\rangle_{M^2}\,dt &= x_0^*U(0)x_0 + 2\mathrm{Re}\,x_0^*\sum_{k=1}^m\int_{-h_k}^0 U(-h_k-\theta)A_k\varphi(\theta)\,d\theta\\ &\quad+\sum_{k=1}^m\int_{-h_k}^0\varphi(\theta_1)^*A_k^\top\sum_{j=1}^m\int_{-h_j}^0 U(h_k-h_j+\theta_1-\theta_2)A_j\varphi(\theta_2)d\theta_2d\theta_1\\ &\quad+\int_{-H}^0\varphi(\tau)^*(H+\tau)W_H\varphi(\tau)d\tau,\end{aligned} \tag{6.29}$$

where $U(\tau) = U^0(\tau) + HU^H(\tau) = \int_0^\infty K(t)^\top(W_0+HW_H)K(t+\tau)\,dt$.
Now, we have to identify (6.29) as the M^2-inner product weighted with P. We evaluate this inner product using (6.19) and (6.20),

$$\begin{aligned}\langle Pz,z\rangle_{M^2} &= x_0^*\left(U(0)x_0+P_1\varphi\right) + \int_{-H}^0\varphi(t)^*\left((P_1^*x_0)(t)+(P_2\varphi)(t)\right)dt\\ &= x_0^*U(0)x_0 + x_0^*\sum_{k=1}^m\int_{-h_k}^0 U(-h_k-\theta)A_k\varphi(\theta)d\theta\\ &\quad+\sum_{k=1}^m\int_{-h_k}^0\varphi(t)^*A_k^\top U(t+h_k)x_0dt\\ &\quad+\Bigg(\sum_{k=1}^m\int_{-h_k}^0\varphi(t)^*A_k^\top\sum_{j=1}^m\int_{-h_j}^0 U(t-\theta+h_k-h_j)A_j\varphi(\theta)\,d\theta\,dt\\ &\quad+\int_{-H}^0\varphi(t)^*(H+t)W_H\varphi(t)\Bigg)dt.\end{aligned} \tag{6.30}$$

When we compare (6.30) with (6.29), and recall that U satisfies the symmetry condition $U(t) = U^\top(-t)$, we see that both expressions are identical, hence (6.22) holds. □

From Theorem 6.20 we conclude that P satisfies (6.13).

Corollary 6.21. *Given the generator A of the exponentially stable solution semigroup of* (6.1) *in M^2, then the operators P and Q of* (6.19) *and* (6.18) *associated with Hermitian weights W_0 and W_H satisfy the Liapunov equation*

$$\langle PAz, z\rangle_{M^2} + \langle z, PAz\rangle_{M^2} = -\langle z, Qz\rangle_{M^2}, \qquad z \in D(A). \tag{6.31}$$

Especially, the derivative of the functional $v : D(A) \to \mathbb{R}_+, z \mapsto v(z) = \langle z, Pz\rangle_{M^2}$ along trajectories of the abstract Cauchy problem (2.6) *is given by*

$$\dot v(z) := \lim_{t\searrow 0} \tfrac{1}{t}(v(\hat x_t(z)) - v(z)) = -\langle z, Qz\rangle_{M^2}, \qquad z \in D(A). \tag{6.32}$$

Proof. Let us first recall that by Proposition 2.10, $z \in D(A)$ implies that $t \mapsto x_t(z) = T(t)z$ is a differential function for all $t \in \mathbb{R}_+$ which satisfies $\frac{d}{dt}x_t(z) = Ax_t(z)$. Moreover for $z \in D(A)$, $T(t)Az = AT(t)z$ for all $t \in \mathbb{R}_+$. Then by Theorem 6.20,

$$\begin{aligned}\langle Pz, Az\rangle_{M^2} + \langle Az, Pz\rangle_{M^2} &= \left\langle \int_0^\infty T(t)^*QT(t)z\,dt, Az\right\rangle_{M^2} + \left\langle Az, \int_0^\infty T(t)^*QT(t)z\,dt\right\rangle_{M^2}\\ &= \int_0^\infty \langle Qx_t(z), Ax_t(z)\rangle_{M^2} + \langle Ax_t(z), Qx_t(z)\rangle_{M^2}\,dt\\ &= \int_0^\infty \langle Qx_t(z), \dot x_t(z)\rangle_{M^2} + \langle \dot x_t(z), Qx_t(z)\rangle_{M^2}\,dt\\ &= \int_0^\infty \tfrac{d}{dt}\langle T(t)z, QT(t)z\rangle_{M^2}\,dt = [\langle T(t)z, QT(t)z\rangle_{M^2}]_{t=0}^\infty = -\langle z, Q, z\rangle_{M^2}.\end{aligned}$$

Hence, $\dot v(z)$ equals $-\langle z, Qz\rangle$ on $z \in D(A)$. □

If the weights W_0 and W_H are positive definite, then P satisfies the Liapunov inequality (6.12) in M^2 because Q is coercive.

6.2.2 Liapunov-Krasovskii Functionals

We now want to derive transient estimates for the solutions of the delay system (6.1). For this we introduce the notion of a Liapunov-Krasovkii functional.

Definition 6.22. A continuous functional $v : M^2 \to \mathbb{R}_+$ is called a *Liapunov-Krasovskii functional* for the delay equation (6.1) if it has the following properties

(i) There exist $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1 \|x_0\|_2^2 \le v(z) \le \alpha_2 \|z\|_{M^2}^2 \qquad \text{for all } z = (x_0, \varphi) \in M^2. \tag{6.33}$$

(ii) For $z \in D(A)$ the derivative $\dot{v}(z) = \lim_{t \searrow 0} \frac{1}{t}\left(v(\hat{x}_t(z)) - v(z)\right)$ along solutions of (6.1) exists, and there exists a constant $\beta < 0$ such that $\dot{v}(z) \leq 2\beta v(z)$.

Theorem 6.23. *Suppose that $v : M^2 \to \mathbb{R}_+$ is a Liapunov-Krasovskii functional satisfying* (i) *and* (ii) *in Definition 6.22. Then the delay system* (6.1) *is exponentially stable and satisfies the exponential estimate*

$$\|x(t,z)\|_2 \leq \sqrt{\tfrac{\alpha_2}{\alpha_1}} e^{\beta t} \|z\|_{M^2}, \qquad z \in M^2,\ t \geq 0. \tag{6.34}$$

On the other hand, if (6.1) *is exponentially stable then for every given pair of positive definite matrices $W_0, W_H \in \mathcal{H}^n_+(\mathbb{R})$ the functional $v(z) = \langle z, Pz \rangle_{M^2}$ defined by* (6.30) *is a Liapunov-Krasovskii functional for* (6.1) *where P is defined in* (6.19).

Proof. By definition, the Liapunov-Krasovskii functional v satisfies $\dot{v}(\hat{x}_t) \leq 2\beta v(\hat{x}_t)$ for all solutions $\hat{x}_t = \hat{x}_t(z)$, $t \geq 0$, with initial value $z = (x_0, \varphi) \in D(A)$. Then the derivative[1] of $e^{-2\beta t} v(x_t)$ is given by

$$\tfrac{d}{dt}\left(e^{-2\beta t} v(\hat{x}_t)\right) = e^{-2\beta t}\left(\dot{v}(\hat{x}_t) - 2\beta v(\hat{x}_t)\right) \leq 0,$$

so that $v(\hat{x}_t) \leq e^{2\beta t} v(z)$. By (6.33) we obtain for $z \in D(A)$

$$\alpha_1 \|x(t,z)\|_2^2 \leq v(\hat{x}_t) \leq e^{2\beta t} v(z) \leq e^{2\beta t} \alpha_2 \|z\|_{M^2}^2, \qquad t \geq 0.$$

Now, v is a continous functional on M^2 and $D(A)$ is dense in M^2. Hence (6.34) holds for all $z \in M^2$ and the delay system (6.1) is exponentially stable by Definition 6.3.
Conversely, if the delay system is exponentially stable we show that the functional $v(z) = \langle z, Pz \rangle_{M^2}$ of Theorem 6.20 is a Liapunov-Krasovskii functional. If W_0 and W_H are positive definite then the operator $Q \in \mathcal{L}(M^2)$ is a coercive self-adjoint operator, and Theorem 6.15 shows that P is also a coercive and bounded linear operator. Thus there exist constants $\alpha_1, \alpha_2 > 0$, $\beta_1, \beta_2 > 0$ such that

$$\alpha_1 \|z\|_{M^2}^2 \leq \langle z, Pz \rangle_{M^2} \leq \alpha_2 \|z\|_{M^2}^2, \qquad \beta_1 \|z\|_{M^2}^2 \leq \langle z, Qz \rangle_{M^2} \leq \beta_2 \|z\|_{M^2}^2.$$

Clearly, $\alpha_1 \|x_0\|_2^2 \leq \alpha_1 \|z\|_{M^2}$ for $z = (x_0, \varphi) \in M^2$, and therefore (6.33) is satisfied. Since $\dot{v}(z) = -\langle z, Qz \rangle_{M^2}$ for $z \in D(A)$ by Corollary 6.21 we have $\dot{v}(z) \geq -\beta_2 \|z\|_{M^2}^2 \geq -\frac{\beta_2}{\alpha_1} v(z)$. Hence $v(z) = \langle z, Pz \rangle_{M^2}$ is a Liapunov-Krasovskii functional for (6.1). □

Theorem 6.23 shows that the existence of a Liapunov-Krasovskii functional defined in Definition 6.22 provides a necessary and sufficient condition for the exponential stability of the solution semigroup T of the delay equation (6.1).

Remark 6.24. Using the terminology of Chapter 2, the inequalities in (6.43) provide an estimate for the eccentricity of the quadratic functional $v(z) = \langle z, Pz \rangle_{M^2}$ compared to the M^2-norm $\|\cdot\|_{M^2} = \sqrt{\langle \cdot, \cdot \rangle_{M^2}}$. An optimal value of β in Definition 6.22 (ii) corresponds

[1] For $t = 0$ this derivative is one-sided.

to the initial growth rate of the generator A (6.10) with respect to the weighted norm $\nu_P(z) = \sqrt{\langle z, Pz\rangle_{M^2}}$, as for $v(z) = \nu_P(z)^2$ we get from Proposition 2.31 that

$$\mu(A) = \sup_{z\in D(A)\setminus\{0\}} \operatorname{Re} \frac{\langle z, PAz\rangle_{M^2}}{\langle z, Pz\rangle_{M^2}} = \sup_{z\in D(A)\setminus\{0\}} \frac{-\langle z, Qz\rangle_{M^2}}{2\langle z, Pz\rangle_{M^2}} = \tfrac{1}{2} \sup_{z\in D(A)\setminus\{0\}} \frac{\dot{v}(z)}{v(z)}.$$

With Corollary 2.17 we have $\mu(A) = \inf\{\beta \in \mathbb{R} \,|\, \text{for all } z \in D(A),\ \dot{v}(z) \le 2\beta v(z)\}$.

We can also consider Liapunov-Krasovskii functionals which operate on continuous segments. Let us define the following continuous counterpart to Definition 6.22.

Definition 6.25. A continuous functional $v : C \to \mathbb{R}^+$ is a Liapunov-Krasovskii functional for (6.1) if the following properties hold

(i) There exist $\alpha_1, \alpha_2 > 0$ such that for all $\varphi \in C$, $\alpha_1 \|\varphi(0)\|_2^2 \le v(\varphi) \le \alpha_2 \|\varphi\|_\infty^2$, where $\|\varphi\|_\infty = \sup_{t\in[-h,0]} \|\varphi(t)\|_2$.

(ii) The derivative along solutions $\dot{v}(\varphi)$ exists for all $\varphi \in D(A_C)$, and there exists $\beta < 0$ such that $\dot{v}(\varphi) \le 2\beta v(\varphi)$.

We obtain the following counterpart to Theorem 6.23.

Corollary 6.26. *Let $v : C \to \mathbb{R}^+$ be a Liapunov-Krasovskii functional satisfying Definition 6.25* (i) *and* (ii)*. Then the delay system* (6.1) *is exponentially stable. Its solutions satisfy the exponential estimate*

$$\|x(t,\varphi)\|_2 \le \sqrt{\tfrac{\alpha_1}{\alpha_2}} e^{\beta t} \|\varphi\|_\infty, \qquad \varphi \in C, t \ge 0. \tag{6.35}$$

On the other hand, if (6.1) *is exponentially stable, then for every given pair of positive definite matrices $W_0, W_H \in \mathcal{H}_+^n(\mathbb{R})$ the functional $v(\varphi) = \langle \hat{\varphi}, P\hat{\varphi}\rangle_{M^2}$ is a Liapunov-Krasovskii functional on C for* (6.1) *where P is defined in* (6.19)*.*

Proof. The proof of the exponential estimate (6.35) follows analogously to (6.34). We will only show that $\varphi \mapsto v(\varphi) = \langle \hat{\varphi}, P\hat{\varphi}\rangle_{M^2}$ is a Liapunov-Krasovskii functional on C. Let us consider a continuous segment $\varphi \in C$. The associated M^2-segment $\hat{\varphi} = (\varphi(0),\ \varphi) \in M^2$ then satisfies the following inequalities

$$\begin{aligned} \|\hat{\varphi}\|_{M^2}^2 &= \|\varphi(0)\|_2^2 + \|\varphi\|_{L^2}^2 \le \|\varphi(0)\|_2^2 + H\|\varphi\|_\infty^2 \le (1+H)\|\varphi\|_\infty^2, \\ \|\hat{\varphi}\|_{M^2}^2 &= \|\varphi(0)\|_2^2 + \|\varphi\|_{L^2}^2 \ge \|\varphi(0)\|_2^2, \end{aligned}$$

so that $\alpha_1 \|\hat{\varphi}\|_{M^2}^2 \le v(\varphi) \le \alpha_2 \|\hat{\varphi}\|_{M^2}^2$ implies that $\alpha_1 \|\varphi(0)\|_2^2 \le v(\varphi) \le \alpha_2(1+H)\|\varphi\|_\infty^2$. The functional $\varphi \mapsto \langle \hat{\varphi}, P\hat{\varphi}\rangle_{M^2}$ is a continuous function for all $\varphi \in C$ and satifies (6.34), hence also (6.35) (with different constants) for $\varphi \in C$. Hence it is a Liapunov-Krasovskii functional on C. □

Note that $\varphi \mapsto P\hat{\varphi}$ gives rise to a continuous function, i.e., $U(0)\varphi(0) + P_1(\varphi) = (P_1^*\varphi(0) + P_2\varphi)(0)$ is satisfied, if and only if $\varphi \in D(A_C)$.

6.2.3 Complete Type Liapunov-Krasovskii Functionals

In a series of articles [81, 79, 78], V. Kharitonov and co-authors study so-called *complete type* Liapunov-Krasovskii functionals $v : C([-H, 0], \mathbb{R}^n) \to \mathbb{R}_+$, for which the the derivative along trajectories, $\dot{v}(\varphi) = -w(\varphi)$, takes the following form

$$w(\varphi) = \varphi(0)^\top R_0 \varphi(0) + \sum_{k=1}^{m} \varphi(-h_k)^\top R_k \varphi(-h_k) + \sum_{k=1}^{m} \int_{-h_k}^{0} \varphi(\theta)^\top R_{m+k} \varphi(\theta) d\theta, \qquad (6.36)$$

where $R_k \in \mathcal{H}_+^n(\mathbb{R})$ are given positive definite weights. If $\varphi \in C$ is a real continuous segment then $\varphi \mapsto \langle \hat{\varphi}, Q\hat{\varphi} \rangle_{M^2}$ is of the form (6.36) with $R_0 = W_0$, $R_{2m} = W_H$, and $R_1 = \cdots = R_{2m-1} = 0$. But this breaks the requirement of positive definite weights. However, we have seen in the previous discussion that $\varphi \mapsto \langle \hat{\varphi}, P\hat{\varphi} \rangle_{M^2}$ is a Liapunov-Krasovskii functional, hence we do not need positive definite weights $R_1, \ldots, R_{2m-1}$ in (6.36).

Proposition 6.27. *For every complete type Liapunov-Krasovskii functional $v : C \to \mathbb{R}_+$ there exist weights W_0 and W_H such that the coercive operators $P, Q : M^2 \to M^2$ given by* (6.19) *and* (6.18) *satisfy for all continuous segments $\varphi \in C$*

$$v(\varphi) \geq \langle \hat{\varphi}, P\hat{\varphi} \rangle_{M^2} \quad \text{and} \quad \dot{v}(\varphi) \leq -\langle \hat{\varphi}, Q\hat{\varphi} \rangle_{M^2}.$$

Proof. The complete type functional v is induced by a quadratic functional w given by (6.36). Setting $W_0 = R_0$ and $W_H = R_{2m}$ we get $Q\hat{\varphi} = (W_0\varphi(0), W_H\varphi) \in M^2$. Clearly, $\langle \hat{\varphi}, Q\hat{\varphi} \rangle_{M^2} \leq w(\varphi) = -\dot{v}(\varphi)$. Now, by Theorem 6.20,

$$\langle \hat{\varphi}, P\hat{\varphi} \rangle_{M^2} = \int_0^\infty \langle T(t)\hat{\varphi}, QT(t)\hat{\varphi} \rangle_{M^2} \, dt \leq \int_0^\infty w(x_t(\hat{\varphi})) \, dt = v(\varphi),$$

where the last equality follows from the construction of complete type Liapunov-Krasovskii functionals, see [81]. □

We can modify the operator Q to account for more terms of the complete type functional. To this end, we replace the matrix W_H with an operator $W : [-H, 0] \to \mathcal{H}_+^n(\mathbb{R}^n)$, given by $W(t) = \sum_{k=1}^m 1_k(t) W_k$ with W_k positive definite. The multiplication of $W(t)$ with $f \in L^2([-H, 0], R^n)$ is defined pointwise. Then for $Q\binom{x}{f} = \binom{W_0 x}{t \mapsto W(t)f(t)}$ we have

$$\begin{aligned} \langle \hat{\varphi}, Q\hat{\varphi} \rangle_{M^2} &= \varphi(0)^* W_0 \varphi(0) + \int_{-H}^0 f(t)^* \sum_{k=1}^m 1_k(t) W_k f(t) dt \\ &= \varphi(0)^* W_0 \varphi(0) + \sum_{k=1}^m \int_{-h_k}^0 f(t)^* W_k f(t) dt. \end{aligned}$$

Hence all integral terms in (6.36) can be reconstructed by introducing a time-varying positive definite matrix $W(t)$, while the weighted point-delays associated with the weights $R_1, \ldots, R_m$ cannot be embedded into an M^2-framework.

6.3 Existence and Uniqueness of Delay Liapunov Matrices

In the current and following sections we present an analysis of the properties of the delay Liapunov matrix based upon a finite-dimensional approach.
We have seen in (6.30) that the delay Liapunov matrix $U(t)$ is the building block in the construction of $\langle\hat{\varphi}, P\hat{\varphi}\rangle_{M^2}$. However, the integral representation $U(t) = \int_0^\infty K(\tau)^\top W K(t+\tau)d\tau$ cannot be used for the numerical computation of the delay Liapunov matrix. We therefore present an alternative characterization of U. The following description of $U(t)$ has been given in Datko [32] for the one-delay case.

Proposition 6.28. *Suppose that* (6.1) *is exponentially stable. The delay Liapunov matrix* $U : \mathbb{R} \to \mathbb{R}^{n\times n}$ *given by* (6.16) *is a function which is differentiable on* $[0,\infty)$ *and satisfies the following matrix delay differential equation*[2]

$$\dot{U}(t) = U(t)A_0 + \sum_{k=1}^{m} U(t-h_k)A_k, \quad t \geq 0, \tag{6.37}$$

and the conditions

$$U(t) = U(-t)^\top, \qquad t \leq 0, \tag{6.38}$$

$$U(0)A_0 + A_0^\top U(0) + \sum_{k=1}^{m} \left(U(h_k)^\top A_k + A_k^\top U(h_k)\right) = -W. \tag{6.39}$$

The condition (6.38) is called the *symmetry condition* as it implies $U(0) = U(0)^\top$, while (6.39) is called the *algebraic condition* associated with the weight W. Using the one-sided derivative of U in $t = 0$ we can rewrite (6.39) as $\dot{U}(0) + \dot{U}(0)^\top = -W$.

Proof. We will verify that the improper integral (6.16) satisfies the delay equation (6.37) and the additional conditions (6.38) and (6.39). The integral is well-defined for all $t \in \mathbb{R}$ because (6.1) is exponentially stable. By Lemma 6.7 we have for $t \geq 0$

$$\begin{aligned}
\tfrac{d}{dt}U(t) &= \int_0^\infty K(\tau)^\top W \tfrac{d}{dt}K(t+\tau)d\tau \\
&= \int_0^\infty K(\tau)^\top W \left(K(t+\tau)A_0 + \sum_{k=1}^{m} K(t+\tau-h_k)\right) d\tau \\
&= \int_0^\infty K(\tau)^\top W K(t+\tau)A_0 dt + \sum_{k=1}^{m} \int_0^\infty K(\tau)^\top W K(t+\tau-h_k)A_k d\tau \\
&= U(t)A_0 + \sum_{k=1}^{m} U(t-h_k)A_k,
\end{aligned}$$

[2] In $t = 0$ we require that the one-sided derivative satisfies $\dot{U}(0+) = U(0)A_0 + \sum_{k=1}^m U(-h_k)A_k$.

whence U satisfies the differential delay equation (6.37). Again, as the differential equation for $K(t)$ does not hold for $t < 0$ and is only one-sided in $t = 0$, we have $\dot{U}(0) = \lim_{t\searrow 0} \dot{U}(t) = U(0)A_0 + \sum_{k=1}^m U(-h_k)A_k$. The symmetry condition (6.38) has been shown in (6.17). Using the symmetry condition (6.38) we rewrite (6.39) as

$$U(0)A_0 + A_0^\top U(0)^\top + \sum_{k=1}^m \left(U(-h_k)A_k + A_k^\top U(-h_k)^\top\right)$$

$$= \int_0^\infty K(\tau)^\top W \left(K(\tau)A_0 + \sum_{k=1}^m K(\tau - h_k)A_k\right) + \left(K(\tau)A_0 + \sum_{k=1}^m K(\tau - h_k)A_k\right)^\top WK(\tau)d\tau$$

$$= \int_0^\infty K(\tau)^\top W \left(\tfrac{d}{d\tau}K(\tau)\right) + \left(\tfrac{d}{d\tau}K(\tau)\right)^\top WK(\tau)d\tau = -K(0)^\top WK(0) = -W,$$

since $\lim_{\tau\to\infty} K(\tau) = 0$ by exponential stability of (6.1). □

For $t = 0$, equation (6.38) shows that $U(0)$ is symmetric. Then the left hand side of (6.39) can be written by $\dot{U}(0) + (\dot{U}(0))^\top$. Moreover, (6.39) is satisfied with $W = 0$ if and only if $\dot{U}(0) = -(\dot{U}(0))^\top$, i.e., $U(0)$ is *skew-symmetric.*

The solutions of (6.37),(6.38), and (6.39) may also be obtained in the following way, see Louisell [96] for the one-delay case.

Proposition 6.29. *Consider the* transfer matrix *of the delay system* (6.1) *given by*

$$G(s) = \left(sI - A_0 - \sum_{k=1}^m e^{-sh_k}A_k\right)^{-1}, \qquad s \in \varrho(A). \tag{6.40}$$

If $i\mathbb{R} \subset \varrho(A)$ then the integral

$$V(t) = \frac{1}{2\pi}\int_{-\infty}^\infty G(i\omega)^* W G(i\omega)e^{i\omega t}d\omega, \qquad t \in \mathbb{R}, \tag{6.41}$$

is well-defined. If the delay equation (6.1) *is exponentially stable then $V(t) = U(t)$ for all $t \in \mathbb{R}$. Hence V satisfies* (6.37),(6.38), *and* (6.39).

Proof. Let us first show that $V(t)$ is well-defined if $\sigma(A)$ does not contain purely imaginary roots. By taking norms in (6.41) and using $\|G(i\omega)\| = \|G(i\omega)^*)\| = \|G(-i\omega)\|$, we have $\|V(t)\| \le \pi^{-1}\|W\| \int_0^\infty \|G(i\omega)\|_2^2 d\omega$. Now $\|G(i\omega)\|$ satisfies

$$\|G(i\omega)\| = |\omega|^{-1}\left\|\left(I_n - \tfrac{1}{i\omega}A_0 - \sum_{k=1}^m \tfrac{1}{i\omega}e^{-i\omega h_k}A_k\right)^{-1}\right\|, \qquad \omega \neq 0. \tag{6.42}$$

For $\omega \to \infty$ the right factor in (6.42) approaches 1, hence $\omega \mapsto G(i\omega)$ is bounded on $\mathbb{R}$ by continuity. Therefore there exists a constant $M > 0$ such that for all $\omega \in \mathbb{R}$ with $|\omega| > 1$, $\|G(i\omega)\| \le M|\omega|^{-1}$. Thus for all $t \in \mathbb{R}$,

$$\|V(t)\| \le \tfrac{1}{\pi}\|W\|\left(\int_0^1 \|G(i\omega)\|^2 d\omega + \int_1^\infty (\tfrac{M}{\omega})^2 d\omega\right) \le \tfrac{1}{\pi}\|W\|\left(\max_{\omega\in[0,1]}\|G(i\omega)\|^2 + M^2\right).$$

Therefore V is uniformly bounded.
Let us now show that V equals U if (6.1) is exponentially stable. As $K \in L^1 \cap L^2$, the Fourier(-Plancherel) transform of K is given by the L^2 function $\omega \mapsto G(i\omega)$. The inverse transformation gives $K(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(i\omega)e^{i\omega t}d\omega$ for $t \geq 0$. Now consider the weighted inner product

$$\langle f, g\rangle_W = \int_{-\infty}^{\infty} g(\theta)^* W f(\theta) d\theta = \int_{-\infty}^{\infty} (W^{1/2} g(\theta))^* (W^{1/2} f(\theta)) d\theta \qquad \text{on } L^2(\mathbb{R}, \mathbb{C}^n).$$

Applying Parseval's formula ([135, Equation (2.1.8)], [34, Section 6.5.2]) to this inner product yields for all $t \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$

$$\langle K(\cdot + t)x, K(\cdot)y\rangle_W = \int_{-\infty}^{\infty} y^* K(\theta)^* W K(t+\theta) x \, d\theta = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} y^* G(i\omega)^* W G(i\omega) e^{it\omega} x \, d\omega.$$

But $\langle K(\cdot + t)x, K(\cdot)y\rangle_W = y^* U(t) x$ for all $x, y \in \mathbb{C}^n$. Therefore V equals the definition of U in (6.16). □

Note that by Proposition 6.29 the integral (6.41) exists if $i\mathbb{R} \subset \varrho(\Sigma)$, hence this formula may provide solutions of (6.37)–(6.39) also in case that the delay system is not exponentially stable. The two Propositions 6.28 and 6.29 show that we have to study the existence and uniqueness of solutions for (6.37),(6.38),(6.39) which we pose in form of the following problem.

Problem 6.30. *For a given symmetric positive definite matrix $W \in \mathcal{H}_+^n(\mathbb{R})$ find a continuous matrix function $U : \mathbb{R} \to \mathbb{R}^{n\times n}$ which solves the delay differential equation* (6.37) *on $\mathbb{R}_+$ (with initial function $U|_{[-H,0]}$) and satisfies the conditions* (6.38) *and* (6.39)*.*

Instead of specifying an initial function directly, the initial function is given by the symmetry condition, and hence by mirroring a part of the solution. In a sense, we deal here with a boundary problem for delay equations. Let us comment on the smoothness of solutions. Assume that we have a solution of (6.37)–(6.39) for the initial matrix segment $U|_{[-H,0]}$. If the initial function $U|_{[-H,0]}$ is continuous, then this solution is continuously differentiable. But by symmetry (6.38), the initial function is itself continuously differentiable. Repeating this argument, we see that U is infinitely differentiable, with a possible exception at $t = 0$ where the delay equation (6.37) only determines the one-sided derivative $\dot{U}(0+)$.
For the choice $W = W_0 + HW_H$ we obtain a functional U that can be used for the construction of Liapunov-Krasovskii functionals in Theorem 6.23. Here U does not depend directly on the terms W_0 and W_H, but only via the sum $W_0 + HW_H$.
We will now show that equation (6.37) and conditions (6.38),(6.39) uniquely determine the delay Liapunov matrix if (6.1) is exponentially stable.

Theorem 6.31. *Suppose that* (6.1) *is exponentially stable. Given a Hermitian W there exists a unique solution of* (6.37) *satisfying the conditions* (6.38) *and* (6.39) *which is given by the matrix $U(t)$ of* (6.16)*.*

Proof. By Proposition 6.28, $U(t)$ given by (6.16) is a solution of (6.37)–(6.39). Let us now assume that Problem 6.30 has two different solutions $U^1(t)$ and $U^2(t)$ for a given W. We define two functionals $v_i : M^2 \to \mathbb{R}_+$, $i = 1, 2$, which operate on $z = (x_0,\, \varphi) \in M^2$,

$$\begin{aligned} v_i(z) = x_0^* U^i(0) x_0 + \sum_{k=1}^{m} 2\mathrm{Re}\, x_0^* \int_{-h_k}^{0} U^i(-h_k - \theta) A_k \varphi(\theta) d\theta + \\ + \sum_{k=1}^{m}\sum_{j=1}^{m} \int_{-h_k}^{0} \varphi(\theta_2)^* A_k^\top \left[\int_{-h_j}^{0} U^i(\theta_2 - \theta_1 + h_k - h_j) A_j \varphi(\theta_1) d\theta_1 \right] d\theta_2 \end{aligned} \tag{6.43}$$

corresponding to U^1 and U^2, respectively. These functionals satisfy $v_i(z) = \langle z, P^i z\rangle_{M^2}$ where $P^i : M^2 \to M^2$ are given by (6.19), (6.20) with U replaced by U^i and where $W_H = 0$. Hence by Corollary 6.21 we have

$$\dot{v}_i(\hat{x}_t(z)) = -x(t,z)^* W x(t,z) \qquad \text{for} \quad t \geq 0,\ z \in D(A),\ i = 1, 2.$$

Thus the difference $v(\hat{x}_t) = v_2(\hat{x}_t) - v_1(\hat{x}_t)$ satisfies the equality $\dot{v}(\hat{x}_t) = 0$, $t \geq 0$. This shows that for all initial segments $z \in D(A)$ and all $t \geq 0$ we have $v(\hat{x}_t(z)) = v(z)$ as v is constant along solutions of (6.1). By exponential stability of (6.1), $\|\hat{x}_t(z)\|_{M^2} \to 0$ as $t \to \infty$, therefore it follows from Definition 6.22 that also $v(\hat{x}_t(z)) \to 0$ for $t \to \infty$ which implies that $v(z) = 0$ for every initial segment $z \in M^2$. Now, $D(A)$ is dense in M^2 and therefore $v(z) = 0$ for all $z \in M^2$. Using $U(t) = U^2(t) - U^1(t)$ in (6.43) for $t = 0$ yields

$$\begin{aligned} 0 = x_0^* U(0) x_0 + \sum_{k=1}^{m} 2\mathrm{Re}\, x_0^* \int_{-h_k}^{0} U(-h_k - \theta) A_k \varphi(\theta) d\theta + \\ + \sum_{k=1}^{m}\sum_{j=1}^{m} \int_{-h_k}^{0} \varphi(\theta_2)^* A_k^\top \left(\int_{-h_j}^{0} U(\theta_2 - \theta_1 + h_k - h_j) A_j \varphi(\theta_1) d\theta_1 \right) d\theta_2, \end{aligned} \tag{6.44}$$

because $U(t) = U^2(t) - U^1(t)$ satisfies the conditions of Problem 6.30 with $W = 0$. Now for $y \in \mathbb{C}^n$ consider the M^2-initial value $z = (y, 0)$. For this z all integrals in (6.44) vanish and hence (6.44) takes the form $y^\top U(0) y = 0$. Since y is an arbitrary vector and $U(0)$ is a symmetric matrix, $U(0) = 0$ must hold. Now, fix an index $i \in \{1, 2\ldots, m\}$ and choose $\tau \in [-h_i, -h_{i-1})$ and $\varepsilon > 0$ such that $\tau + \varepsilon < -h_{i-1}$. For any two given vectors $y, y' \in \mathbb{C}^n$ consider now the initial value

$$z = (y,\, \varphi) \in M^2, \qquad \varphi(t) = \begin{cases} y', & t \in [\tau, \tau + \varepsilon], \\ 0, & \text{for all other} \quad t \in [-H, 0). \end{cases}$$

For this $z \in M^2$, condition (6.44) now reads

$$\begin{aligned} 0 = \sum_{k=i}^{m} 2\mathrm{Re}\, y^* \left(\int_{\tau}^{\tau+\varepsilon} U(-h_k - \theta) A_k d\theta \right) y' + \\ + \sum_{k=i}^{m}\sum_{j=i}^{m} y'^* A_k^\top \left(\int_{\tau}^{\tau+\varepsilon} \int_{\tau}^{\tau+\varepsilon} U(\theta_1 - \theta_2 - h_k + h_j) d\theta_1 d\theta_2 \right) A_j y'. \end{aligned}$$

If $\varepsilon > 0$ is small then the first integral is proportional to ε while the double integral is proportional to ε^2 so that the last equation can be written as

$$0 = 2\mathrm{Re}\,\varepsilon y^* \left(\sum_{k=i}^{m} U(-h_k - \tau) A_k \right) y' + o(\varepsilon),$$

where $\frac{o(\varepsilon)}{\varepsilon} \to 0$ as $\varepsilon \to 0$. As y and y' are arbitrary vectors and as ε can be made arbitrarily small,

$$\sum_{k=i}^{m} U(t - h_k) A_k = 0 \quad \text{for} \quad t \in (h_{i-1}, h_i]. \tag{6.45}$$

Now (6.45) holds for all $i = 1, 2, \ldots, m$. For $i = 1$ we therefore obtain from (6.37) the differential equation $\dot{U}(t) = U(t)A_0$ for $t \in (0, h_1]$ as $\sum_{k=1}^{m} U(t - h_k)A_k = 0$. But we already know $U(0) = 0$, and hence $U(t) = 0$ for all $t \in [0, h_1]$. On the interval $(h_1, h_2]$ equations (6.37) and (6.45) for $i = 2$ now yield the delay equation $\dot{U}(t) = U(t)A_0 + U(t - h_1)A_1$. But on the interval $[0, h_1]$, $U(t)$ is constantly 0, therefore $U(t) = 0$ for $t \in (h_1, h_2]$. Continuing this process we conclude that $U(t) = 0$, $t \in [0, H]$, i.e., $U^1(t) = U^2(t)$ for all $t \in [-H, H]$. Hence every solution of Problem 6.30 is given by the integral (6.16) whenever (6.1) is exponentially stable. □

Let us now investigate under which conditions equation (6.37) has no solution satisfying the conditions (6.38) and (6.39). Of course, by the previous Theorem 6.31 such a situation may only occur if system (6.1) is not exponentially stable. We first discuss the relationship between solvability of (6.37)–(6.39) and the uniqueness of its solutions.

Proposition 6.32. *The solution set $\mathcal{U}_0$ of* (6.37)–(6.39) *associated with $W = 0$ forms a real linear subspace. If $\mathcal{U}_0$ is non-trivial, then the solution set $\mathcal{U}_W$ for a given $W \in \mathcal{H}^n(\mathbb{R})$ is either empty or given by $\mathcal{U}_W = \mathcal{U}_0 + U_W$ where U_W is a particular solution of* (6.37)–(6.39) *associated with W.*

Proof. Equations (6.37)–(6.39) represent a system of affine equations for continuous and apart from 0 differentiable matrix functions $U : \mathbb{R} \to \mathbb{R}^{n\times n}$. The associated homogeneous system of equations is given by (6.37)–(6.39) with $W = 0$. □

In Corollary 6.21 we constructed a solution of the operator Liapunov equation given the explicit integral formula (6.16) of the delay Liapunov matrix U. Let us now show that the same construction can be accomplished given a solution of equations (6.37)–(6.39) in Proposition 6.28.

Theorem 6.33. *Let $W_0 = W$ and $W_H = 0$. If $U : [-H, H] \to \mathbb{R}^{n\times n}$ is a solution of* (6.37)–(6.39) *then for $z = (x_0, \varphi)$, $\tilde{z} = (y_0, \psi)$ in $D(A)$ and for solutions $x_t = x(t + \cdot, z)$, $y_t = x(t + \cdot, \tilde{z})$ of* (6.1) *we have*

$$\tfrac{d}{dt}\langle \hat{x}_t, P\hat{y}_t \rangle_{M^2} = -\langle x(t), Wy(t) \rangle_2, \qquad t > 0, \tag{6.46}$$

where P is given by (6.19) *and* (6.20)*.*

Especially for $z = \tilde{z} = \binom{x}{\varphi} \in M^2$, P is a solution of the Liapunov equation with right hand side $\langle x, Wx\rangle_2$. However, as this it not coercive in M^2 , it is not clear if P is a coercive self-adjoint bounded linear operator.

Proof. We have to verify that the derivative of $\langle P\hat{y}_t, \hat{x}_t\rangle_{M^2}$ equals $-\langle Wy(t), x(t)\rangle_2 = \langle (\dot{U}(0) + \dot{U}(0)^\top)y(t), x(t)\rangle_2$. We write down the derivative in (6.46) more explicitly,

$$\begin{aligned}
\tfrac{d}{dt}\langle P\hat{y}_t, \hat{x}_t\rangle_{M^2} &= \Big\langle P\tbinom{y_t(0)}{y_t(\cdot)}, A\tbinom{x_t(0)}{x_t(\cdot)}\Big\rangle_{M^2} + \Big\langle PA\tbinom{y_t(0)}{y_t(\cdot)}, \tbinom{x_t(0)}{x_t(\cdot)}\Big\rangle_{M^2} \\
&= \dot{x}(t)^* \left(U(0)y(t) + (P_1 y_t)\right) + \int_{-H}^{0} \dot{x}_t(\theta)^* \left((P_1^* y(t))(\theta) + (P_2 y_t)(\theta)\right) d\theta \\
&\quad + x(t)^* \left(U(0)\dot{y}(t) + (P_1\dot{y}_t)\right) + \int_{-H}^{0} x_t(\theta)^* \left((P_1^* \dot{y}(t))(\theta) + (P_2 \dot{y}_t)(\theta)\right) d\theta \\
&= \left(U(0)\dot{x}(t) + (P_1\dot{x}_t)\right)^* y(t) + \int_{-H}^{0} \left((P_1^* \dot{x}(t))(\theta) + (P_2\dot{x}_t)(\theta)\right)^* y_t(\theta) d\theta \\
&\qquad + x(t)^* \left(U(0)\dot{y}(t) + (P_1\dot{y}_t)\right) + \int_{-H}^{0} x_t(\theta)^* \left((P_1^* \dot{y}(t))(\theta) + (P_2\dot{y}_t)(\theta)\right) d\theta, \qquad (6.47)
\end{aligned}$$

where we used the duality of P_1 and P_1^*, see Lemma 6.19. Let us start by computing the difference of $\dot{U}(0)x(t) - U(0)\dot{x}(t)$. By (6.37) U satisfies $\dot{U}(t) = U(t)A_0 + \sum_{k=1}^m U(t-h_k)A_k$ on $t \geq 0$, while the solution $x(t)$ of (6.3) satisfies $\dot{x}(t) = A_0 x(t) + \sum_{k=1}^m A_k x(t-h_k)$ on $t \geq 0$. By partial integration we obtain

$$\begin{aligned}
\dot{U}(0)x(t) - U(0)\dot{x}(t) &= \sum_{k=1}^{m} U(-h_k)A_k x(t) - U(0)A_k x(t-h_k) \\
&= \sum_{k=1}^{m} \big[U(-\theta - h_k)A_k x(t+\theta)\big]_{\theta=-h_k}^{0} \\
&= \sum_{k=1}^{m} \int_{-h_k}^{0} U(-\theta - h_k)A_k \dot{x}(t+\theta) + \tfrac{d}{d\theta}(U(-\theta-h_k))A_k x(t+\theta)\, d\theta \\
&= (P_1\dot{x}_t) + \sum_{k=1}^{m} \int_{-h_k}^{0} \tfrac{d}{d\theta}(U(-\theta - h_k))A_k x(t+\theta)\, d\theta.
\end{aligned} \qquad (6.48)$$

For the ease of notation let us introduce the operator

$$P_3 : \mathbb{R} \times L^2 \to \mathbb{C}^n, \qquad P_3(\tau, x_t) = \sum_{k=1}^{m} \int_{-h_k}^{0} \tfrac{d}{d\theta}(U(\tau - \theta - h_k))A_k x_t(\theta)\, d\theta,$$

so that (6.48) can now be written as $U(0)\dot{x}(t) + P_1(\dot{x}_t) = \dot{U}(0)x(t) - P_3(0, x_t)$.

We now study the term $Z := \int_{-H}^{0} x_t(\theta)^* \left((P_1^*\dot{y}(t))(\theta) + (P_2\dot{y}_t)(\theta)\right) d\theta$. We obtain the following expression,

$$Z = \sum_{k=1}^{m} \int_{-h_k}^{0} x_t(\theta_1)^* A_k^\top \bigg(U(\theta_1 + h_k)\dot{y}(t) + \sum_{j=1}^{m} \int_{-h_j}^{0} U(\theta_1 - \theta_2 + h_k - h_j)A_j \dot{y}_t(\theta_2)\, d\theta_2 \bigg) d\theta_1.$$

If we now replace $\dot{y}(t)$ by $A_0y(t) + \sum_{j=1}^{m} A_jy(t-h_j)$ in Z we get

$$\begin{aligned} Z = \sum_{k=1}^{m} \int_{-h_k}^{0} x_t(\theta_1)^* A_k^\top \Bigg(U(\theta_1 + h_k)A_0y(t) + \sum_{j=1}^{m} U(\theta_1 + h_k)A_jy(t-h_j) \\ + \int_{-h_j}^{0} U(\theta_1 - \theta_2 + h_k - h_j)A_j\dot{y}_t(\theta_2)\,d\theta_2 \Bigg) d\theta_1. \end{aligned}$$

The inner sum can be modified by partial integration analogously to (6.48), thus

$$\begin{aligned} &\sum_{j=1}^{m} U(\theta_1 + h_k)A_jy(t-h_j) + \int_{-h_j}^{0} U(\theta_1 - \theta_2 + h_k - h_j)A_j\dot{y}_t(\theta_2)d\theta_2 \\ &= \sum_{j=1}^{m} U(\theta_1 + h_k - h_j)A_jy(t) - \int_{-h_j}^{0} \tfrac{d}{d\theta_2}(U(\theta_1 - \theta_2 + h_k - h_j))A_jy_t(\theta_2)d\theta_2 \\ &= \left(\sum_{j=1}^{m} U(\theta_1 + h_k - h_j)A_jy(t) \right) - P_3(\theta_1 + h_k). \end{aligned}$$

We therefore obtain using the symmetry of U

$$\begin{aligned} Z &= \sum_{k=1}^{m} \int_{-h_k}^{0} x_t(\theta_1)^* A_k^\top \Bigg(U(\theta + h_k)A_0y(t) + \sum_{j=1}^{m} U(\theta + h_k - h_j)A_jy(t) - P_3(\theta + h_k) \Bigg) d\theta \\ &= \sum_{k=1}^{m} \int_{-h_k}^{0} x_t(\theta)^* A_k^\top \Big(\dot{U}(\theta + h_k)y(t) - P_3(\theta + h_k, y_t) \Big) d\theta \\ &= \left(\sum_{k=1}^{m} \int_{-h_k}^{0} \tfrac{d}{d\theta}(U(-\theta - h_k)A_kx_t(\theta)d\theta \right)^* y(t) - \sum_{k=1}^{m} \int_{-h_k}^{0} x_t(\theta)^* A_k^\top P_3(\theta + h_k, y_t) \\ &= P_3(0, x_t)^* y(t) - \sum_{k=1}^{m} \int_{-h_k}^{0} x_t(\theta)^* A_k^\top P_3(\theta + h_k, y_t). \end{aligned} \tag{6.49}$$

A dual result holds when exchanging x_t and y_t so that we can treat both integral terms in (6.47). We now have all the results ready to write $\frac{d}{dt}\langle \hat{x}_t, P\hat{y}_t \rangle_{M^2}$ without explicit dependency on the derivative of the trajectories. Using (6.48) and (6.49), (6.47) now reads

$$\begin{aligned} \tfrac{d}{dt}\langle \hat{x}_t, P\hat{y}_t \rangle_{M^2} &= \Big(\dot{U}(0)x(t) - P_3(0, x_t) \Big)^* y(t) \\ &\quad + x(t)^* P_3(0, y_t) + x(t)^* \Big(\dot{U}(0)y(t) - P_3(0, y_t) \Big) + P_3(0, y_t)^* y(t) \\ &\quad - \sum_{k=1}^{m} \int_{-h_k}^{0} \big(P_3(\theta + h_k, x_t)^* A_k y(\theta) + x_t(\theta)^* A_k^\top P_3(\theta + h_k, y_t) \big)\, d\theta \\ &= x(t)^* \Big(\dot{U}(0)^\top + \dot{U}(0) \Big) y(t) = -x(t)^* W y(t). \end{aligned}$$

To verify this equality we have to check that the sum involving the P_3 operators vanishes. But this follows after a small calculation from

$$\left(\tfrac{d}{d\theta_2}U(\theta_1 + h_k - \theta_2 - h_j)\right)^\top = -\tfrac{d}{d\theta_1}U(\theta_2 + h_j - \theta_1 - h_k).$$

Hence (6.46) holds for all M^2-initial conditions. □

In the following we describe a situation where there exists a weight $W \in \mathcal{H}^n(\mathbb{R})$ such that there exists no associated solution. We need the following technical lemma.

Lemma 6.34. *For two non-trivial vectors $x, y \in \mathbb{C}^n$ there exists a real symmetric matrix $W \in \mathbb{R}^{n\times n}$ such that $x^*Wy \neq 0$.*

Proof. Assume that x and y are linearly independent vectors. Then the Cauchy-Schwarz inequality yields $|x^*y|^2 < \|x\|^2 \|y\|^2$, hence

$$x^*(xy^* + yx^*)y = \|x\|^2 \|y\|^2 + (x^*y)^2 \neq 0.$$

For linearly dependent vectors, choose $W = I_n$. □

Proposition 6.35. *If there exists $\lambda_0 \in \mathbb{C}$ such that $\{\lambda_0, -\lambda_0\} \subset \sigma(A)$, i.e.,*

$$\det\left(\pm\lambda_0 I_n - A_0 - \sum_{k=1}^m e^{\mp\lambda_0 h_k} A_k\right) = 0, \tag{6.50}$$

then there exists a symmetric matrix W for which (6.37) *has no solution satisfying the conditions* (6.38)–(6.39)*. Moreover, in this case there exists a non-trivial solution of* (6.37)–(6.39) *with $W = 0$.*

Proof. Assume by contradiction that for every symmetric matrix W, equation (6.37) has a solution satisfying conditions (6.38)–(6.39). Note that as the matrices A_k are all real, $\lambda \in \sigma(A)$ implies that $\bar{\lambda} \in \sigma(A)$. Thus we can pick two eigenmotions of system (6.1) associated with the eigenvalues $\lambda_1 = \lambda_0$ and $\lambda_2 = -\bar{\lambda}_0$ (see Definition 6.4) which are of the form

$$x^{(1)}(t) = e^{\lambda_1 t}x, \qquad x^{(2)}(t) = e^{\lambda_2 t}y, \qquad x, y \in \mathbb{C}^n, x, y \neq 0, \quad t \geq -H, \tag{6.51}$$

and which are solutions of (6.1). By Lemma 6.34 there exists a symmetric matrix W such that $x^*Wy \neq 0$. Now by assumption, (6.37) has a solution $U(t)$ which satisfies the conditions (6.38)–(6.39). Let us define the bilinear functional for $z = (x_0, \varphi)$, $\tilde{z} = (\tilde{x}_0, \tilde{\varphi})$

$$\begin{aligned} p(z, \tilde{z}) = x_0^* U(0)\tilde{x}_0 + \sum_{j=1}^m x_0^* \int_{-h_j}^0 U(-h_j - \theta)A_j \tilde{\varphi}(\theta)d\theta \\ + \sum_{k=1}^m \int_{-h_k}^0 \varphi(\theta)^* A_k^\top U(h_k + \theta)d\theta \tilde{x}_0 \\ + \sum_{k=1}^m \sum_{j=1}^m \int_{-h_k}^0 \varphi(\theta_2)^* A_k^\top \int_{-h_j}^0 U(\theta_2 - \theta_1 + h_k - h_j)A_j \tilde{\varphi}(\theta_1)d\theta_1 d\theta_2, \end{aligned} \tag{6.52}$$

i.e., $p(z,\tilde{z}) = \langle P\tilde{z}, z\rangle_{M^2}$ where $P : M^2 \to M^2$ is given by (6.19) and (6.20) with $W_H = 0$. To see this, compare (6.52) with (6.30). Note that we do not assume that U is of the form (6.16). However, U solves Problem 6.30. The solutions $x^{(1)}(t)$ and $x^{(2)}(t)$ defined by (6.51) are in the domain of A, hence we can apply Theorem 6.33. We obtain

$$\tfrac{d}{dt}p(\hat{x}_t^{(1)}, \hat{x}_t^{(2)}) = -x^{(1)}(t)^* W x^{(2)}(t) = -e^{(\bar{\lambda}_1+\lambda_2)t} x^* W y = -x^* W y \neq 0. \tag{6.53}$$

On the other hand, direct substitution of these solutions into the bilinear functional yields

$$\begin{aligned} p(\hat{x}_t^{(1)}, \hat{x}_t^{(2)}) = e^{(\bar{\lambda}_1+\lambda_2)t} x^* \Big[U(0) + \sum_{j=1}^{m} \int_{-h_j}^{0} U(-h_j-\theta) A_j e^{\lambda_2\theta} + A_j^\top U(h_j+\theta) e^{\bar{\lambda}_1\theta} d\theta + \\ + \sum_{k=0}^{m}\sum_{j=0}^{m} \int_{-h_k}^{0}\int_{-h_j}^{0} e^{\lambda_2\theta_1+\bar{\lambda}_1\theta_2} A_k^\top U(\theta_2-\theta_1+h_k-h_j) A_j d\theta_1 d\theta_2 \Big] y. \end{aligned}$$

Observe that the matrix in square brackets does not depend on t. The condition $\bar{\lambda}_1 + \lambda_2 = \lambda_0 - \lambda_0 = 0$ therefore implies that

$$\tfrac{d}{dt}p(\hat{x}_t^{(1)}, \hat{x}_t^{(2)}) = 0. \tag{6.54}$$

But this is in contradiction to (6.53). Hence there exists no solution of (6.37) satisfying (6.38)–(6.39) for the special choice of W.

If $W = 0$ then for any non-trivial solution U of (6.37)–(6.39), $\dot{U}(0)$ is skew-symmetric while $U(0)$ is symmetric, see the discussion following Proposition 6.28. We now construct such a solution. By (6.50) there exist vectors $v_i \in \mathbb{C}^n$, $v_i \neq 0$, such that $v_i^\top(A_0 + \sum_{k=1}^m e^{-\lambda_i h_k} A_k) = \lambda_i v_i^\top$ for $i = 1,2$ where $\lambda_1 = \lambda_0, \lambda_2 = -\lambda_0$. Define

$$U(t) = e^{\lambda_0 t} v_2 v_1^\top + e^{-\lambda_0 t} v_1 v_2^\top + e^{\bar{\lambda}_0 t} \bar{v}_2 v_1^* + e^{-\bar{\lambda}_0 t} \bar{v}_1 v_2^* = \mathrm{Re}\,\left(e^{\lambda_0 t} v_2 v_1^\top + e^{-\lambda_0 t} v_1 v_2^\top\right), \qquad t \in \mathbb{R},$$

which is real and satisfies the symmetry condition (6.38). Here the real (or Hermitian) part of a matrix $A \in \mathbb{C}^{n\times n}$ is given by $\mathrm{Re}\, A = \frac{1}{2}(A + A^*)$. Now U also satisfies the differential equation (6.37) for every $t \in \mathbb{R}$, as

$$\begin{aligned} U(t)A_0 + \sum_{k=1}^{m} U(t-h_k)A_k &= 2\mathrm{Re}\,\left(e^{\lambda_0 t} v_2 v_1^\top (A_0 + \sum_{k=1}^{m} e^{-\lambda_0 h_k} A_k) + e^{-\lambda_0 t} v_1 v_2^\top (A_0 + \sum_{k=1}^{m} e^{\lambda_0 h_k} A_k) \right) \\ &= 2\mathrm{Re}\,\left(\lambda_0 e^{\lambda_0 t} v_2 v_1^\top - \lambda_0 e^{-\lambda_0 t} v_1 v_2^\top\right) = \dot{U}(t). \end{aligned}$$

For $t = 0$ the derivative $\dot{U}(0) = 2\mathrm{Re}\,\left(\lambda_0(v_2 v_1^\top - v_1 v_2^\top)\right)$ is skew-symmetric, hence (6.39) is satisfied with $W = 0$. □

Remark 6.36. The proof Proposition 6.35 shows that under the condition that there are two eigenvalues of (6.1) with sum 0 or $0 \in \sigma(A)$, Problem 6.30 has a non-trivial solution of $W = 0$. It is generally not known if the condition of Proposition 6.35 is the only critical condition.

We will investigate this question for systems with one delay ($m = 1$) in the next section.

6.4 The One-Delay Case

Let us now assume that system (6.1) has only one delay term $h > 0$ ($m = 1$),

$$\dot{x}(t) = A_0 x(t) + A_1 x(t-h), \qquad t \geq 0. \tag{6.1'}$$

We study existence and uniqueness issues for this case.
The symmetry condition (6.38) allows us to reformulate the differential equation (6.37) for U as a delay-free ordinary differential matrix equation. This has already been studied in Infante and Castellan [73] and Datko [32]. A recent analysis of this approach may be found in Luisell [98, 99] where it is used to locate those eigenvalues of (6.1') which lie on the imaginary axis. Consider the following problem formulation.

Problem 6.37. *For a given symmetric matrix $W \in \mathbb{R}^{n\times n}$ find a solution $U : [-h, h] \to \mathbb{R}^{n\times n}$ satisfying*

$$\begin{aligned}
&\dot{U}(t) = U(t)A_0 + U(t-h)A_1, && && t \in [0, h], \qquad (6.55)\\
&U(t) = U(-t)^\top, && t \in [-h, h] && \textit{(symmetry condition)},\\
&U(0)A_0 + U(h)^\top A_1 + A_0^\top U(0) + A_1^\top U(h) = -W && && \textit{(algebraic condition)}.
\end{aligned}$$

As this problem is just the restriction of Problem 6.30 to $t \in [-h, h]$ and $m = 1$, any solution of Problem 6.37 is called a delay Liapunov matrix for (6.1'). Note that we do not assume exponential stability, so the integral representation (6.16) is not applicable. Therefore not only uniqueness, but also existence of delay Liapunov matrices must be investigated. We do so by introducing the following boundary value problem for a delay-free system for which the solution set is basically equivalent to the one of Problem 6.37.

Problem 6.38. *For a given symmetric matrix $W \in \mathbb{R}^{n\times n}$ find solutions $U, V : [0, h] \to \mathbb{R}^{n\times n}$ of the ordinary differential system*

$$\dot{U}(t) = U(t)A_0 + V(t)A_1, \qquad \dot{V}(t) = -A_1^\top U(t) - A_0^\top V(t), \tag{6.56}$$

which satisfy the two conditions

$$\dot{U}(0) - \dot{V}(h) = -W, \qquad U(0) - V(h) = 0. \tag{6.57}$$

Here $\dot{U}(0)$ and $\dot{V}(h)$ serve as a shorthand notation for the one-sided derivatives,

$$\dot{U}(0) := \lim_{t\searrow 0} \dot{U}(t) = U(0)A_0 + V(0)A_1 \quad \text{and} \quad \dot{V}(h) := \lim_{t\nearrow h} \dot{V}(t) = -A_1^\top U(h) - A_0^\top V(h).$$

The differential equation for V is called the *counterflow* equation, see Marshall et al. [105]. Let us reformulate equations (6.56) and (6.57) by introducing linear operators working on pairs of matrices,

$$\mathcal{A}:\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n},\quad \begin{pmatrix}U\\V\end{pmatrix}\mapsto\begin{pmatrix}UA_0+VA_1\\-A_1^\top U-A_0^\top V\end{pmatrix},\tag{6.58}$$

$$\pi_i:\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n},\quad \begin{pmatrix}X_1\\X_2\end{pmatrix}\mapsto X_i,\ i=1,2,$$

$$\mathcal{B}_1,\mathcal{B}_2:\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n},\qquad \mathcal{B}_1=(\pi_1-\pi_2e^{\mathcal{A}h})\mathcal{A},\qquad \mathcal{B}_2=\pi_1-\pi_2e^{\mathcal{A}h},$$

$$\mathcal{B}=\begin{pmatrix}\mathcal{B}_1\\\mathcal{B}_2\end{pmatrix}:\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}.\tag{6.59}$$

Then Problem 6.38 can be written compactly as $\dot{x}(t)=\mathcal{A}x(t)$ on $t\in[0,h]$ with the boundary condition

$$\mathcal{B}x(0)=\left(\begin{smallmatrix}-W\\0\end{smallmatrix}\right)\qquad\text{for}\quad x(0)\in\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}.\tag{6.60}$$

From (6.59) we see that the kernel of $\mathcal{B}$ satisfies $\ker\mathcal{B}=\ker\mathcal{B}_1\cap\ker\mathcal{B}_2$. If this kernel is trivial then $\mathcal{B}$ is a vector-space automorphism of $\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$.

Corollary 6.39. *A solution of Problem 6.38 is obtained by prescribing an initial value $x(0)=\left(\begin{smallmatrix}U_0\\V_0\end{smallmatrix}\right)=\mathcal{B}^{-1}\left(\begin{smallmatrix}-W\\0\end{smallmatrix}\right)$ for $\dot{x}(t)=\mathcal{A}x(t)$. Then $x(t)=\left(\begin{smallmatrix}U(t)\\V(t)\end{smallmatrix}\right)$ where $U(t)$ and $V(t)$ are a solution of Problem 6.38. Moreover, a solution of Problem 6.38 is uniquely determined if and only if the* boundary operator $\mathcal{B}$ *is invertible.*

In particular, if $x(t)$ is such a solution then the symmetry condition is regained by

$$\mathcal{B}_2x(0)=\pi_1x(0)-\pi_2x(h)=\pi_1\left(\begin{smallmatrix}U(0)\\V(0)\end{smallmatrix}\right)-\pi_2\left(\begin{smallmatrix}U(h)\\V(h)\end{smallmatrix}\right)=U(0)-V(h)=0,$$

and the algebraic condition reappears from

$$\mathcal{B}_1x(0)=\mathcal{B}_2\mathcal{A}x(0)=\mathcal{B}_2\dot{x}(0)=\mathcal{B}_2\left(\begin{smallmatrix}\dot{U}(0)\\\dot{V}(0)\end{smallmatrix}\right)=\dot{U}(0)-\dot{V}(h)=-W.$$

Here we used that if $x(t)$ is a solution of the linear system $\dot{x}=\mathcal{A}x$ then $\dot{x}(t)$ is also a solution.

The solution sets of Problems 6.37 and 6.38 are equivalent in the following sense.

Proposition 6.40. *If $U:[-h,h]\to\mathbb{R}^{n\times n}$ is a solution of Problem 6.37 then the pair $(U,V):[0,h]\to\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$ with $V(t)=U(h-t)^\top$ solves Problem 6.38. If the pair $(U,V):[0,h]\to\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$ solves Problem 6.38 then*

$$\tilde{U}(t)=\tfrac{1}{2}\begin{cases}U(t)+V(h-t)^\top, & t\in[0,h],\\ U(-t)^\top+V(h+t), & t\in[-h,0),\end{cases}\qquad\textit{solves Problem 6.37.}$$

Proof. Suppose that $U(t)$ solves Problem 6.37. Set $V(t) = U(h-t)^\top$. By symmetry,

$$\begin{aligned} \dot U(t) &= U(t)A_0 + U(h-t)^\top A_1 = U(t)A_0 + V(t)A_1, && t \in [0,h], \\ \dot V(t) &= -A_0^\top U(h-t)^\top - A_1^\top U(h-(h-t))^\top = -A_1^\top U(t) - A_0^\top V(t), && t \in [0,h]. \end{aligned} \tag{6.61}$$

Moreover, the symmetry condition $U(0) = U(0)^\top$ gives $U(0) = V(h)$. Applying this equality and $V(0) = U(h)^\top$ to the algebraic condition in (6.55) yields

$$\begin{aligned} -W &= U(0)A_0 + U(h)^\top A_1 + A_0^\top U(0) + A_1^\top U(h) = \\ &\qquad = U(0)A_0 + V(0)A_1 + A_0^\top V(h) + A_1^\top U(h) = \dot U(0) - \dot V(h). \end{aligned}$$

Therefore the pair $(U(t), V(t))$ solves Problem 6.38.
On the other hand, given a solution pair (U, V) of Problem 6.38, the pair $(\hat U(t), \hat V(t)) = (V(h-t)^\top, U(h-t)^\top)$ also solves Problem 6.38 since

$$\begin{aligned} \dot{\hat U}(t) &= -\left(-A_1^\top U(h-t) - A_0^\top V(h-t)\right)^\top = \hat U(t)A_0 + \hat V(t)A_1, \\ \dot{\hat V}(t) &= -\left(U(h-t)A_0 + V(h-t)A_1\right)^\top = -A_1^\top \hat U(t) - A_0^\top \hat V(t). \end{aligned}$$

Furthermore we have $\hat U(0) - \hat V(h) = V(h)^\top - U(0)^\top = 0$ and by symmetry of W

$$\begin{aligned} \dot{\hat U}(0) - \dot{\hat V}(h) &= \hat U(0)A_0 + \hat V(0)A_1 + A_1^\top \hat U(h) + A_0^\top \hat V(h) = \\ &= V(h)^\top A_0 + U(h)^\top A_1 + A_1^\top V(0)^\top + A_0^\top U(0)^\top = \\ &= \left(A_0^\top U(0) + A_1^\top U(h) + V(0)A_1 + V(h)A_0\right)^\top = \left(\dot U(0) - \dot V(h)\right)^\top = -W. \end{aligned}$$

We will now show that $\tilde U$ defined in the proposition solves Problem 6.37. Note that $\tilde U = \frac{1}{2}((U(t) + \hat U(t))$ on $[0,h]$. Hence for $t \in [0,h]$

$$\dot{\tilde U}(t) = \tfrac{1}{2}(U(t) + V(h-t)^\top)A_0 + \tfrac{1}{2}(V(t) + U(h-t)^\top)A_1 = \tilde U(t)A_0 + \tilde U(h-t)^\top A_1. \tag{6.62}$$

To verify the symmetry condition for $\tilde U$ it only remains to check $\tilde U(0) = \tilde U(0)^\top$ since by definition $\tilde U(t) = \tilde U(-t)^\top$ on $[-h, 0)$. But the condition $U(0) = V(h)$ of (6.57) implies that

$$\tilde U(0) = \tfrac{1}{2}\left(U(0) + V(h)^\top\right) = \tfrac{1}{2}\left(V(h) + U(0)^\top\right) = \tilde U(0)^\top. \tag{6.63}$$

Let us now verify the algebraic condition. Since W is symmetric, we have by (6.57) that $-W = \frac{1}{2}\left(\left(\dot U(0) - \dot V(h)\right) + \left(\dot U(0) - \dot V(h)\right)^\top\right)$. From this equation we obtain by using the symmetry of $\tilde U(0)$, (6.63), and the differential equations for U and V in (6.61) that

$$\begin{aligned} -W &= \tfrac{1}{2}\left((U(0) + V(h)^\top)A_0 + (V(0) + U(h)^\top)A_1\right) \\ &\qquad + \tfrac{1}{2}\left(A_1^\top(U(h) + V(0)^\top) + A_0^\top(V(h) + U(0)^\top)\right) \\ &\qquad\qquad = \tilde U(0)A_0 + \tilde U(h)^\top A_1 + A_1^\top \tilde U(h) + A_0^\top \tilde U(0)^\top, \end{aligned}$$

which is the algebraic condition for $\tilde U$ of Problem 6.37. Hence $\tilde U$ is a solution of Problem 6.37. □

From the proof of Proposition 6.40 we get the following corollary.

Corollary 6.41. *Given a solution pair $(U(t), V(t))$ of Problem 6.38 with $t \in [0,h]$.*

1. *The pair $(\tilde{U}(t), \tilde{V}(t)) = {}^1\!/\!_2(U(t)+V(h-t)^\top, V(t)+U(h-t)^\top)$ also solves Problem 6.38 and satisfies $\tilde{U}(t) = \tilde{V}(h-t)^\top$ for $t \in [0,h]$, and $\tilde{U}(0) = \tilde{U}(0)^\top$.*

2. *If the solution pair of problem 6.38 is uniquely determined then $U(t) = V(h-t)^\top$ for $t \in [0,h]$.*

The last item raises the uniqueness problem, for which we present the following uniqueness theorem.

Theorem 6.42. *The following statements are equivalent.*

(i) There exists a non-trivial solution pair (U, V) of Problem 6.38 associated with $W = 0$.

(ii) The boundary condition of Problem 6.38 is singular, i.e., $\ker \mathcal{B} \neq \{0\}$.

(iii) There exists $\lambda \in \mathbb{C}$ such that

$$\det(\lambda I - A_0 - A_1 e^{-\lambda h}) = 0 \quad \textit{and} \quad \det(-\lambda I - A_0 - A_1 e^{\lambda h}) = 0. \tag{6.64}$$

In this case, $\lambda, -\lambda \in \sigma(\mathcal{A})$.

(iv) There exists $\lambda \in \sigma(\mathcal{A})$ for which an associated eigenvector of $\mathcal{A}$ takes the form $\binom{U_0}{\zeta U_0} \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times n}$, $U_0 \neq 0$, $\zeta \in \mathbb{C}$, with $\zeta = e^{-\lambda h}$.

For the proof we recall the following technical lemma, see e.g., Arnold [5].

Lemma 6.43 (Unique Representation of Quasi-Polynomials)**.** *Given a quasi-polynomial $\varphi(t) = \sum_{i=1}^{\ell} e^{\lambda_i t} p_i(t)$ where $\lambda_i \in \mathbb{C}$, $\lambda_i \neq \lambda_j$ for $i \neq j$, and $p_i \in \mathbb{C}[t]$ are polynomials. Then $\varphi \equiv 0$ implies $p_i \equiv 0$ for all $i = 1, \dots, \ell$.*

Proof (of Theorem 6.42). *(iii) $\Longrightarrow$ (iv).* Let $\lambda \in \mathbb{C}$ such that $\det(\pm\lambda I - A_0 - A_1 e^{\mp\lambda h}) = 0$. Then there exist non-trivial vectors $v, w \in \mathbb{C}^n$ such that

$$\begin{gathered} v^\top\big(\lambda I - A_0 - A_1 e^{-\lambda h}\big) = 0 \quad \text{and} \quad w^\top\big(-\lambda I - A_0 - A_1 e^{\lambda h}\big) = 0, \quad \text{i.e.,} \\ \lambda v^\top = v^\top\big(A_0 + A_1 e^{-\lambda h}\big) \quad \text{and} \quad \lambda w^\top = w^\top\big(-A_0 - A_1 e^{\lambda h}\big). \end{gathered} \tag{6.65}$$

Setting $U_0 = wv^\top$ and $V_0 = e^{-\lambda h} wv^\top$ the pair $\binom{U_0}{V_0} \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times n}$ is an eigenvector of $\mathcal{A}$ corresponding to λ, as

$$\mathcal{A}\begin{pmatrix} U_0 \\ V_0 \end{pmatrix} = \begin{pmatrix} wv^\top A_0 + e^{-\lambda h} wv^\top A_1 \\ -A_1^\top wv^\top - A_0^\top e^{-\lambda h} wv^\top \end{pmatrix} = \begin{pmatrix} wv^\top (A_0 + e^{-\lambda h} A_1) \\ (-A_0^\top e^{-\lambda h} - A_1^\top) wv^\top \end{pmatrix} = \lambda \begin{pmatrix} wv^\top \\ e^{-\lambda h} wv^\top \end{pmatrix} = \lambda \begin{pmatrix} U_0 \\ V_0 \end{pmatrix},$$

where we used (6.65) to extract $\lambda \in \mathbb{C}$. Since $V_0 = e^{-\lambda h} U_0$, we have found an eigenvector of the required structure. If $\lambda = 0$ then $U_0 = V_0 = vv^\top$.

(iv) $\Longrightarrow$ (ii). An eigenvector $\binom{U_0}{V_0}$ of $\mathcal{A}$ corresponding to $\lambda \in \sigma(\mathcal{A})$ which satisfies $V_0 = e^{-\lambda h}U_0$ also satisfies the boundary condition $\mathcal{B}\binom{U_0}{V_0} = 0$ as

$$\mathcal{B}_2\binom{U_0}{V_0} = U_0 - \pi_2 e^{\mathcal{A}h}\binom{U_0}{V_0} = U_0 - \pi_2 e^{\lambda h}\binom{U_0}{V_0} = U_0 - e^{\lambda h}V_0 = U_0 - e^{\lambda h}(e^{-\lambda h}U_0) = 0,$$

whence $\mathcal{B}_1\binom{U_0}{V_0} = \mathcal{B}_2\mathcal{A}\binom{U_0}{V_0} = \lambda\mathcal{B}_2\binom{U_0}{V_0} = 0$. Therefore $\ker\mathcal{B}_1 \cap \ker\mathcal{B}_2 = \ker\mathcal{B} \neq \{0\}$, i.e., the boundary condition is not regular.
(ii) $\Longrightarrow$ (i). Suppose that the boundary condition of Problem 6.38 is singular. Then there exists a non-trivial pair of vectors (U_0, V_0) which satisfies $\mathcal{B}_1\binom{U_0}{V_0} = 0 = \mathcal{B}_2\binom{U_0}{V_0}$. But choosing (U_0, V_0) as the initial value for the differential equation $\frac{d}{dt}\binom{U(t)}{V(t)} = \mathcal{A}\binom{U(t)}{V(t)}$ gives a non-trivial solution $(U, V) : \mathbb{R}_+ \to \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ of Problem 6.38 corresponding to $W = 0$ as $U(0) - V(h) = 0, \dot{U}(0) - \dot{V}(h) = 0$, see Corollary 6.39.
(i) $\Longrightarrow$ (iii). By Corollary 6.41 a non-trivial solution can be chosen in such way that $U(t) = V(h-t)^\top$, $U(0) = U(0)^\top$ and $\dot{U}(0) = \dot{V}(h)$ (i.e., $W = 0$). Clearly, the solutions of the differential equations (6.56) not only exist on $[0, H]$ but on the whole real line. Then $\dot{U}(0)$ and $\dot{V}(h)$ are two-sided derivatives. We now show that the symmetry condition $U(-t) = U(t)^\top$ automatically holds for all $t \in \mathbb{R}$. For this we prove $U(t) = V(t+h)$. To see this consider the second order derivatives

$$\begin{aligned}\ddot{U}(t) &= \dot{U}(t)A_0 + \dot{V}(t)A_1 = \dot{U}(t)A_0 - \left(A_1^\top U(t) + A_0^\top V(t)\right)A_1\\ &= \dot{U}(t)A_0 - A_0^\top\dot{U}(t) + A_0^\top U(t)A_0 - A_1^\top U(t)A_1,\\ \ddot{V}(t) &= -A_1^\top\dot{U}(t) - A_0^\top\dot{V}(t) = -A_1^\top\left(U(t)A_0 + V(t)A_1\right) - A_0^\top\dot{V}(t)\\ &= \dot{V}(t)A_0 - A_0^\top\dot{V}(t) + A_0^\top V(t)A_0 - A_1^\top V(t)A_1.\end{aligned}$$

Hence U and V satisfy the same second order differential equation

$$\ddot{X}(t) = \dot{X}(t)A_0 - A_0^\top\dot{X}(t) + A_0^\top X(t)A_0 - A_1^\top X(t)A_1. \tag{6.66}$$

By the time-invariance of (6.66) it follows that $t \mapsto V(t+h)$ is also a solution of (6.66). Since $U(0) = V(h)$ and $\dot{U}(0) = \dot{V}(h)$, this solution satisfies the same initial conditions as U and therefore $U(t) = V(t+h)$ for all $t \in \mathbb{R}$. Corollary 6.41 then yields the symmetry result $U(t) = V(t+h) = U(-t)^\top$.
Furthermore, the solution pair (U, V) is given by a sum of eigenmotions of the finite-dimensional system (6.56). Hence, when projecting on the first component, there exist $\lambda_i \in \mathbb{C}$ and matrices $Z_{ik} \in \mathbb{C}^{n\times n}$, $i = 1, \ldots, \ell$, $k = 0, \ldots, N_i$, such that $\left\{e^{\lambda_i t}t^k Z_{ik}\right\}$ is a basis of the solution space for the U-component of (6.56) where $\lambda_i \in \sigma(\mathcal{A})$ are the associated eigenvalues and $Z_{ik} \in \mathbb{C}^{n\times n}$ are the U-components of generalized eigenvectors of (6.56). Therefore

$$U(t) = \sum_{i\in I} e^{\lambda_i t}\sum_{k\in K_i} t^k Z_{ik}, \qquad I \subset \{1, \ldots, \ell\}, K_i \subset \{0, \ldots, N_i\}, \quad t \in \mathbb{R},$$

where the coefficients are incorporated in $Z_{ik} \neq 0$. Since $U(t) = V(h-t)^\top = V(h+t)$, the matrix function $U(t)$ satisfies $\dot{U}(t) = U(t)A_0 + U(t-h)A_1$ on $\mathbb{R}$ because the differential equation is satisfied on $[0, h]$ and U and V are analytic functions. As the components of $\dot{U}(t)$ are formed by quasi-polynomials we obtain from Lemma 6.43 that

$$\lambda_i\left(\sum_{k\in K_i} t^k Z_{ik}\right)+\left(\sum_{k\in K_i\setminus\{0\}} k t^{k-1} Z_{ik}\right)=\left(\sum_{k\in K_i} t^k Z_{ik}\right)A_0+e^{-\lambda_i h}\left(\sum_{k\in K_i}(t-h)^k Z_{ik}\right)A_1, \quad i\in I.$$

Now consider for a fixed index i the coefficient matrix of $t^{\hat{k}_i}$ belonging to the highest degree $\hat{k}_i = \max K_i$. Then $Z_{i\hat{k}_i}(\lambda_i I - A_0 - e^{-\lambda_i h}A_1) = 0$. As $Z_{i\hat{k}_i} \neq 0$ we conclude that $\det(\lambda_i I - A_0 - e^{-\lambda_i t}A_1) = 0$. Projecting the eigenmotions of the solution pair (U, V) of (6.56) onto the V component and then repeating the above argument yields $\det(-\lambda_i I - A_0 - e^{\lambda_i t}A_1) = 0$. Thus we have found an eigenvalue λ of $\mathcal{A}$ which satisfies $\det(\pm\lambda I - A_0 - A_1 e^{\mp\lambda h}) = 0$. □

If the conditions of Theorem 6.42 do not hold we obtain the following result concerning the solution set of Problem 6.37.

Corollary 6.44. *For all symmetric $W \in \mathcal{H}^n(\mathbb{R})$ there exists a uniquely determined solution of Problem 6.37 if and only if there exists no $\lambda \in \mathbb{C}$ satisfying* (6.64)*, i.e., all eigenvalues $\lambda \in \sigma(A)$ of the generator A of the solution semigroup associated with* (6.1') *satisfy $-\lambda \notin \sigma(A)$.*

Proof. By Corollaries 6.39 and 6.41 (2) a unique solution of Problem 6.37 exists if and only if the boundary operator $\mathcal{B}$ is invertible. The proof of the equivalence of Theorem 6.42 *(iii)* and *(iv)* shows that there exists a one-to-one correspondence between eigenvalues $\lambda \in \sigma(\mathcal{A})$ of the finite-dimensional system (6.56) such that there exists an associated eigenvector with the special structure $(U_0, e^{-\lambda h}U_0)$ and eigenvalues $\lambda, -\lambda \in \sigma(A)$ of the semigroup generator A, see (6.64) and (6.11). Hence the uniqueness issue (and therefore the invertibility of $\mathcal{B}$) can be answered by considering the zeros of the characteristic equation associated with (6.1'). □

By using *Kronecker products* Problem 6.38 can be vectorized and the resulting equations can then be utilized in the numerical computation of solutions. The Kronecker product satisfies $\operatorname{vec} AXB = (B^\top \otimes A)\operatorname{vec} X$, where $\operatorname{vec} X \in \mathbb{R}^{n^2}$ is obtained from $X \in \mathbb{R}^{n\times n}$ by stacking up its columns, see [71]. Problem 6.38 takes the following vectorized form, where we denote the vectorization of the matrices U, V, W with the corresponding small letters u, v, w. As usual, we identify the operator $\mathcal{A}$ with its matrix representation with respect to the standard basis on $\mathbb{R}^{2n^2}$.

Problem 6.45. *Given a symmetric matrix $W \in \mathbb{R}^{n\times n}$. Find a solution pair $u, v : [0, h] \to \mathbb{R}^{n^2}$ such that*

$$\begin{pmatrix}\dot{u}(t)\\ \dot{v}(t)\end{pmatrix} = \mathcal{A}\begin{pmatrix}u(t)\\ v(t)\end{pmatrix}, \qquad \mathcal{A} = \begin{pmatrix}A_0^\top \otimes I & A_1^\top \otimes I\\ -I\otimes A_1^\top & -I\otimes A_0^\top\end{pmatrix}, \tag{6.67}$$

holds with boundary conditions

$$M\begin{pmatrix}u(0)\\ v(0)\end{pmatrix}+N\begin{pmatrix}u(h)\\ v(h)\end{pmatrix}=\begin{pmatrix}-w\\ 0\end{pmatrix},\ M=\begin{pmatrix}A_0^\top\otimes I & A_1^\top\otimes I\\ I_{n^2} & 0\end{pmatrix},\ N=\begin{pmatrix}I\otimes A_1^\top & I\otimes A_0^\top\\ 0 & -I_{n^2}\end{pmatrix}, \tag{6.68}$$

where $u = \operatorname{vec} U$, $v = \operatorname{vec} V$, and $w = \operatorname{vec} W$.

Let us show that this problem is the matrix counterpart of Problem 6.38.

Lemma 6.46. *The pair* $(U,V):\mathbb{R}\to\mathbb{R}^{n\times n}\times\mathbb{R}^{n\times n}$ *solves* (6.58) *with* $\mathcal{B}\binom{U(0)}{V(0)}=\binom{-W}{0}$ *if and only if* $\binom{\operatorname{vec}U}{\operatorname{vec}V}=\binom{u}{v}:\mathbb{R}\to\mathbb{R}^{2n^2}$ *is a solution of* (6.67) *with* $(M+Ne^{\mathcal{A}h})\binom{u(0)}{v(0)}=\binom{-\operatorname{vec}W}{0}$.

Proof. It is easy to see that the system matrix in (6.67) is the matrix representation of the operator $\mathcal{A}$ of (6.58). Let us show that the matrix representation of $\mathcal{B}$ is given by $M+e^{\mathcal{A}h}N$, hence showing that both problems are equivalent. The matrix representation of the projections π_1,π_2 is given by $(I_{n^2}\quad 0_{n^2})$ and $(0_{n^2}\quad I_{n^2})$, respectively. Hence $\mathcal{B}_2=\pi_1-\pi_2e^{\mathcal{A}h}$ has a matrix representation given by $(I_{n^2}\quad 0)-(0\quad I_{n^2})e^{\mathcal{A}h}$. Now $\mathcal{B}_1$ satisfies $\mathcal{B}_1=\mathcal{B}_2\mathcal{A}$. As $\mathcal{A}$ commutes with $e^{\mathcal{A}h}$, we get the following matrix representation of $\mathcal{B}_1$,

$$\left[(I_{n^2}\ 0)-(0\ I_{n^2})e^{\mathcal{A}h}\right]\mathcal{A}=(I_{n^2}\ 0)\mathcal{A}-(0\ I_{n^2})\mathcal{A}e^{\mathcal{A}h}=(A_0^\top\otimes I\ A_1^\top\otimes I)+(I\otimes A_1^\top\ I\otimes A_0^\top)e^{\mathcal{A}h},$$

where we already used the matrix representation of $\mathcal{A}$. Now $\mathcal{B}=\binom{\mathcal{B}_1}{\mathcal{B}_2}$ so that we can identify it with the matrix

$$\mathcal{B}=\begin{pmatrix}A_0^\top\otimes I & A_1^\top\otimes I\\ I_{n^2} & 0\end{pmatrix}+\begin{pmatrix}I\otimes A_1^\top & I\otimes A_0^\top\\ 0 & -I_{n^2}\end{pmatrix}e^{\mathcal{A}h}=M+Ne^{\mathcal{A}h}.$$

Thus both the system operator $\mathcal{A}$ and the boundary operator $\mathcal{B}$ are represented by their matrix counterpart. Hence Problems 6.38 and 6.45 are equivalent. □

From Problem 6.45 and the discussion of the boundary operator $\mathcal{B}$ following Problem 6.38 we immediately obtain the following existence and uniqueness result.

Corollary 6.47. *Problem 6.45 has a uniquely determined solution if and only if the boundary operator* $\mathcal{B}=M+Ne^{\mathcal{A}h}$ *is invertible. If* $\mathcal{B}$ *is singular then for a given* w *there exist multiple solutions if* $\binom{-w}{0}$ *is contained in the image of* $\mathcal{B}$, *otherwise there does not exist any solution satisfying* (6.67) *and* (6.68).

We now take a closer look at structure of the eigenvectors of the system matrix $\mathcal{A}$ in (6.67) or equivalently, of the operator $\mathcal{A}$ defined in (6.58).

Lemma 6.48. *Let* $\mathcal{A}$ *be the linear operator given by* (6.58). *If* $(\lambda_0,\binom{U_0}{V_0})$ *is an eigenpair of* $\mathcal{A}$, *then* $\binom{V_0^\top}{U_0^\top}$ *is an eigenvector of* $\mathcal{A}$ *corresponding to the eigenvalue* $-\lambda_0$.

Proof. Suppose that $\left(\lambda_0,\binom{U_0}{V_0}\right)$ is an eigenpair of $\mathcal{A}$ then

$$\mathcal{A}\begin{pmatrix}V_0^\top\\ U_0^\top\end{pmatrix}=\begin{pmatrix}V_0^\top A_0+U_0^\top A_1\\ -A_1^\top V_0^\top-A_0^\top U_0^\top\end{pmatrix}=\begin{pmatrix}(A_0^\top V_0+A_1^\top U_0)^\top\\ (-V_0A_1-U_0A_0)^\top\end{pmatrix}=-\lambda_0\begin{pmatrix}V_0^\top\\ U_0^\top\end{pmatrix},$$

i.e., $-\lambda_0$ is also an eigenvalue of $\mathcal{A}$ and the pair $\binom{V_0^\top}{U_0^\top}$ is a corresponding eigenvector. □

One can even show that $\mathcal{A}$ and $-\mathcal{A}$ are similiar, hence the Jordan structure of $\lambda \in \sigma(\mathcal{A})$ is identical to the Jordan structure of $-\lambda \in \sigma(\mathcal{A})$.
As the matrices A_0, A_1 are real, the spectrum of $\mathcal{A}$ contains with every λ also $-\lambda, \bar{\lambda}, -\bar{\lambda}$. This is reminiscent of the spectral properties of real Hamiltonian matrices. And indeed, if we consider the change of the arguments of the Kronecker product $A \otimes B \rightsquigarrow B \otimes A$ as some "quasi-transposition" then $\mathcal{A}$ is a "quasi-Hamiltonian" matrix.

Proposition 6.49. *Suppose that λ_0 is an eigenvalue of the linear operator $\mathcal{A}$ given by* (6.58) *and that $-\lambda_0 \notin \sigma(A_0)$. Then there exists an eigenvector of $\mathcal{A}$ corresponding to the eigenvalue λ_0 which is given by a pair of the form $\binom{Y_0}{\zeta_0 Y_0}$ where $Y_0 \in \mathbb{C}^{n\times n}$, $Y_0 \neq 0$, and $\zeta_0 \in \mathbb{C}$.*

Proof. The pair $\binom{U_0}{V_0}$ is an eigenvector of $\mathcal{A}$ corresponding to λ_0 if and only if the following system of equations is satisfied,

$$U_0(A_0 - \lambda_0 I) + V_0 A_1 = 0, \qquad A_1^\top U_0 + (\lambda_0 I + A_0^\top)V_0 = 0. \tag{6.69}$$

Let us introduce the linear operator

$$\mathcal{L} : \mathbb{C} \times \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}, \qquad \mathcal{L}(\lambda)X = (\lambda I + A_0^\top)X(\lambda I - A_0) + A_1^\top X A_1$$

Then using both equations of (6.69)

$$\begin{aligned}\mathcal{L}(\lambda_0)(U_0) &= (\lambda_0 I + A_0^\top)U_0(\lambda_0 I - A_0) - (\lambda_0 I + A_0^\top)V_0 A_1 = 0,\\ \mathcal{L}(\lambda_0)(V_0) &= (\lambda_0 I + A_0^\top)V_0(\lambda_0 I - A_0) - A_1^\top U_0(A_0 - \lambda_0 I) = 0.\end{aligned}$$

Hence both components of an eigenvector corresponding to λ_0 are contained in $\ker \mathcal{L}(\lambda_0)$. We can therefore define the following linear operator on the kernel of $\mathcal{L}(\lambda_0)$ (cf. (6.69))

$$\mathcal{M}(\lambda_0):\ \ker \mathcal{L}(\lambda_0) \to \ker \mathcal{L}(\lambda_0),\ U \mapsto V = -(\lambda_0 I + A_0^\top)^{-1} A_1^\top U. \tag{6.70}$$

For any $U \in \ker \mathcal{L}(\lambda_0)$, $U \neq 0$, the pair $\binom{U}{\mathcal{M}(\lambda_0)U}$ is an eigenvector of $\mathcal{A}$ corresponding to λ_0, because it satisfies (6.69),

$$\begin{aligned}U(A_0 - \lambda_0 I) + M(\lambda_0)UA_1 &= U(A_0 - \lambda_0 I) - (\lambda_0 I + A_0^\top)^{-1}A_1^\top U A_1\\ &= (\lambda_0 I + A_0^\top)^{-1}\big((\lambda_0 I + A_0^\top)U(A_0 - \lambda_0 I) - A_1^\top U A_1\big) = 0,\\ A_1^\top U + (\lambda_0 I + A_0^\top)M(\lambda_0)U &= A_1^\top U - (\lambda_0 I + A_0^\top)(\lambda_0 I + A_0^\top)^{-1}A_1^\top U = 0.\end{aligned}$$

Now, the linear operator $\mathcal{M}(\lambda_0)$ possesses an eigenvector Y_0 with $\mathcal{M}(\lambda_0)Y_0 = \zeta_0 Y_0$. Hence there exists an eigenvector $\binom{Y_0}{\zeta_0 Y_0}$ of $\mathcal{A}$ corresponding to λ_0 which is constructed from the eigenpair (ζ_0, Y_0) of $\mathcal{M}(\lambda_0)$. □

Remark 6.50. The condition $-\lambda \notin \sigma(A_0)$ can be replaced with the following alternatives.

1. If $\lambda_0 \in \sigma(\mathcal{A})$, but $\lambda_0 \notin \sigma(A_0)$ then there exists an eigenvector of $\mathcal{A}$ corresponding to λ_0 which is of the form $\binom{\zeta_0 Y_0}{Y_0}$, $\zeta_0 \in \mathbb{C}$. This can be seen by replacing $\mathcal{M}(\lambda_0)$ of (6.70) by
$$\mathcal{M}'(\lambda_0): \ \ker \mathcal{L}(\lambda_0) \to \ker \mathcal{L}(\lambda_0), \ V \mapsto VA_1(\lambda_0 I - A_0)^{-1}.$$

2. If A_1 is a regular matrix then the conditions $-\lambda_0 \notin \sigma(A_0)$ or $\lambda_0 \notin \sigma(A_0)$ can be dropped. Here $\mathcal{M}''(\lambda_0): \ U \mapsto U(\lambda_0 I - A_0)A_1^{-1}$ replaces (6.70).

3. If A_1 is singular and $\lambda_0 \in \sigma(A_0) \cap \sigma(-A_0)$ then eigenvectors of $\sigma(\mathcal{A})$ corresponding to λ_0 can be constructed explicitly: they are formed by pairs $(yv^\top, 0)$ and $(0, xy^\top)$ where $A_1^\top y = 0$, $(A_0^\top - \lambda_0 I)v = 0$, and $(A_0^\top + \lambda_0 I)x = 0$.

Hence without any assumption on the locations of the eigenvalues, the result of Proposition 6.49 can be generalized in such way that for every $\lambda_0 \in \mathcal{A}$ there exists a corresponding eigenvector which is of the form $\binom{U}{\zeta U}$, $\binom{\zeta V}{V}$, $\binom{U}{0}$ or $\binom{0}{V}$.

Remark 6.51. Let us now comment on how to compute the delay Liapunov function U for the one-delay system (6.1').

1. Set up the system matrix $\mathcal{A}$ and the boundary matrices M, N according to the data given in Problem 6.45.

2. Test if the boundary matrix $\mathcal{B} = M + Ne^{\mathcal{A}h}$ is invertible. If it is not invertible then the solution of linear equation $\mathcal{B}x_0 = \binom{-w}{0}$ for x_0 in the next step may fail.

3. Compute an initial value $x_0 \in \mathbb{R}^{2n^2}$ via $\mathcal{B}x_0 = \binom{-w}{0}$.

4. Solve the system of linear ordinary differential equations $\dot{x}(t) = \mathcal{A}x(t)$ on $t \in [0, h/2]$ with $x(0) = x_0$.

5. Join the solution segments contained in $x(t) = \binom{u(t)}{v(t)}$: As $U(t) = V^\top(h-t)$ we have $\operatorname{vec} U(t) = u(t)$ and $\operatorname{vec} U^\top(h-t) = v(t)$ for for $t \in [0, h/2]$.

In Step 3, the quasi-Hamiltonian structure of $\mathcal{A}$ (see the discussion following Lemma 6.48) implies that the computation of the matrix exponential can be ill-conditioned for even relative small values of h. Namely for every $\lambda \in \sigma(\mathcal{A})$ with $\operatorname{Re}\lambda > 0$, the negative value $-\lambda \in \sigma(\mathcal{A})$ is also contained in the spectrum. Hence the spectrum of the matrix exponential contains both eigenvalues of small modulus, $e^{-\lambda h}$, and eigenvalues of large modulus, $e^{\lambda h}$. The matrix $\mathcal{A}$ has a sparse structure. This should be honored when solving $\dot{x} = \mathcal{A}x$ in order to keep computational costs and storage requirements small.
We close this section with an example.

Example 6.52. Consider the 2×2 delay equation

$$\dot{x}(t) = \begin{pmatrix} 0 & 1 \\ -4 & -1 \end{pmatrix} x(t) + \begin{pmatrix} 0 & 0 \\ 2 & 1 \end{pmatrix} x(t-h), \tag{6.71}$$

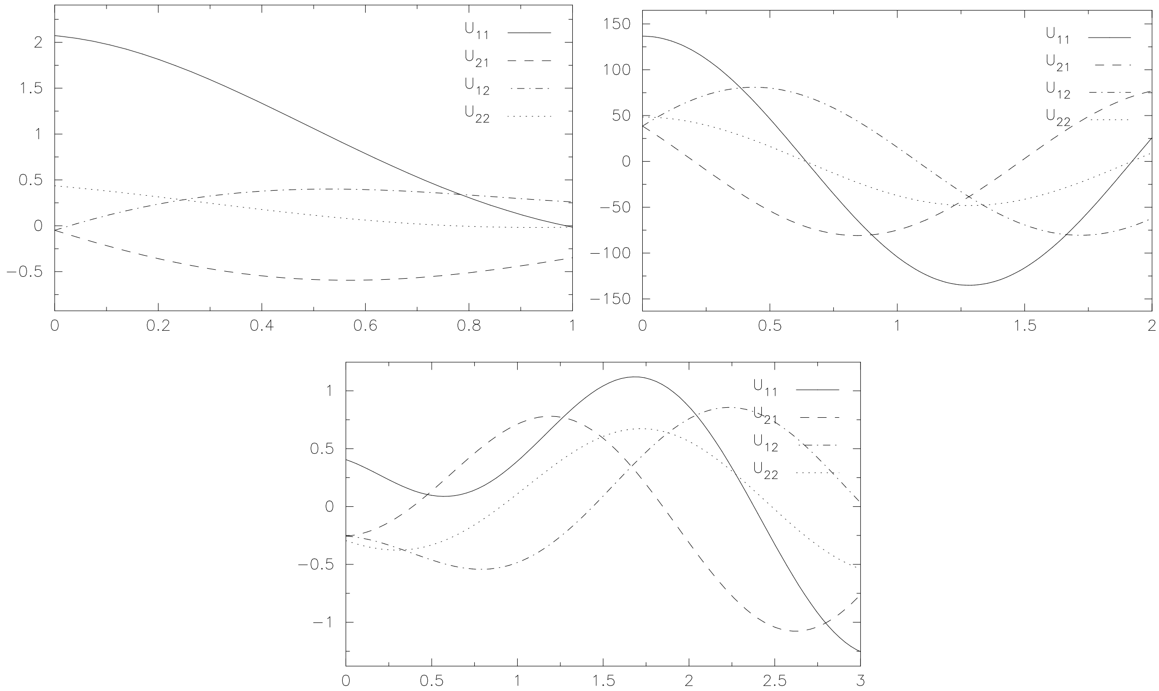

Figure 6.4: Delay Liapunov matrices for $h = 1, 2, 3$.

which has been discussed in [98]. Without delay, it is a pure oscillator, and it is stable for positive delays $h < 2.006$. Now following Remark 6.51 we can compute a solution of Problem 6.45 by solving an initial value problem.
Figure 6.4 shows the components of a delay Liapunov function for $h = 1, 2, 3$ corresponding to the weight $W = I$. For $h = 2$ the delay system is close to instability, and the norm of U is relatively large. For $h = 3$ there still exists a uniquely determined U, but note that the matrix $U(0)$ is not positive definite. ■

6.5 Uncertain Delays

In this section we consider the case when the delay $h \geq 0$ of the system

$$\Sigma_h : \qquad \dot{x}(t) = A_0 x(t) + A_1 x(t-h), \qquad t \geq 0, \tag{6.1'}$$

is not exactly known. Hence we are looking for some robust results for the existence of delay Liapunov matrices and for robust results on stability. As it turns out, these two problems are closely related.
To indicate the dependence on h, let us define the spectrum and the spectral abscissa of the delay system (6.1')

$$\begin{aligned} \sigma(\Sigma_h) &= \left\{ s \in \mathbb{C} \,\middle|\, \det(sI - A_0 - e^{-sh} A_1) = 0 \right\}, \\ \alpha(\Sigma_h) &= \sup \left\{ \operatorname{Re} s \,|\, s \in \sigma(\Sigma_h) \right\}. \end{aligned}$$

With this definition, (6.1') is exponentially stable if and only if $\alpha(\Sigma_h) < 0$.
Let us study the dependence of the spectrum on the delay h. Most authors dealing with variable delays implicitly use the following conjecture for which there seems to be currently no rigouros proof.

Conjecture 6.53. The map $h \mapsto \alpha(\Sigma_h)$ is continuous on $\mathbb{R}_+$.

In the following we will assume that this conjecture holds true. Let us present some ideas which could be used for a proof. The main obstacle for a proof is a missing global parameterisation of the spectra $h \mapsto \sigma(\Sigma_h)$. However, we have the following local result.

Lemma 6.54. *Given $\alpha \in \mathbb{R}$. Then the spectra of Σ_h restricted to $\mathbb{C}_{>\alpha}$ decompose into finitely many continuous branches. In particular, let us define*

$$\mathcal{N}_\alpha(h) = \left\{ s \in \mathbb{C} \,\middle|\, \det(sI - A_0 - e^{-sh}A_1) = 0 \text{ and} \operatorname{Re} s > \alpha \right\}.$$

Then for a given $h^0 \geq 0$ there exist r continuous functions $\lambda_i : I_i^* \to \mathbb{C}_{\geq\alpha}$, $i = 1, \dots, r$ with a suitable interval of maximal existence $I_i^* = [h_-^i, h_+^i] \subset \mathbb{R}_+$, $h^0 \in I_i^*$, such that $\mathcal{N}_\alpha(h^0) = \{\lambda_i(h^0) \,|\, i = 1, \dots, r\}$ and $\det(\lambda_i(h)I - A_0 - e^{-h\lambda_i(h)}A_1) = 0$ for $h \in I_i^*, i = 1, \dots, r$. Here, if $h_-^i > 0$ then $\lambda_i(h_-^i) = \alpha$ and if $h_+^i < \infty$ then $\lambda_i(h_+^i) = \alpha$.

Proof. For fixed α and h^0 the set $\mathcal{N}_\alpha(h^0)$ is finite, see Theorem 6.13. Hence the remarks from [77, Section IV.3.5] allow us to apply the finite-dimensional decomposition [77, Theorem II.5.2] into continuous functions. If a branch ceases to exist then it has to leave $\mathbb{C}_{>\alpha}$ which gives the conditions on h_-^i and h_+^i. □

Now to prove the continuity of the spectral abscissa we have to take new snapshots of parameterisations whenever the number of zeros in $\mathbb{C}_{\geq\alpha}$ changes. For this, let us assume that α is such that $\mathbb{C}_{>\alpha}$ contains an eigenvalue of $A_0 + A_1$. By Lemma 6.54 the spectral abscissa is continuous if the number of zeros inside $\mathbb{C}_{\geq\alpha}$ does not change since the maximum of finitely many continuous functions is continuous. The change of zeros can be detected as follows. Consider the *shifted system*

$$\Sigma_h^\alpha : \quad \dot{x}(t) = (A_0 - \alpha I)x(t) + e^{-\alpha h_1} A_1 x(t - h_1)$$

which satisfies $\sigma(\Sigma_h^\alpha) = \sigma(\Sigma_h) - \alpha$. Hence instead of detecting zeros of $h \mapsto \sigma(\Sigma_h)$ passing through $\alpha + i\mathbb{R}$ we consider zeros of $h \mapsto \sigma(\Sigma_h^\alpha)$ passing through the imaginary axis. Theorem 6.42 *(iii)* and *(iv)* provides a method of testing for this situation which yields critical points if the equation $\zeta = e^{-\lambda h}$ is satisfied where both ζ and $\lambda \in i\mathbb{R}$ are obtained from the finite-dimensional operator $\mathcal{A}$ derived from Σ_h^α. Thus new snapshots of the spectrum have be taken at isolated critical delays, and therefore the spectral abscissa is a continuous function of the delay h.

6.5.1 Critical Delays

From our previous analysis in Theorem 6.42 we obtain the following sets of critical delays.

Definition 6.55. The set of *critical delays* of (6.1') is given by

$$H_{\text{crit}} = \left\{ h \in \mathbb{R}_+ \,\middle|\, \det(M + Ne^{\mathcal{A}h}) = 0 \right\}.$$

By Theorem 6.42 we have the following characterizations of the set of critical delays.

Corollary 6.56. *The following statements are equivalent.*

(i) $h \in H_{crit}$.

(ii) Problem 6.37 has either no or infinitely many solutions for the delay h.

(iii) The minimal singular value of the boundary matrix satisfies $\sigma_{\min}(M + Ne^{\mathcal{A}h}) = 0$.

(iv) There exists $\lambda \in \mathbb{C}$ *which satisfies* $\det(\pm\lambda I - A_0 - e^{\mp\lambda h}A_1) = 0$. *In this case,* $\lambda \in \sigma(\mathcal{A})$.

(v) There exists an eigenpair $(\lambda, \binom{u}{\zeta u}) \in \mathbb{C} \times \mathbb{C}^{2n^2}$ *of* $\mathcal{A}$ *such that* $\zeta = e^{-\lambda h}$.

Proof. Clearly, since $\mathcal{A}$ of (6.67) is the matrix representation of $\mathcal{A}$ of (6.58), their spectra coincide. We have already verified that the matrix representation of the boundary condition $\mathcal{B}$ is $M + Ne^{\mathcal{A}h}$. Thus the statements follow directly from Theorem 6.42, Corollary 6.44 and Lemma 6.46. □

By Corollary 6.56 *(v)* we can rewrite H_{crit} as

$$H_{\text{crit}} = \left\{ h \geq 0 \,\middle|\, \text{there exists an eigenpair } (\lambda, \tbinom{u}{\zeta u}) \text{ of } \mathcal{A} \text{ with } \zeta = e^{-\lambda h} \right\}. \tag{6.72}$$

If $\sigma(\Sigma_h) \cap i\mathbb{R} \neq \emptyset$ then h is critical. Let us therefore consider those critical delays which belong to purely imaginary eigenvalues $\lambda = i\omega$, of $\mathcal{A}$. We first discuss the case $\omega = 0$.

Corollary 6.57. *If* $0 \in \sigma(\mathcal{A})$ *then* $H_{crit} = \mathbb{R}_+$.

Proof. If $0 \in \sigma(\mathcal{A})$ then there exists an eigenvector of the form $\binom{U}{U}$ associated with the eigenvalue 0 of $\mathcal{A}$. This yields $U(A_0 + A_1) = 0$ and $-(A_1^\top + A_0^\top)U = 0$. Therefore $\det(-A_0 - A_1) = \det(0I - A_0 - e^{0 \cdot h}A_1) = 0$ which is independent of $h \geq 0$. Therefore all $h \geq 0$ are critical. □

For the rest of this discussion let us assume that $A_0 + A_1$ is *regular*. We now consider eigenvalues $\lambda = i\omega$, $\lambda \neq 0$, of $\mathcal{A}$.

Proposition 6.58. *The following statements are equivalent.*

(i) There exists $i\omega \in \sigma(\Sigma_h) \setminus \{0\}$.

(ii) There exists an eigenvector $\binom{U}{\zeta U}$ *of* $\mathcal{A}$ *associated with* $i\omega$ *such that* $U = U^*$.

In this case, all delays of the form $h + \frac{2\pi}{|\omega|}\ell \geq 0$ *with* $\ell \in \mathbb{Z}$ *are critical.*

Proof. (i) $\Longrightarrow$ (ii). Let us assume that h is a critical delay of Σ_h such that $\det(i\omega I - A_0 - e^{-i\omega h}A_1) = 0$ for a suitable $\omega \neq 0$, see also Corollary 6.56 *(iv)*. We find a vector $v \in \mathbb{C}^n$ such that

$$v^*(i\omega I - A_0 - A_1 e^{-i\omega h}) = 0.$$

Complex conjugation yields $-v^\top(i\omega + A_0 + A_1 e^{i\omega h}) = 0$. The Hermitian matrix $U = vv^*$ then induces an eigenvector $\left(\begin{smallmatrix} U \\ e^{-i\omega h}U \end{smallmatrix}\right)$ associated with the eigenvalue $i\omega$ of $\mathcal{A}$, see the proof of Theorem 6.42. This shows *(ii)*.

For *(ii) $\Longrightarrow$ (i)* consider an eigenvector $\left(\begin{smallmatrix} U \\ \zeta U \end{smallmatrix}\right)$ of $\mathcal{A}$ with Hermitian component U. The scaling factor $\zeta \in \mathbb{C}$ satisfies $|\zeta| = 1$ as by Lemma 6.48, $\zeta^{-1} = \bar{\zeta}$ holds. If $v \in \mathbb{C}^n$ is an eigenvector associated with a non-trivial eigenvalue of U then $\mathcal{A}\left(\begin{smallmatrix} U \\ \zeta U \end{smallmatrix}\right) = i\omega\left(\begin{smallmatrix} U \\ \zeta U \end{smallmatrix}\right)$ implies that $v^*(i\omega I - A_0 - \zeta A_1) = 0$ and $(i\omega I + A_0^\top + \zeta^{-1}A_1^\top)v = 0$. Hence as $|\zeta| = 1$, there are infinitely many solutions $h > 0$ with $\zeta = e^{-i\omega h}$ which are all critical delays of Σ_h by Corollary 6.56 *(iv)*. These solutions occur periodically with a period length of $\frac{2\pi}{|\omega|}$. □

Unfortunately it is currently not clear if an imaginary eigenvalue of $\mathcal{A}$ always leads to critical delays. It is easy to see that if $\left(\begin{smallmatrix} U \\ \zeta U \end{smallmatrix}\right)$ is an eigenvector corresponding to $i\omega \in \sigma(\mathcal{A})$ then $\left(\begin{smallmatrix} \bar{\zeta}U^* \\ U^* \end{smallmatrix}\right)$ is an eigenvector corresponding to the eigenvalue $-\overline{i\omega} = i\omega$ of $\mathcal{A}$, see Lemma 6.48 and Proposition 6.49 including Remark 6.50. If the eigenspace is assumed to be of dimension 1, $U = U^*$ and $\bar{\zeta} = \zeta^{-1}$, which shows that in this case any $h \geq 0$ satisfying $\zeta = e^{-i\omega h}$ is critical. For eigenspaces of higher dimension, the situation is not clear. However, note that

$$\ker \mathcal{L}(i\omega) = \{X \in \mathbb{C}^{n\times n} \,|\, (i\omega I - A_0)^* X (i\omega I - A_0) = A_1^\top X A_1\}$$

enforces a Hermitian/skew-Hermitian structure on the components of the eigenvectors of $\mathcal{A}$ corresponding to $i\omega \in \sigma(\mathcal{A})$. In particular, if $X \in \ker \mathcal{L}(i\omega)$ then also $X + X^*, X - X^* \in \ker \mathcal{L}(i\omega)$.

We can split H_{crit} into periodic and aperiodic critical delays $H_{\text{crit}} = H_{\text{per}} \cup H_{\text{aper}}$. The set of *periodic critical delays* is given by

$$H_{\text{per}} = \bigcup_{i\omega\in\sigma(\mathcal{A})\setminus\{0\}} \left\{ h + \tfrac{2\pi\ell}{|\omega|} \in \mathbb{R}_+ \,\middle|\, h \in [0, \tfrac{2\pi}{|\omega|}), \ell \in \mathbb{N}, \, (i\omega, \left(\begin{smallmatrix} U \\ e^{-i\omega h}U \end{smallmatrix}\right)) \text{ is an eigenpair of } \mathcal{A} \right\}.$$

The set of *aperiodic critical delays*

$$H_{\text{aper}} = \bigcup_{\lambda\in\sigma(\mathcal{A})\setminus i\mathbb{R}^*} \left\{ h \in \mathbb{R}_+ \,\middle|\, \text{there exists an eigenpair } (\lambda, \left(\begin{smallmatrix} U \\ e^{-\lambda h}U \end{smallmatrix}\right)) \text{ of } \mathcal{A} \right\}$$

only contains finitely many points under the regularity assumption of $A_0 + A_1$. Aperiodic critical delays can only occur in unstable systems, hence they are of no importance for stability considerations. To see this, assume that $\lambda \in \sigma(\Sigma_h)$ is not a purely imaginary eigenvalue associated with a critical delay h. By Corollary 6.56, $-\lambda \in \sigma(\Sigma_h)$ which implies that Σ_h has an instable eigenvalue with positive real part.

Hence if $H_{\mathrm{crit}} \neq \mathbb{R}_+$ then it does not contain any finite accumulation point, and its elements can be ordered increasingly. Let us introduce for all $h \geq 0$,

$$\lfloor h \rfloor_{\mathrm{crit}} := \sup \{h' \in H_{\mathrm{crit}} \,|\, h' < h\}, \qquad \lceil h \rceil_{\mathrm{crit}} := \inf \{h' \in H_{\mathrm{crit}} \,|\, h' > h\},$$

with the convention that $\inf \emptyset = +\infty$ and $\sup \emptyset = -\infty$.
Let us now determine maximal delay intervals for which Σ_h is stable provided that Conjecture 6.53 holds.

Proposition 6.59. *Suppose that Σ_{h_0} is exponentially stable for some $h_0 \geq 0$, then Σ_h is exponentially stable for all $h \in (\lfloor h_0 \rfloor_{crit}, \lceil h_0 \rceil_{crit}) \cap \mathbb{R}_+$.*

Proof. The set H_{crit} is a discrete subset of $\mathbb{R}_+$, as by assumption Σ_{h_0} is exponentially stable, hence $H_{\mathrm{crit}} \neq \mathbb{R}_+$. Now let us suppose that $h > h_0$ and Σ_h is not exponentially stable. Using Conjecture 6.53 the minimal $\tilde{h} \in (h_0, h]$ with this property satisfies $\alpha(\Sigma_{\tilde{h}}) = 0$. Hence there exists $\omega > 0$ with $\pm i\omega \in \sigma(\Sigma_{\tilde{h}})$. Corollary 6.44 and Corollary 6.56 *(iv)* imply that $\tilde{h} \in H_{\mathrm{crit}}$, hence $\tilde{h} = \lceil h_0 \rceil_{\mathrm{crit}}$. For the lower bound, analogous results hold. □

If $\alpha(\Sigma_h) < 0$ and $\lceil h \rceil_{\mathrm{crit}} = +\infty$ then the delay system $\Sigma_{h'}$ is exponentially stable for all $h' \geq h$, and if $\alpha(\Sigma_h) < 0$ and $\lfloor h \rfloor_{\mathrm{crit}} = -\infty$ then $\Sigma_{h'}$ is exponentially stable for all $h' \in [0, h]$. We obtain the following criterion for the exponential stability independent of delay. For other conditions of delay-independent stability see Bliman [19],and Hertz et al. [55].

Corollary 6.60. *Suppose that the eigenspaces associated with imaginary eigenvalues of $\mathcal{A}$ are of dimension 1. Then the delay equation* (6.1') *is exponentially stable independent of delay (i.o.d.) if and only if $A_0 + A_1$ is exponentially stable and $\sigma(\mathcal{A}) \cap i\mathbb{R} = \emptyset$.*

Proof. Using Proposition 6.59 and its proof it remains to show the necessity of the condition $\sigma(\mathcal{A}) \cap i\mathbb{R} = \emptyset$. But by Proposition 6.49 (including Remark 6.50) and Proposition 6.58 each purely imaginary eigenvalue $i\omega$, $\omega \neq 0$ of $\mathcal{A}$ gives rise to some critical delay h' with $\alpha(\Sigma_{h'}) \geq 0$. Hence (6.1') is exponentially stable i.o.d if and only if $H_{\mathrm{crit}} = \emptyset$. □

Note that a stability criterion listed in [55] requires to test $Q(s,z) := \det(sI - A_0 - zA_1) \neq 0$ for all $\{(s,z) \,|\, \mathrm{Re}\, s = 0, |z| = 1\}$ while Corollary 6.60 only has a finite number of tests.

Remark 6.61. (i) There are eigenvalues of the delay equation (6.1') which coincide with eigenvalues of the matrix $\mathcal{A}$ if the delay Liapunov matrix is not uniquely determined, see Corollary 6.44. Hence the finite-dimensional system $\dot{x} = \mathcal{A}x$ together with the symmetry condition has some kind of resonance with the delay system, more or less like a candle placed between two parallel mirrors gives the impression of infinitely many candles. This coincidence also implies that for a varying delay all critical spectra of the delay equation (6.1') have to pass through finitely many holes in $i\mathbb{R}$ punched by $\sigma(\mathcal{A})$ which is illustrated in Figure 6.5, see the next section.

(ii) Corollary 6.56 allows us to compute critical delays directly from an analysis of the matrix $\mathcal{A}$. For every eigenvalue λ of $\mathcal{A}$ one has to find the scaling factor ζ between the two components of an associated eigenvectors $\binom{U}{\zeta U}$ The set of critical delays can be computed via (6.72). For a stability analysis, only purely imaginary eigenvalues of $\mathcal{A}$ are of interest, see Corollary 6.60.

Definition 6.62. Suppose that Σ_h is exponentially stable for $h = 0$. The constant $0 < h^* \leq \infty$ is called the *delay margin* for Σ_h if $[0, h^*)$ is the maximal interval such that Σ_h is exponentially stable for all $h \in [0, h^*)$.

From Proposition 6.59, we immediately obtain the following corollary.

Corollary 6.63. *Suppose that $A_0 + A_1$ is exponentially stable. Then the delay margin of* (6.1') *is given by $h^* = \min H_{crit}$.*

Note that the delay margin is a periodic critical delay, as h can only be an aperiodic critical delay if the delay system Σ_h satisfies $\sigma(\Sigma_h) \cap \mathbb{C}_+ \neq \emptyset$.

Example 6.64. This example shows the existence of an aperiodic critical point, i.e., $h \in H_{\text{crit}}$ does not correspond to a purely imaginary eigenvalue of $\mathcal{A}$. Consider the matrices

$$A_0 = \begin{pmatrix} -\alpha & 3-\alpha \\ -\alpha & 2-\alpha \end{pmatrix} \quad \text{and} \quad A_1 = \begin{pmatrix} -\alpha/2 & 1+\alpha \\ 0 & \alpha-4 \end{pmatrix}.$$

Using these matrices for the delay equation (6.1') we can start a numerical parameter study. And indeed, for $\alpha_0 = 1.17003$ we obtain a real solution $h = 1.6048$ of $\zeta = e^{-\lambda h}$ where both $\lambda, \zeta \in \mathbb{C}$ are derived from the spectrum of $\mathcal{A}$, but contrary to periodic solutions, $\lambda = 0.77330 + 1.3434i$ is *not* purely imaginary. Here $\zeta = -0.15965 - 0.24103i$. Varying α about the critical value α_0 , we observe a sign change in the imaginary part of the solution h, so that this solution is not only numerically close to a real solution, but there exists indeed a real solution in the vicinity of the given parameter α_0. Figure 6.5 shows the root locus for varying h, some roots leave the left half-plane through the holes at $\pm 3.71i \in \sigma(\mathcal{A})$ for critical values of $0.317 + 1.69\ell$, $\ell = 1, 2, 3, \ldots$ But for $h = 1.6048$ four roots of the delay equation hit the spectrum of $\mathcal{A}$ which is marked by circles in Figure 6.5. For this delay there exist no uniquely determined delay Liapunov matrix. ■

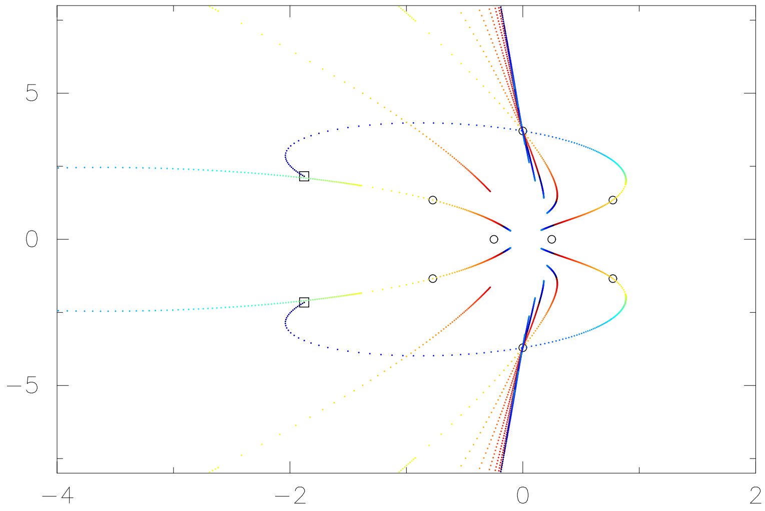

Figure 6.5: Root locus of a delay equation with varying h.

6.5.2 Spectrum under the Variation of the Delay

We have seen in Proposition 6.59 that some accumulation points of the eigenvalues of Σ_h as $h \to \infty$ are given by eigenvalues $i\omega \in \sigma(\mathcal{A})$. Let us now study other properties of $\sigma(\Sigma_h)$ as the delay term h varies. The following example shows that there exist delay equation where the variation of the delay has no influence on the spectrum.

Example 6.65. Consider the matrices $A_0 = \left(\begin{smallmatrix} \lambda_1 & \alpha_1 \\ 0 & \lambda_2 \end{smallmatrix}\right)$ and $A_1 = \left(\begin{smallmatrix} 0 & \alpha_2 \\ 0 & 0 \end{smallmatrix}\right)$ where $\lambda_1, \lambda_2, \alpha_1, \alpha_2 \in \mathbb{C}$. Then $\sigma(A_0) = \sigma(A_0 + A_1) = \sigma(\Sigma_h)$ as

$$\sigma(\Sigma_h) = \left\{ s \in \mathbb{C} \,\middle|\, \det \begin{pmatrix} s - \lambda_1 & -\alpha_1 - \alpha_2 e^{-sh} \\ 0 & s - \lambda_2 \end{pmatrix} = 0 \right\} = \{\lambda_1, \lambda_2\}. \tag{6.73}$$

■

We have already seen the continuous dependency of some branches of the spectrum of Σ_h on the delay h in Proposition 6.54. For the following analysis, let us recall the Implicit Function Theorem, see Dieudonné [34], which provides us with the following result.

Lemma 6.66. *Define the continuously differentiable function*

$$f(h, \lambda) = \det(\lambda I - A_0 - e^{-h\lambda} A_1). \tag{6.74}$$

Suppose that $(h_*, \lambda_*) \in \mathbb{R}_+ \times \mathbb{C}$ satisfy $f(h_*, \lambda_*) = 0$ and $f_\lambda(h_*, \lambda_*) \neq 0$ where $f_\lambda = \frac{\partial f}{\partial \lambda}$. Then there exists a continuously differentiable function $\lambda(h) : I_* \to \mathbb{C}$ on an open interval $I_* \ni h_*$ which satisfies $\lambda(h_*) = \lambda_*$ and $f(h, \lambda(h)) = 0$ for all $h \in I_*$. The derivative in h_* is given by $\lambda'(h_*) = -\frac{f_h}{f_\lambda}(h_*, \lambda_*)$.

We now analyse if the roots enter or leave the left half-plane when passing through the imaginary axis.

Proposition 6.67. *Define $f(h, \lambda) := \det(\lambda I - A_0 - e^{-h\lambda} A_1)$. If $i\omega_* \in \sigma(\mathcal{A})$ is a simple eigenvalue of Σ_{h_*}, i.e., $f_\lambda(h_*, i\omega_*) \neq 0$, for some $h_* > 0$ then the direction of a (local) root branch $\lambda(h)$ of $h \mapsto \sigma(\Sigma_h)$ crossing the imaginary axis through $i\omega_*$ is independent of h, i.e., the roots of Σ_h either always leave or always enter the left half plane through $i\omega_*$.*

Proof. We can write $f(h, \lambda)$ as a polynomial $p(\lambda, \zeta) = \det(\lambda I - A_0 - \zeta A_1)$ in λ and $\zeta = e^{-h\lambda}$. The derivative of a root branch $\lambda(h)$ in h_* is then given by

$$\lambda'(h_*) = -\frac{f_h(h_*, \lambda_*)}{f_\lambda(h_*, \lambda_*)} = \frac{\lambda_* \zeta_* p_\zeta(\lambda_*, \zeta_*)}{p_\lambda(\lambda_*, \zeta_*) - h_* \zeta_* p_\zeta(\lambda_*, \zeta_*)}, \tag{6.75}$$

where $\lambda_* = \lambda(h_*)$ and $\zeta_* = e^{-h_* \lambda_*}$, see Lemma 6.66. Here the partial derivatives satisfy $f_h(h_*, \lambda_*) = p_\zeta(\lambda_*, \zeta_*)\zeta_*(-\lambda_*)$ and $f_\lambda(h_*, \lambda_*) = p_\lambda(\lambda_*, \zeta_*) + p_\zeta(\lambda_*, \zeta_*)\zeta_*(-h_*)$. As we are interested in the direction in which a root branch crosses the imaginary axis, we are looking for the sign of the real part of the derivative in h_* with $\lambda_* = i\omega_*$. Since $\operatorname{sgn} \operatorname{Re} z = \operatorname{sgn} \operatorname{Re} z^{-1}$ we obtain from (6.75) that

$$\operatorname{sgn} \operatorname{Re} \lambda'(h_*) = \operatorname{sgn} \operatorname{Re} \left(\frac{p_\lambda(\lambda_*, \zeta_*)}{\lambda_* \zeta_* p_\zeta(\lambda_*, \zeta_*)} - \frac{h_*}{\lambda_*} \right). \tag{6.76}$$

But $h_*/\lambda_* = h_*/(i\omega_*)$ does not contribute to the real part of (6.76) as it is purely imaginary. Hence for $\omega_* > 0$ using $\operatorname{Re} \frac{z}{i} = \operatorname{Im} z$,

$$\operatorname{sgn} \operatorname{Re} \lambda'(h_*) = \operatorname{sgn} \operatorname{Re} \frac{p_\lambda(\lambda_*, \zeta_*)}{i\omega_* \zeta_* p_\zeta(\lambda_*, \zeta_*)} = \operatorname{sgn} \operatorname{Im} \frac{p_\lambda(\lambda_*, \zeta_*)}{p_\zeta(\lambda_*, \zeta_*)} \zeta_*^{-1} = -\operatorname{sgn} \operatorname{Im} \frac{p_\zeta(\lambda_*, \zeta_*)}{p_\lambda(\lambda_*, \zeta_*)} \zeta_*.$$

Clearly, this expression for $\operatorname{sgn} \operatorname{Re} \lambda'(h_*)$ only involves the data $\lambda_* = i\omega_*$ and ζ_* which can be obtained from $\mathcal{A}$ without calculating any critical delays h_* first, see Proposition 6.49. □

A rough method to compute the crossing direction is given by approximating the partial derivatives of $p(\lambda_*, \zeta_*)$ by difference quotients. One proceeds by choosing a small $\varepsilon > 0$ and computing

$$-\operatorname{sgn} \operatorname{Im} \zeta_* \frac{p(\lambda_*, \zeta_* + \varepsilon)}{p_\lambda(\lambda_* + \varepsilon, \zeta_*)} \quad \text{or} \quad -\operatorname{sgn} \operatorname{Im} \zeta_* \frac{p(\lambda_*, \zeta_* + \varepsilon) - p(\lambda_*, \zeta_*)}{p_\lambda(\lambda_* + \varepsilon, \zeta_*) - p(\lambda_*, \zeta_*)} \tag{6.77}$$

for all critical values obtained by an analysis of the matrix $\mathcal{A}$ of Problem 6.45. In the second formulation of (6.77) the term $p(\lambda_*, \zeta_*)$ is included which should theoretically be 0, but due to numerical errors it is not. The inclusion of this term may robustify the computation.

Numerical experiments suggest the following asymptotic behaviour of the root branches of $\sigma(\Sigma_h)$. However, due to the missing global parameterisation of these branches we do not provide a proof.

Conjecture 6.68. Suppose that $A_0 + A_1$ is exponentially stable and A_1 is regular. If $h \mapsto \lambda(h)$ is a continuous function from $\mathbb{R}_+$ to $\mathbb{C}$ such that $\lambda(h) \in \sigma(\Sigma_h)$ for all sufficiently large h then $\lim_{h\to\infty} \lambda(h) = 0$.

Remark 6.69. With the help of Proposition 6.49 and Proposition 6.67 we can trace the destabilization process of $h \mapsto \sigma(\Sigma_h)$. We introduce the inertia of a retarded linear delay system $(\pi(\Sigma_h), \iota(\Sigma_h)) \in \mathbb{N}^2$ which counts the number of zeros of the characteristic function $\chi_h = f(h, \cdot)$, see (6.74), in the right half plane and on the imaginary axis, respectively. As the number of zeros in the left half plane is infinite, we do not consider it part of the inertia. Now, an analysis of the operator $\mathcal{A}$ gives all periodic critical values h, for which $\sigma(\Sigma_h) \cap i\mathbb{R} \neq \emptyset$, and Proposition 6.67 shows in which direction the imaginary axis is traversed. Thus, for each delay h the pair $(\pi(\Sigma_h), \iota(\Sigma_h))$ is known.

6.6 Multiple Delays

Let us now consider the case of multiple delays, i.e., the delay equation is given by

$$\dot{x}(t) = \sum_{k=0}^{m} A_k x(t - h_k) \quad \text{where} \quad 0 = h_0 < h_1 < \cdots < h_m = H \quad \text{and} \quad A_k \in \mathbb{R}^{n\times n}. \tag{6.78}$$

The associated delay Liapunov function then satisfies the matrix delay equation

$$\dot{U}(t) = \sum_{k=0}^{m} U(t - h_k) A_k, \qquad t \geq 0, \tag{6.79}$$

the symmetric and algebraic conditions now read

$$U(-t) = U(t)^\top,\ t \geq 0, \qquad -W = U(0)A_0 + A_0^\top U(0) + \sum_{k=1}^{m} U(h_k)^\top A_k + A_k^\top U(h_k), \quad (6.80)$$

where $W \in \mathcal{H}_+^n(\mathbb{R})$ is a symmetric positive definite weight matrix. Unfortunately, for this general case there are no results available. If we do not assume asymptotic stability of (6.78) then existence and uniqueness issues of a delay Liapunov matrix have not been addressed yet in the literature.

6.6.1 Systems with Commensurable Delays

The results for the one-delay case can be extended to the multi-delay case if the symmetric condition allows us to extract a finite dimensional linear ODE from the matrix delay equation (6.79). This is possible in case of *commensurable delays*, i.e., the delays are given by $h_k = kh$, $k = 0, \ldots, m$. Let us assume that U is a solution of (6.79),(6.80) associated with the weight W. By defining

$$U_k(t) = U(t+kh), \qquad V_k(t) = U((k+1)h-t)^\top = U_k(h-t)^\top \qquad k = 0, \ldots, m-1 \quad (6.81)$$

the matrix delay equation (6.79) with respect to shifted time arguments can be written as

$$\begin{aligned}\dot{U}_k(t) = \dot{U}(t+kh) &= \sum_{j=0}^{m} U(t+kh-jh)A_j = \sum_{j=0}^{k} U_{k-j}(t)A_j + \sum_{j=k+1}^{m} U((j-k)h-t)^\top A_j \\ &= \sum_{j=0}^{k} U_{k-j}(t)A_j + \sum_{j=k+1}^{m} V_{j-k-1}(t)A_j, \qquad t \in [0,h]. \qquad (6.82)\end{aligned}$$

The differential equation for the counterflow V_k, $k = 0, \ldots, m-1$, is then given by

$$\begin{aligned}\dot{V}_k(t) = \tfrac{d}{dt}\left(U_k(h-t)\right)^\top &= -\sum_{j=0}^{k} A_j^\top U_{k-j}(h-t)^\top - \sum_{j=k+1}^{m} A_j^\top V_{j-k-1}(h-t)^\top \\ &= -\sum_{j=0}^{k} A_j^\top V_{k-j}(t) - \sum_{j=k+1}^{m} A_j^\top U_{j-k-1}(t), \qquad t \in [0,h]. \qquad (6.83)\end{aligned}$$

Hence the use of the symmetry condition $U(t) = U(-t)^\top$ gives us a system of $2m$ ordinary differential matrix equations, where all U_k and V_k are defined on $[0,h]$. The boundary conditions $U_{k-1}(h) = U_k(0)$ and $V_k(h) = V_{k-1}(0)$ for $k = 1, \ldots, m-1$ are needed to concatenate the solution segments. The symmetry condition is given by $U_0(0) = V_0(h)$ and the algebraic condition by

$$-W = \dot{U}_0(0) - \dot{V}_0(h) = U_0(0)A_0 + A_0^\top V_0(h) + \sum_{j=1}^{m} V_{j-1}(0)A_j + A_j^\top U_{j-1}(h).$$

Again, if there exists a unique solution of (6.82), (6.83) with the above-mentioned boundary conditions for a given W then it coincides with the solution of a delay Liapunov equation. This will be shown in the next proposition. Let us first collect the differential equations and the boundary conditions in the following problem formulation.

Problem 6.70. *For a given $W \in \mathcal{H}^n_+(\mathbb{R})$ find a solution of*

$$\begin{aligned}\dot U_k(t) &= \sum_{j=0}^{k} U_{k-j}(t)A_j + \sum_{j=k+1}^{m} V_{j-(k+1)}(t)A_j, \qquad && t \in [0,h], k = 0,1,\dots,m-1,\\ \dot V_k(t) &= -\sum_{j=0}^{k} A_j^\top V_{k-j}(t) - \sum_{j=k+1}^{m} A_j^\top U_{j-(k+1)}(t), && t \in [0,h], k = 0,1,\dots,m-1,\end{aligned}$$

which satisfies the following conditions

$$\begin{aligned}&U_0(0) = V_0(h), \quad U_{k-1}(h) = U_k(0), \quad V_{k-1}(0) = V_k(h), \qquad k = 1,\dots,m-1,\\ &\dot U_0(0) - \dot V_0(h) = U_0(0)A_0 + \sum_{j=1}^{m} V_{j-1}(0)A_j + A_0^\top V_0(h) + \sum_{j=1}^{m} A_j^\top U_{j-1}(h) = -W.\end{aligned}$$

As already mentioned, solutions of Problem 6.70 are intimately related to delay Liapunov matrices for (6.1) with commensurable delays.

Proposition 6.71. *Suppose the pairs $(U_k(\tau), V_k(\tau))_{k=0,\dots,m-1}$ solve Problem 6.70 on $[0,h]$ for a given symmetric matrix $W \in \mathcal{H}^n_+(\mathbb{R})$. Using the Gauss integer bracket $\lfloor t \rfloor = \max\{n \in \mathbb{Z} \mid n \le t\}$ the function*

$$U(t) = \begin{cases} \frac{1}{2}\left(U_{\lfloor t/h \rfloor}(t - \lfloor t/h \rfloor h) + V_{\lfloor t/h \rfloor}((1 + \lfloor t/h \rfloor)h - t)^\top\right), & t \ge 0,\\ U(-t)^\top, & t < 0,\end{cases} \tag{6.84}$$

is a solution of (6.79),(6.80). *On the other hand, if $U(t)$ is delay Liapunov matrix for* (6.78) *with $h_k = kh$ given as a solution to* (6.79),(6.80), *then*

$$U_k(\tau) = U(\tau+kh), \qquad V_k(\tau) = U((k+1)h-\tau)^\top = U_k(h-\tau)^\top, \qquad k = 0,\dots m-1, \tag{6.85}$$

form a solution of Problem 6.70. If this solution is uniquely determined, then $U_k(t) = V_k(h-t)^\top$ for all $k = 0,\dots,m-1$, $t \in [0,h]$.

Proof. Let $\{(U_k, V_k)\}$ be a solution of Problem 6.70 and U be given by (6.84). As $U_0(0) = V_0(h)$ the matrix $U(0) = \frac{1}{2}(U_0(0) + V_0(h)^\top)$ is symmetric, hence $t \mapsto U(t)$ is continuous in $t_0 = 0$ when setting $U(t) = U(-t)^\top$ for $t < 0$. For each $t \in [kh, (k+1)h)$ we have

$$U(t) = \tfrac{1}{2}\left(U_k(\tau) + V_k(h-\tau)^\top\right) \quad \text{with} \quad \tau = t - \lfloor t/h \rfloor h = t - kh.$$

Hence

$$\begin{aligned}\dot{U}(t) &= \tfrac{1}{2}\left(\sum_{j=0}^{k}(U_{k-j}(\tau) + V_{k-j}(h-\tau)^\top)A_j + \sum_{j=k+1}^{m}(V_{j-(k+1)}(\tau) + U_{j-(k+1)}(h-\tau)^\top)A_j\right)\\ &= \sum_{j=0}^{k} U(t-jh)A_j + \sum_{j=k+1}^{m} U(jh-t)^\top A_j = \sum_{j=0}^{m} U(t-jh)A_j.\end{aligned}$$

Therefore $U(t)$ satisfies the matrix differential equation (6.79). We now have to verify the algebraic condition (6.80). Since $U(0)$ is symmetric and by setting $U_m(0) := U_{m-1}(h), V_m(h) := V_{m-1}(0)$ we have

$$\begin{aligned}U(0)A_0 + A_0^\top U(0) + \sum_{k=1}^{m} U(kh)^\top A_k + A_k^\top U(kh) &= \sum_{k=0}^{m} U(kh)^\top A_k + A_k^\top U(kh)\\ &= \tfrac{1}{2}\left(\sum_{k=0}^{m}(U_k(0)^\top + V_k(h))A_k + A_k^\top(U_k(0) + V_k(h)^\top)\right)\\ &= \tfrac{1}{2}\sum_{k=0}^{m}\left(A_k^\top U_k(0) + V_k(h)A_k\right) + \tfrac{1}{2}\sum_{k=0}^{m}\left(U_k(0)^\top A_k + A_k^\top V_k(h)^\top\right)\\ &= -\tfrac{1}{2}(W + W^\top) = -W.\end{aligned}$$

Hence U is a delay Liapunov matrix for the weight W. On the other hand, since we obtained the formulation of Problem 6.70 by chopping a delay Liapunov matrix into pieces of length h, any Liapunov matrix for the weight W will also be a solution of Problem 6.70. If this solution is unique, then we obtain —analogously to Corollary 6.41— that $U_k(t) = V_k(h-t)^\top$ for all $k = 0, \dots, m-1$, $t \in [0, h]$. □

Now for further analysis, the equations of Problem 6.70 can be brought into matrix form by using Kronecker products.

Problem 6.72. *Consider the tuple $(U_{m-1}, \dots, U_1, U_0, V_0, V_1, \dots, V_{m-1})$. The system of ordinary differential equations corresponding to the vectorization of the conditions in Problem 6.70 then takes the form*

$$\dot{x} = \begin{pmatrix} A_0^\top \otimes I & \dots & A_{m-1}^\top \otimes I & A_m^\top \otimes I & & \\ & \ddots & & & \ddots & \\ & & A_0^\top \otimes I & A_1^\top \otimes I & \dots & A_m^\top \otimes I \\ -I \otimes A_m^\top & \dots & -I \otimes A_1^\top & -I \otimes A_0^\top & & \\ & \ddots & & & \ddots & \\ & & -I \otimes A_m^\top & -I \otimes A_{m-1}^\top & \dots & -I \otimes A_0^\top \end{pmatrix} x, \qquad x = \begin{pmatrix} \operatorname{vec} U_{m-1} \\ \vdots \\ \operatorname{vec} U_0 \\ \operatorname{vec} V_0 \\ \vdots \\ \operatorname{vec} V_{m-1} \end{pmatrix}.$$

The boundary matrices M and N satisfying $M + Ne^{\mathcal{A}h} = -\binom{\operatorname{vec} W}{0}$ are given by

$$M = \begin{pmatrix} & A_0^\top \otimes I & \dots & A_{m-1}^\top \otimes I & A_m^\top \otimes I \\ I_{n^2} & & & & \\ & \ddots & & & \\ & & I_{n^2} & & \\ & & & \ddots & \\ & & & & I_{n^2} \end{pmatrix}, N = \begin{pmatrix} I \otimes A_m^\top & I \otimes A_{m-1}^\top & \dots & I \otimes A_0^\top & \\ & -I_{n^2} & & & \\ & & \ddots & & \\ & & & -I_{n^2} & \\ & & & & \ddots \\ & & & & -I_{n^2} \end{pmatrix}.$$

The system matrix $\mathcal{A}$ in Problem 6.72 has the structure of a block Sylvester resultant matrix, see Lang [92] for a general discussion of resultants. Note how the symmetry condition $U_0(0) = V_0(h)$ nicely fits into the conditions which join the solution segments together. The existence and uniqueness issues for this boundary value problem are again attached to the regularity of $M + Ne^{\mathcal{A}h}$.

We study the properties of such block resultant matrices. The following proposition contains the basic facts.

Proposition 6.73. *Suppose that $p(s) = \sum_{i=0}^{r} A_i s^i$ and $q(s) = \sum_{j=0}^{r} B_j s^j$ are polynomial matrices of degree $r \geq 1$ with $A_i, B_j \in \mathbb{C}^{n\times n}$ and $A_i B_j = B_j A_i$ for all $i, j = 0, \dots, r$ where A_0 and B_r are regular. There exists a common root $s \in \mathbb{C}$ and a non-trivial vector $x \in \mathbb{C}^{n\times n}$ such that $p(s)x = 0 = q(s)x$ if and only if the determinant of*

$$\mathfrak{A} := \begin{pmatrix} A_0 & \dots & A_{r-1} & A_r & & \\ & \ddots & & & \ddots & \\ & & A_0 & A_1 & \dots & A_r \\ B_0 & \dots & B_{r-1} & B_r & & \\ & \ddots & & & \ddots & \\ & & B_0 & B_1 & \dots & B_r \end{pmatrix}_{2rn^2 \times 2rn^2}$$

is zero.

Proof. If $s \in \mathbb{C}$ is common root s of p and q and $x \neq 0$ is a suitable vector such that $p(s)x = 0 = q(s)x$, we construct the column vector $z := (x, sx, \dots s^{2r-1}x)$. Then

$$\mathfrak{A}z = \begin{pmatrix} A_0 & \dots & A_{r-1} & A_r & & \\ & \ddots & & & \ddots & \\ & & A_0 & A_1 & \dots & A_r \\ B_0 & \dots & B_{r-1} & B_r & & \\ & \ddots & & & \ddots & \\ & & B_0 & B_1 & \dots & B_r \end{pmatrix} \begin{pmatrix} x \\ \vdots \\ s^{r-1}x \\ s^r x \\ \vdots \\ s^{2r-1}x \end{pmatrix} = \begin{pmatrix} p(s)x \\ \vdots \\ s^{r-1}p(s)x \\ q(s)x \\ \vdots \\ s^{r-1}q(s)x \end{pmatrix} = 0.$$

Thus $\mathfrak{A}z = 0$ for $z \neq 0$ implies that $\det(\mathfrak{A}) = 0$. On the other hand, if $\det(\mathfrak{A}) = 0$ then there exists a non-trivial vector x partitioned into $x = (x_1, \dots, x_{2r})$ with $x_i \in \mathbb{C}^n$ such that

$\mathfrak{A}x = 0$. As A_0 and B_r are invertible, let us set $A_{0,i} = A_0^{-1}A_i$ and $B_{r,j} = B_r^{-1}B_j$. Then the structure of $\mathfrak{A}$ implies that for $i = 1, \ldots, r$

$$\begin{pmatrix} x_{i+1} \\ x_{i+2} \\ \vdots \\ x_{r+i} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & I & & \\ & & \ddots & \\ & & & I \\ -B_{r,0} & -B_{r,1} & \ldots & -B_{r,r-1} \end{pmatrix}}_{=:B^+} \begin{pmatrix} x_i \\ x_{i+1} \\ \vdots \\ x_{r+i-1} \end{pmatrix}, \quad \begin{pmatrix} x_i \\ x_{i+1} \\ \vdots \\ x_{r+i-1} \end{pmatrix} = \underbrace{\begin{pmatrix} -A_{0,1} & -A_{0,2} & \ldots & -A_{0,r} \\ 0 & I & & \\ & & \ddots & \\ & & & I \end{pmatrix}}_{=:A^-} \begin{pmatrix} x_{i+1} \\ x_{i+2} \\ \vdots \\ x_{r+i} \end{pmatrix}.$$

If we define $\tilde{x}_i = (x_i, \ldots x_{r+i-1}) \in \mathbb{C}^{rn}$ then this can be written compactly as $\tilde{x}_{i+1} = B^+\tilde{x}_i$ and $\tilde{x}_i = A^-\tilde{x}_{i+1}$ for all $i = 1, \ldots, r$. Now consider the product A^-B^+. The vector $\tilde{x}_i$ is an eigenvector of this product corresponding to the eigenvalue 1. Considering just the first block-row in this product gives us for $i = 1, \ldots, r$

$$-\sum_{k=1}^{r-1} A_{0,k}x_{i+k} + A_{0,r}\sum_{k=0}^{r-1} B_{r,k}x_{i+k} = x_i. \tag{6.86}$$

Since A_i and B_j commute, multiplying (6.86) with A_0 and B_r yields

$$\sum_{k=0}^{r-1}(A_rB_k - A_kB_r)x_{i+k} = 0.$$

Hence all $\tilde{x}_i$, $i = 1, \ldots, r$, are contained in the kernel of the matrix

$$Z := [A_rB_0 - A_0B_r, A_rB_1 - A_1B_r, \ldots, A_rB_{r-1} - A_{r-1}B_r].$$

Therefore $B^+, A^- : \ker Z \to \ker Z$. We have found a subspace which is invariant under B^+ and A^-. Thus there exists an eigenvector in $\ker Z$, say $\tilde{z}$, corresponding to an eigenvalue $\zeta \in \mathbb{C}$ of B^+. But if $\tilde{z}$ is an eigenvector of B^+ then it is also an eigenvector of A^- as $\tilde{z} = A^-B^+\tilde{z} = \zeta A^-\tilde{z}$. Due to the structure of B^+, this eigenvector is given by $\tilde{z} = (z, \zeta z, \ldots, \zeta^{r-1}z)$. Clearly, for powers of B^+, $\tilde{z}$ will also be an eigenvector. Therefore $\mathfrak{A}(z, \zeta z, \ldots, \zeta^{2r-1}z) = 0$, which implies $p(\zeta)z = 0 = q(\zeta)z$. □

As the system matrix $\mathcal{A}$ of Problem 6.72 satisfies the assumptions of Proposition 6.73, there are eigenvectors x_i of $\mathcal{A}$ which are given by vectorizations of tuples $(Z_i, \zeta_iZ_i, \ldots, \zeta_i^{2m-1}Z_i)$ where $\zeta_i \in \mathbb{C}$ and $Z_i \in \mathbb{C}^{n\times n}$, corresponding to an eigenvalue λ_i of $\mathcal{A}$. These eigenvectors satisfy

$$\sum_{k=0}^{m} \zeta_i^k Z_i A_k = \lambda_i Z_i \quad \text{and} \quad -\sum_{k=0}^{m} A_{m-k}^\top \zeta_i^k Z_i = \lambda_i \zeta_i^m Z_i. \tag{6.87}$$

Now assume that the boundary matrix $B = M + Ne^{\mathcal{A}h}$ is singular and that there exists $x = \sum_i \alpha_i x_i$, $x \neq 0$, such that

$$\begin{aligned} 0 = Bx = \sum_i \alpha_i(Mx_i + Ne^{\lambda_i h}x_i) = \sum_i \alpha_i\big(&\mathrm{vec}(\lambda_i\zeta_i^{m-1}Z_i, Z_i, \zeta_iZ_i, \ldots, \zeta_i^{2m-2}Z_i) \\ &-e^{\lambda_i h}\,\mathrm{vec}(-\lambda_i\zeta_i^m Z_i, \zeta_iZ_i, \zeta_i^2Z_i, \ldots, \zeta_i^{2m-1}Z_i)\big), \end{aligned} \tag{6.88}$$

where we used (6.87) to obtain

$$Mx_i = \operatorname{vec}(\sum_{k=0}^{m} \zeta_i^{m-1+k} Z_i A_k, Z_i, \zeta_i Z_i, \dots, \zeta_i^{2m-2} Z_i) = \operatorname{vec}(\lambda_i \zeta_i^{m-1} Z_i, Z_i, \dots, \zeta_i^{2m-2} Z_i),$$

$$Nx_i = \operatorname{vec}(\sum_{k=0}^{m} A_{m-k}^\top \zeta_i^k Z_i, -\zeta_i Z_i, \dots, -\zeta_i^{2m-1} Z_i) = \operatorname{vec}(\lambda_i \zeta_i^{m} Z_i, -\zeta_i Z_i, \dots, -\zeta_i^{2m-1} Z_i).$$

From (6.88) we have $\zeta_i = e^{-\lambda_i h}$ for all indices i with $\alpha_i \neq 0$. We still have to verify the first block-row of (6.88) which reads

$$0 = \sum_i \alpha_i \left(\zeta_i^{m-1} \sum_{k=0}^{m} \zeta_i^k Z_i A_k + \zeta_i^{-1} \sum_{k=0}^{m} A_{m-k}^\top \zeta_i^k Z_i \right).$$

But by (6.87) this term vanishes if

$$Z_i \left(\lambda_i I - \sum_{k=0}^{m} e^{-\lambda_i k h} A_k \right) = 0 \quad \text{and} \quad \zeta_i^m Z_i^\top \left(-\lambda_i - \sum_{k=0}^{m} e^{-\lambda_i (m-k) h} A_{m-k} \right) = 0.$$

Therefore $\det(\pm \lambda_i I - \sum_{k=0}^{m} e^{-\lambda_i k h} A_k) = 0$ for all i with $\alpha_i \neq 0$. Hence we have no uniquely determined delay Liapunov matrix if for a suitable index i both λ_i and $-\lambda_i$ are eigenvalues of the delay equation (6.1) for commensurable delays.

Example 6.74. We now calculate the delay Liapunov matrix $U(t)$ on $t \in [0,2]$ which is associated with the equation

$$\dot{x}(t) = A_0 x(t) + A_1 x(t-1) + A_2 x(t-2), \text{ where } A_0 = -\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right), A_1 = \left(\begin{smallmatrix} 0 & 0.7 \\ 0.7 & 0 \end{smallmatrix}\right), A_2 = -\left(\begin{smallmatrix} 0.49 & 0 \\ 0 & 0.49 \end{smallmatrix}\right)$$

with $h = 1$. Consider the vectorization of the tuple (U_1, U_0, V_0, V_1) as state. Then the system matrix $\mathcal{A}$ and the boundary matrices M, N are given as in Problem 6.70. Now given a weight W we find an initial value by computing $x_0 = (M + N e^{\mathcal{A} h})^{-1} \left(\begin{smallmatrix} -\operatorname{vec} W \\ 0 \end{smallmatrix}\right)$. Due to the symmetry in the solution segments we need to solve the associated initial value problem only on $[0, h/2]$. Rearranging the solution segments via

$$U(t) = \begin{cases} U_k(t - kh) & \text{if } t \in [kh, (k + \frac{1}{2})h], \\ V_k(h - (t - kh))^\top & \text{if } t \in [(k + \frac{1}{2})h, (k+1)h], \end{cases}$$

we obtain the solution depicted in Figure 6.6 with weight $W = 6I$. ■

Remark 6.75. We have not addressed the problem of delay margins for stability of delay systems with commensurable delays. The main problem with the presented approach is that it only considers the “unit” delay h. Hence a stability analysis starting from a solution of Problem 6.70 would have to ensure that the commensurability of the delays stays intact, regardless of perturbations in the unit delay h. Under this premise, the results for critical delays can also be used in the commensurable case.

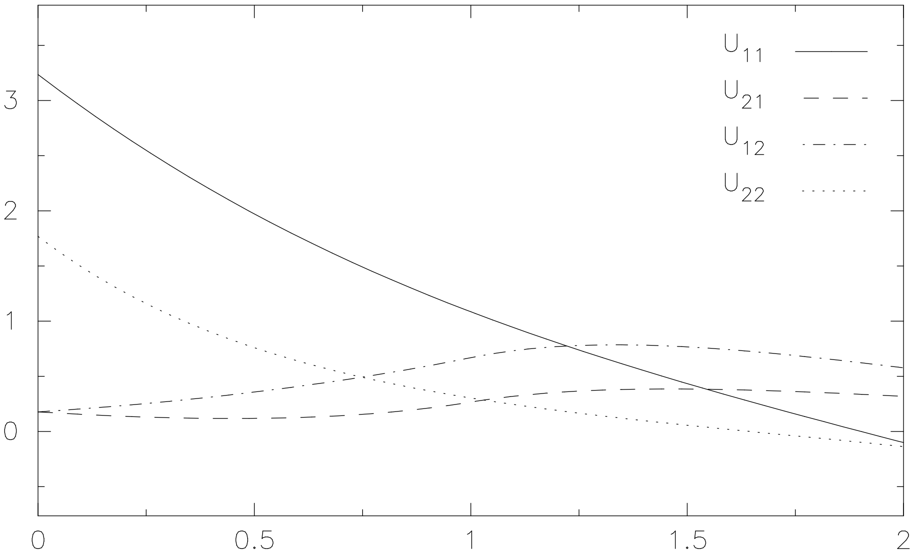

Figure 6.6: Components of the multi delay Liapunov matrix $U(t), t \in [0, 2]$.

6.7 Scalar Differential Delay Equations

In this section we study the scalar delay equation with one delay

$$\dot{x}(t) = a_0 x(t) - a_1 x(t-h) \quad \text{with} \quad a_0 < 0, a_1 \neq 0, \tag{6.89}$$

and the scalar equation with commensurable time lags

$$\dot{x}(t) = a_0 x(t) + \sum_{k=0}^{m} a_k x(t - kh), \tag{6.90}$$

for which we apply the obtained results.
Let us first consider the one-delay case (6.89). The continuous dependency of the spectrum of equation (6.89) on the delay is seen easily in this case. In the real scalar case the partial derivatives of $f(h, \lambda) = \lambda - a_0 - a_1 e^{-\lambda h}$ are given by

$$\tfrac{\partial f}{\partial \lambda} = 1 + a_1 h e^{-\lambda h}, \qquad \tfrac{\partial f}{\partial h} = a_1 \lambda e^{-\lambda h}.$$

We only consider solutions (h, λ) with $f(h, \lambda) = 0$, i.e., $\lambda - a_0 = a_1 e^{-\lambda h}$, hence we obtain

$$\tfrac{\partial f}{\partial \lambda} = 1 + h(\lambda - a_0), \qquad \tfrac{\partial f}{\partial h} = \lambda(\lambda - a_0).$$

In particular, the only critical point with $f(h, \lambda) = 0$ and $\frac{\partial f}{\partial \lambda}(h, \lambda) = 0$ is located at $\lambda = a_0 - h^{-1} \in \mathbb{R}$. This can only occur if $a_1 < 0$, as additionally $e^{a_0 h - 1} = -a_1 h$ has to hold. A monotonicity argument shows that in this case there exists only one real solution h_0 of $e^{a_0 h - 1} = -a_1 h$. Hence the implicit function $\lambda(h)$ with $f(h, \lambda(h)) = 0$ is continuous on $h \in (0, h_0)$ or $h \in (h_0, \infty)$ depending on whether $h < h_0$ or $h > h_0$ as there are no other

critical values for which the partial derivative vanishes. Moreover, if $a_1 > 0$ then there exist no critical values, and $\lambda(h)$ is continuous for $h > 0$.
Let us now set up the data according to Problem 6.45 in order to solve for a delay Liapunov matrix. We obtain

$$\mathcal{A} = \begin{pmatrix} a_0 & a_1 \\ -a_1 & -a_0 \end{pmatrix}, \quad M = \begin{pmatrix} a_0 & a_1 \\ 1 & 0 \end{pmatrix}, \quad N = \begin{pmatrix} a_1 & a_0 \\ 0 & -1 \end{pmatrix}.$$

By Corollary 6.60, equation (6.89) is stable independent of delay if the spectrum of $\mathcal{A}$ contains no imaginary eigenvalues. Now $\det(sI - \mathcal{A}) = s^2 - a_0^2 + a_1^2$ has only real roots for $a_0^2 \geq a_1^2$ and only purely imaginary roots for $a_0^2 > a_1^2$. Therefore we have stability independent of delay if (6.89) is exponentially stable, $a_0 + a_1 < 0$, and $|a_0| < |a_1|$.
Let us now consider the case $|a_1| > |a_0|$. We determine the delay margin for (6.89). For this analysis we need the eigenvectors of $\mathcal{A}$ which are given by $\left(\begin{smallmatrix} a_1 \\ -a_0 \pm i\sqrt{a_1^2 - a_0^2} \end{smallmatrix}\right)$ corresponding to the eigenvalues $\pm i\sqrt{a_1^2 - a_0^2}$. Hence the scaling factor associated with $\lambda = i\sqrt{a_1^2 - a_0^2}$ is $\zeta = e^{-\lambda h} = \frac{\lambda - a_0}{a_1}$ which is of modulus 1. We obtain the delay margin, the first critical value for h, as the smallest positive solution of $\zeta = \frac{\lambda - a_0}{a_1} = e^{-\lambda h}$.

Example 6.76. Let us take a look at the hot-shower Example 6.1. There we had $a_0 = 0$ and $a_1 < 0$. From the spectral analysis of $\mathcal{A}$ we obtain $\lambda = i\,|a_1|$ and therefore $\zeta = -i = e^{ia_1 h}$. The smallest positive solution of this equation is given by $h^* = -\frac{\pi}{2a_1} > 0$, which is the delay margin.
With the data from Example 6.1, namely $a_1 = -1$, the delay margin $h^* = \frac{\pi}{2}$ satisfies $1 < h^* < 2$. Hence our assertions about the exponential stability of (6.2) for the delay $h = 1$ and about the instability for the delay $h = 2$ are correct. ■

For the multi-delay case (6.90), we set up the data according to Problem 6.72. The system matrix $\mathcal{A}$ is given by

$$\mathcal{A} = \begin{pmatrix} a_0 & \dots & a_{m-1} & a_m & & \\ & \ddots & & & \ddots & \\ & & a_0 & a_1 & \dots & a_m \\ -a_m & \dots & -a_1 & -a_0 & & \\ & \ddots & & & \ddots & \\ & & -a_m & -a_{m-1} & \dots & -a_0 \end{pmatrix} \in \mathbb{R}^{2m \times 2m}.$$

This matrix is a resultant matrix. Hence $\det(\lambda I - \mathcal{A}) = 0$ if and only if the polynomials $p_\lambda(s) := (a_0 - \lambda) + \sum_{k=1}^{m} a_k s^k$ and $q_\lambda(s) := \sum_{k=0}^{m-1} a_{m-k} s^k + (a_0 + \lambda)s^m$ contain a common root, see Gantmacher [45]. Now $q_\lambda(s) = s^m p_{-\lambda}(s^{-1})$ which again demonstrates the intrinsic symmetry of the problem. Moreover, the common root $\zeta \in \mathbb{C}$ coincides with the scaling factor from which an eigenvector $(1, \zeta, \dots, \zeta^{2m-1})^\top$ of $\mathcal{A}$ is constructed. For a stability analysis, we have to consider purely imaginary eigenvalues $\lambda = i\omega$ of $\mathcal{A}$ and solve for h in

$\zeta = e^{-\lambda h}$ to get candidates for the delay margin. If we set

$$A_0 = \begin{pmatrix} a_0 & \dots & a_{m-1} \\ & \ddots & \vdots \\ & & a_0 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} a_m & & \\ \vdots & \ddots & \\ a_1 & \dots & a_m \end{pmatrix},$$

then we can rewrite (6.90) as the one-delay matrix equation $\dot{x}(t) = A_0 x(t) + A_1 x(t - mh)$. However, it has not been investigated in which way the delay Liapunov matrices of both formulations are related.

6.8 Notes and References

This chapter has grown out of the article [80] with V. Kharitonov, see also the references therein. In contrast to this article, where Liapunov-Krasovskii functionals over C have been considered, we embed them here into an M^2-framework. Functional analytic approaches to delay-differential systems can be found in many books, e.g. Hale and Verduyn Lunel [51] and Diekmann et al. [33]. Our discussion of the M^2-inner product follows Curtain and Zwart [29]. For application of this calculus to neutral-type delay systems, see Salamon [122], for an application to partial differential equations with delay, see Bátkai and Piazzero [10]. The book [110] by Niculescu is a valuable resource for results on delay equations. M^p-spaces for delay equations are discussed in Bensoussan et al. [16].
To embed Liapunov-Krasovskii functionals v which satisfy

$$-\dot{v}(\varphi) = \varphi(0)^\top W_0 \varphi(0) + \sum_{k=1}^{m} \varphi(-h_k)^\top W_k \varphi(-h_k) + \sum_{k=1}^{m} \int_{-h_k}^{m} \varphi(\theta)^\top W_{m+k} \varphi(\theta)\, d\theta,$$

like the ones discussed in [81, 79] into an M^2-framework one needs an augmented M^2-space which also respects all delayed states, $(\varphi(0), \varphi(-h_1), \dots, \varphi(-h_m), \varphi) \in (\mathbb{R}^n)^{m+1} \times L^2([-H, 0], \mathbb{R}^n)$. A different approach of prescribing the terms of the functional $v(\hat{\varphi})$ is chosen in most papers dealing with the construction of Liapunov-Krasovskii functionals [97, 19]. Such an approach does not automatically lead to positive definite functionals $\dot{v}(\hat{\varphi})$, hence the functional v might not provide exponential estimates for the solutions of the delay system.
Proposition 6.28 can be found in Datko [32], a systematic study can be found in Kharitonov and Zhabko [81]. Smith [127] has shown that a solution of the classical delay-free quadratic Liapunov function $-Q = PA + A^* P$ is given by

$$P = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(i\omega)^* Q G(i\omega) d\omega, \qquad \text{where } G(s) = (sI - A)^{-1}.$$

The counterpart for Liapunov-Krasovskii functionals associated with systems with one delay is obtained in Louisell [96], which we extended to the multi-delay case in in Proposition 6.29.

The Infante-Castellan approach of solving the one-delay Liapunov matrix goes back to the articles [73, 25], see also [32]. Unfortunately, Infante and Castellan state in their article [73] that the existence and uniqueness issue has been solved in their earlier paper [25]. However, those results are not applicable.

The computation of delay margins is discussed in Hertz et al. [55]. A method working with matrix pencils can be found in Chen et al. [26], see also the overview in Niculescu [110]. For a method working with LMIs, see Bliman [19].

Computational issues for the numerical solution of differential delay equations can be found in Bellen and Zennaro [13]. Numerical methods to obtain the spectrum of a linear differential-delay system are discussed in Breda et al. [23].

Chapter 7

(M, β)-Stabilization

In this chapter we study the synthesis of state feedback matrices for linear dynamical systems such that transient effects are taken into account. Let us first extend Definition 3.1 to dynamical systems with inputs. We will only consider the case of real data.

Definition 7.1. A linear time-invariant system of the form

$$\dot{x}(t) = Ax(t) + Bu(t), \quad t \geq 0, \qquad A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, \tag{7.1}$$

is said to be (strictly/uniformly)(M, β)-*stabilizable* by state feedback, if there exists a matrix $F \in \mathbb{R}^{n\times m}$ such that the closed loop system $\dot{x}(t) = (A-BF)x(t)$ is (strictly/uniformly) (M, β)-stable. For the special case $M = 1$, $\beta = 0$ we will call the pair (A, B) *contractible.*

Our main tool for the investigation will again be the initial growth rate with respect to some norm as for $\mu(A - BF) < 0$ the closed loop system generates a uniform contraction semigroup with respect to this norm. We opted for the synthesis of uniform contraction semigroups, as the inital growth rate does not provide us with methods to differentiate between strict and weak contractions. In the following we identify those systems which allow for a uniform closed loop contraction. To allow additional freedom for the transient bound M, we consider general norms $\|\cdot\|$. Later on, we fix the Euclidean norm and study quadratic (M, β)-stabilizability.

7.1 Synthesis of Contractions

If the system (7.1) admits a feedback matrix $F \in \mathbb{R}^{m\times n}$ such that the closed loop system matrix $A_F = A - BF$ generates a uniform contraction then this feedback has to satisfy $\mu(A_F) < 0$. Let us first investigate the vector case $(m = 1)$.

$$\dot{x}(t) = Ax(t) + bu(t), \qquad A \in \mathbb{R}^{n\times n}, \, b \in \mathbb{R}^{n}. \tag{7.2}$$

We assume that a norm of interest $\|\cdot\|$ is given and we denote its dual norm by $\|\cdot\|^*$, see (2.16). For a given b we set

$$\begin{aligned} V^+ &:= \left\{x \in \mathbb{R}^n \middle| \text{ for all dual vectors } y \text{ of } x,\ y^\top b > 0\right\}, \\ V^- &:= \left\{x \in \mathbb{R}^n \middle| \text{ for all dual vectors } y \text{ of } x,\ y^\top b < 0\right\}, \\ V^0 &:= \left\{x \in \mathbb{R}^n \middle| \text{ there exists a dual vector } y \text{ of } x \text{ with } y^\top b = 0\right\}. \end{aligned} \tag{7.3}$$

For our first result the following assumptions are needed.

(A1) V^0 contains a real hyperplane H^0 defined by a suitable vector $h \neq 0$ through

$$H^0 := \left\{x \in \mathbb{R}^n \middle| \ h^\top x = 0\right\}.$$

(A2) For all $x \in H^0$, $\|x\| = 1$, there is a uniquely determined vector y_* such that (x, y_*) is a normed dual pair.

The hyperplane H^0 separates the sets V^+ and $V^- = -V^+$. The assumption (A2) is satisfied for quadratic norms but not necessarily for arbitrary norms. In general, given a specific norm only a few vectors will have this property as the following Lemma shows.

Lemma 7.2. *If (A2) holds, then for every dual pair (x, y) of $\|\cdot\|$ with $x \in H^0$ and $y^\top b = 0$ the normed vector $y/\|y\|^*$ is an extremal point of $\|\cdot\|^*$.*

Proof. Assume that (x, y) is a unitary dual pair of $\|\cdot\|$ with $x \in H^0$ and $y^\top b = 0$, $\|y\|^* = 1$ such that y is not an extremal point of $\|\cdot\|^*$. By Definition 4.15, the unit sphere of $\|\cdot\|^*$ then contains a face given by $\operatorname{conv}(y_1, \ldots, y_k) \ni y$ with $\|y_i\|^* = 1, i = 1, \ldots, k$. But if (x', y) is a dual pair of $\|\cdot\|$ then (x', y') is a dual pair for all $y' \in \operatorname{conv}(y_1, \ldots, y_k)$, see Proposition 2.27 *(iii)*. Hence the dual vector of x' is not uniquely determined. By definition $x' \in V^0$, and for $x = x'$ we see that there are vectors in H^0 that have no uniquely determined dual vector. □

Corollary 7.3. *If the unit sphere $\mathbb{S}$ of $\|\cdot\|$ is smooth (see p. 9) then (A2) holds.*

The role of (A1) is investigated in the following lemma.

Lemma 7.4. *Given a vector $b \in \mathbb{R}^n$ and a vector norm $\|\cdot\|$ in $\mathbb{R}^n$. The following two statements are equivalent for $h \in \mathbb{R}^n$, $h \neq 0$.*

(i) For all dual pairs (x, y) of $\|\cdot\|$,

$$y^\top b h^\top x \geq 0 \qquad \text{and} \qquad y^\top b h^\top x > 0, \quad \forall x \notin \ker h^\top,\ y \notin \ker b^\top. \tag{7.4}$$

(ii) The set V^0 in (7.3) contains a hyperplane $H^0 \subset V^0$ given by $H^0 = \{x \in \mathbb{R}^n \mid h^\top x = 0\}$.

Proof. $(i) \implies (ii)$. For a given $h \in \mathbb{R}^n$ we take $x \in H^0 = \ker h^\top$. Then $h^\top x = 0$. By (7.4), $y \in \ker b^\top$ holds for all dual vectors y of x. Hence $H^0 \subset V^0$.
$(ii) \implies (i)$. We define the real halfspaces

$$H^+ := \left\{x \in \mathbb{R}^n \middle|\, h^\top x > 0\right\}, \qquad H^- := \left\{x \in \mathbb{R}^n \middle|\, h^\top x < 0\right\}. \tag{7.5}$$

Now, $h^\top b \neq 0$. Namely, suppose that $h^\top b = 0$. Then $b \in H^0 \subset V^0$, whence there exists a dual pair (b, a) with $a^\top b = 0$. But this is a contradiction as dual pairs (b, a) always satisfy $a^\top b > 0$. We therefore can assume without loss of generality that $h^\top b > 0$. First note that if (x, y) is a normed dual pair then $y^\top b$ is a subgradient of the convex function $g : t \mapsto \|x + tb\|$ at $t = 0$, since

$$g(t) = \|x + tb\| = y_t^\top(x + tb) \geq y^\top(x + tb) = \|x\| + ty^\top b \tag{7.6}$$

holds for dual pairs $(x + tb, y_t)$, $\|y_t\|^* = 1$, see Proposition 2.27 *(iv)*. If $x \in V^+$, i.e., $y^\top b > 0$, this implies that the function g is strictly increasing in $t = 0$ and therefore for all $t \geq 0$. If $x \in V^+ \cap H^-$ then as $h^\top b > 0$ we have that $x + t_1 b \in H^0$ for some $t_1 > 0$. By assumption there exists a normed dual vector y_1 of $x + t_1 b$ which satisfies $b^\top y_1 = 0$. This implies that g has a minimum in t_1, as

$$\tfrac{d}{dt} g(t)\big|_{t=t_1} = \tfrac{d}{dt}\, \|x + tb\|\, \big|_{t=t_1} = y_1^\top b = 0.$$

This is in contradiction to the convexity of g, as g is monotonously increasing on $t > 0$ and has a minimum in $t_1 > 0$. Hence the intersection $V^+ \cap H^-$ and analogously $V^- \cap H^+$ are empty. As $V^+ \subset H^+$, $V^- \subset H^-$ we have that $y^\top b h^\top x \geq 0$ for all dual pairs (x, y) and it is easy to see that $y^\top b h^\top x > 0$ for all dual pairs (x, y) such that $y \notin \ker b^\top$, $x \notin \ker h^\top$. □

Hence (A1) guarantees the existence of a "quasi-semidefinite" matrix $bh^\top$ for the norm $\|\cdot\|$. With the addition of Assumption (A2) we can conclude the following.

Theorem 7.5. *Consider system* (7.2) *and the norm* $\|\cdot\|$, *and assume that (A1) and (A2) hold. Then the pair* (A, b) *is uniformly contractible with respect to* $\|\cdot\|$ *if and only if*

$$y^\top A x < 0 \qquad \textit{for all}\, y \in \ker b^\top \;\textit{and}\; x \in \mathbb{R}^n \;\textit{such that}\, (x, y) \;\textit{is a dual pair.} \tag{7.7}$$

Proof. If (A, b) is uniformly contractible then there exists a vector $f \in \mathbb{R}^{1\times n}$ such that $A - bf$ generates a uniform contraction semigroup, or equivalently, for all dual pairs (x, y) of $\|\cdot\|$ the strict inequality $y^\top(A - bf)x < 0$ holds. For dual vectors $y \in \ker b^\top$ of x this implies that $y^\top(A - bf)x = y^\top Ax < 0$, hence proving necessity of (7.7). We now show the existence of a suitable feedback if (7.7) holds under the assumptions (A1) and (A2). By Lemma 7.4 there exists a hyperplane H^0 induced by the vector h which separates V^+ and V^- of (7.3). We now claim that for α sufficiently large we have

$$y^\top(A - \alpha b h^\top)x < 0$$

for all dual pairs (x, y). Note that it is sufficient to prove this on the compact set

$$Z := \left\{(x, y) \,\middle|\, \|y\|^* = \|x\| = y^\top x = 1\right\} = \left\{(x, y) \text{ unitary dual pair}\right\}.$$

By continuity, the set $Z_- \subset Z$ of points satisfying $y^\top Ax < 0$ is open in Z. Assumption (A2) now implies that Z_- contains a set of the form

$$Z_\varepsilon := \left\{ (x, y) \in Z_- \mid -\varepsilon < h^\top x < \varepsilon \right\}$$

for $\varepsilon > 0$ sufficiently small. Now, if $(x, y) \in Z \setminus Z_\varepsilon$ then $\left| h^\top x \right| \geq \varepsilon$. Furthermore, there exists a $\delta > 0$ so that $y^\top Ax \geq 0$ implies $\left| y^\top b \right| > \delta$, otherwise we obtain a contradiction to (7.7). We have by Lemma 7.4 that $y^\top bh^\top x > 0$ on $Z \setminus Z_\varepsilon$. If additionally $y^\top Ax \geq 0$ holds then $y^\top bh^\top x > \delta\varepsilon > 0$. Hence setting

$$\alpha := \frac{2}{\delta\varepsilon} \max_{(x,y) \in Z} \left| y^\top Ax \right| > 0, \tag{7.8}$$

we easily see that $y^\top Ax - \alpha y^\top bh^\top x < 0$ for all $(x, y) \in Z$. □

Remark 7.6. Note that the term $\max_{(x,y)\in Z} \left| y^\top Ax \right|$ can be replaced by $\max\{\mu(A), 0\}$ as only the positive terms $y^\top Ax$ have to be bounded. Moreover, note that the construction in the previous proof relies on a high gain type argument. We have constructed h such that always $y^\top bh^\top x \geq 0$. This implies that if $A - \alpha_0 bh^\top$ generates a uniform contraction semigroup then the same is true for $A - \alpha bh^\top$ for *all* $\alpha \geq \alpha_0$. Such a high gain idea is not feasible in all situations. For example, consider a system (7.2) given by

$$A = \begin{pmatrix} -1 & c \\ -2 & d \end{pmatrix}, \qquad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and assume that we want to generate a uniform contraction with respect to the 1-norm $\|\cdot\|_1$. The kernel is given by $\ker b^\top = \mathbb{R}\binom{1}{0}$ and the (unique) vector x such that $(x, \binom{1}{0})$ is a dual pair is given by $x = e_1$. An easy calculation shows $\binom{1}{0}^\top A \binom{1}{0} = -1$ so that condition (7.7) is satisfied. Note that H^0 from (A1) has to be $H^0 = \mathbb{R}e_1$, but (A2) is not satisfied for H^0. Also Ae_1 is not pointing inside the unit ball of $\|\cdot\|_1$ which can be seen by calculating $\binom{1}{-1}^\top A \binom{1}{0} = 1$ and noting that $\binom{1}{-1}$ is dual to e_1. If we now consider possible feedback matrices $(f_1\, f_2)$ then we see that

$$A - bf = \begin{pmatrix} -1 & c \\ -2 - f_1 & d - f_2 \end{pmatrix}.$$

Hence for $f_1 \in (-3, -1)$ the matrix $A - bf$ is diagonally dominant in the first column. Similarly, $f_2 > \max\{d + c, d - c\}$ ensures that $A - bf$ is pointing inward at e_2. Hence $A - bf$ generates a contraction semigroup with respect to $\|\cdot\|_1$ if and only if

$$f \in \left\{ [f_1, f_2] \in \mathbb{R}^{1\times 2} \middle| \, f_1 \in (-3, -1), f_2 > \max\{d + c, d - c\} \right\}.$$

In particular, for any choice of f that leads to a uniform contraction semigroup there is an α_0 such that for *all* $\alpha \geq \alpha_0$, $A - \alpha bf$ is *not* dissipative. Hence there is no high-gain feedback as in Theorem 7.5.

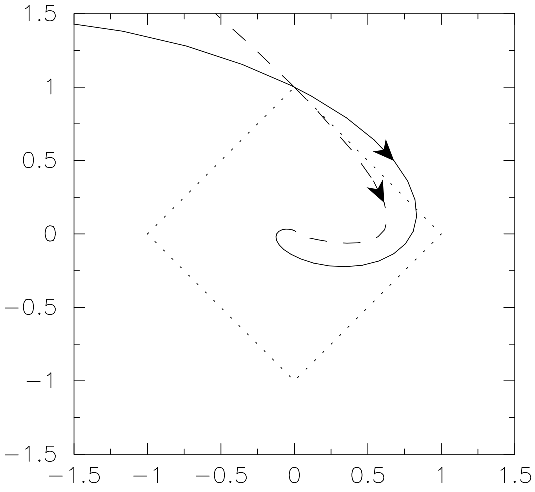

Figure 7.1: A closed loop contraction with respect to $\|\cdot\|_1$.

Example 7.7. Choosing $c = 6, d = -3$ in the previous remark gives $A = \left(\begin{smallmatrix} -1 & 6 \\ -2 & -3 \end{smallmatrix}\right)$. The allowed feedback matrices $[f_1, f_2]$ can be selected from $f_1 \in (-3, -1)$ and $f_2 > 3$. Figure 7.1 shows a trajectory of $\dot{x} = Ax$ which leaves the (dotted) unit box of $\|\cdot\|_1$ and a trajectory of the closed loop system with $f = [-1, 3]$. Here $A - bf = \left(\begin{smallmatrix} -1 & 6 \\ -1 & -6 \end{smallmatrix}\right)$ is only marginally diagonally dominant, and the closed loop system generates a strict contraction but not a uniform contraction. For $f = [-3, 3]$ the closed loop becomes only marginally stable as $A - bf = \left(\begin{smallmatrix} -1 & 6 \\ 1 & -6 \end{smallmatrix}\right)$ is singular. ■

To treat the case of higher dimensional input spaces the following result can be easily obtained from Theorem 7.5. Again the assumption (A1) and (A2) are crucial. To apply the same arguments as before we have to assume that for each of the columns of B the assumptions $(A1), (A2)$ are satisfied individually. Note, however, that using a state transformation R on the input space, this property might be obtained for the matrix BR, while it is false for B.

Theorem 7.8. *Consider system* (7.1) *with* $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}$ *and the norm* $\|\cdot\|$. *Assume that for each column* b_j *of* B, $j = 1, \dots, m$ *the properties (A1) and (A2) are satisfied. Then the pair* (A, B) *is uniformly contractible if and only if*

$$y^\top Ax < 0 \qquad \textit{for all}\, y \in \ker B^\top \textit{ and}\, x \in \mathbb{R}^n \textit{ such that}\, (x, y) \textit{ is a dual pair.} \tag{7.9}$$

Proof. The necessity of (7.9) is obvious. For sufficiency, consider the matrices $H = [h_1, \dots, h_m]$ and $\Delta = \operatorname{diag}(\alpha_1, \dots, \alpha_m)$ obtained from Theorem 7.5 for each column b_j of B, $j = 1, \dots, m$. Then for $F = \frac{1}{m}\Delta H^\top$

$$y^\top (A - BF)x = y^\top \left(A - \sum_{j=1}^{m} \tfrac{\alpha_j}{m} b_j h_j^\top \right) x = \sum_{j=1}^{m} \tfrac{1}{m} y^\top \left(A - \alpha_j b_j h_j^\top \right) x < 0.$$

Hence the pair (A, B) is uniformly contractible. □

By replacing A with $A-\beta I$ in Theorem 7.8 we obtain the following result for arbitrary decay rates.

Corollary 7.9. *Under the assumptions of Theorem 7.8 there exists a feedback matrix $F \in \mathbb{R}^{m\times n}$ for a given decay rate $\beta < 0$ such that $\mu(A-BF) < \beta$ if and only if all dual pairs (x,y) with $y \in \ker B^\top$ satisfy $y^\top Ax < \beta\, y^\top x$.*

Let us return to the system (7.2). We now discuss stabilization results which can be obtained without postulating (A2). As we have already seen, to guarantee the existence of a feedback f with $\mu(A-bf) < 0$ we have to assume that

$$\text{for all dual pairs } (x,y) \text{ of } \|\cdot\| \text{ with } y \in \ker b^\top : y^\top Ax < 0. \tag{7.10}$$

Let us now define for every vector $x \in \mathbb{R}^n$ the following set of feasible feedback vectors,

$$\mathfrak{F}_x = \left\{ f \in \mathbb{R}^{1\times n} \;\middle|\; \text{for all dual vectors } y \in \mathbb{R}^n \text{ of } x,\; y^\top(A-bf)x < 0 \right\}.$$

Depending on the sign of $y^\top b$ this definition can be reformulated, namely, if $x \in V^+$ we have

$$\mathfrak{F}_x = \left\{ f \in \mathbb{R}^{1\times n} \;\middle|\; \tfrac{y^\top Ax}{y^\top b} < fx \quad \text{for all dual pairs } (x,y) \right\},$$

if $x \in V^-$ we set

$$\mathfrak{F}_x = \left\{ f \in \mathbb{R}^{1\times n} \;\middle|\; fx < \tfrac{y^\top Ax}{y^\top b} \quad \text{for all dual pairs } (x,y) \right\},$$

and if $x \in V^0$ both of the above conditions have to be considered, i.e.

$$\begin{aligned} \mathfrak{F}_x = \Big\{ f \in \mathbb{R}^{1\times n} \;\Big|\; & \tfrac{y^\top Ax}{y^\top b} < fx \quad \text{for dual pairs } (x,y) \text{ with } y^\top b > 0, \text{and} \\ & fx < \tfrac{y^\top Ax}{y^\top b} \quad \text{for dual pairs } (x,y) \text{ with } y^\top b < 0 \Big\}. \end{aligned} \tag{7.11}$$

Now suppose that $x \in V^0$ and that there are dual vectors y_1, y_2 of x satisfying $y_1^\top b > 0$ and $y_2^\top b < 0$. Then there exists $\lambda \in (0,1)$ such that (x, y_3) is a dual pair with $y_3 = \lambda y_1 + (1-\lambda) y_2$ and $y_3^\top b = 0$. By (7.10), $y_3^\top Ax < 0$, i.e., $\lambda y_1^\top Ax < (\lambda-1) y_2^\top Ax$. Now, dividing by $\lambda y_1^\top b = (\lambda-1) y_2^\top b > 0$ gives $\frac{y_1^\top Ax}{y_1^\top b} < \frac{y_2^\top Ax}{y_2^\top b}$, hence for the case $x \in V^0$ there is always a feasible set of feedback matrices.

By construction, a feedback matrix $f \in \mathbb{R}^{1\times n}$ which is taken from the set $\mathfrak{F} = \bigcap_{x\in\mathbb{B}} \mathfrak{F}_x$ gives a closed-loop matrix $A - bf$ which generates a contraction. But the set $\mathfrak{F}$ could be empty. We study this in more detail when the norm of interest is a polytopic norm.

7.2 Contractibility for Polytopic Norms

We have seen in Lemma 4.16, that dissipativity with respect to a polytopic norm needs only to be checked for a finite set of extremal points. The test for contractibility therefore generates a finite set of linear inequalities. However, Lemma 7.2 shows that in general (A2) does not hold for polytopic norms, since points inside the faces of its unit sphere are not extremal.

If $\|\cdot\|_C$ is a polytopic norm given by its set of vertices (extremal points), $C \subset \mathbb{R}^n$, then the dual norm is also a polytopic norm, given by the set $C^* \subset \mathbb{R}^n$ which corresponds to the normals of the faces of $\mathbb{S}_C := \{x \in \mathbb{R}^n \mid \|x\|_C = 1\} = \operatorname{conv}(C)$. We want to determine conditions such that there exists a feedback matrix $F \in \mathbb{R}^{m\times n}$ with $\mu_C(A - BF) < 0$. By Lemma 4.16 this is equivalent to $y^\top(A+BF)x < 0$ for all $x \in C$ and $y \in C^*$ with $y^\top x = 1$. However, for $y \notin \ker B^\top$, we have to find a feedback matrix F such that $y^\top BFx < y^\top Ax$. As both C and C^* are finite sets, this condition generates a set of finitely many inequalities for F. But it is not clear if this set of inequalities is feasible.

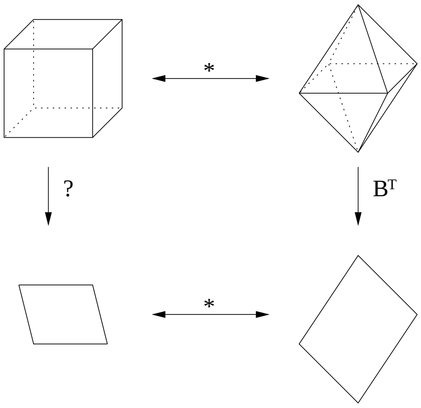

Figure 7.2: Polytopic Norms.

Figure 7.2 demonstrates this situation. The set C is given as the vertices of the left cube, the set of dual extremal points is given by the vertices of the octahedron on the right. $B^\top$ now projects the dual vectors $y \in C^*$ down into some subspace, where they again form some polytopic norm (not all vertices will stay extremal). We now have to find a map F on the C- side such that $(\operatorname{conv} FC)^* = \operatorname{conv}(B^*C^*)$. Then $y^*BFx > 0$ for all $y \in C^*$ and $x \in C$. This general problem is as yet unsolved. For a result involving linear transformations of convex sets, which leads the way to a possible solution, see [120, Corollary 16.3.1].

Let us return to the case $m = 1$, hence $b \in \mathbb{R}^n$ is a column vector in (7.2). We now discuss the set

$$\mathcal{F} = \bigcap_{x \in \mathbb{S}_C} \left\{ f \in \mathbb{R}^{1\times n} \,\middle|\, y \text{ is a dual vector of } x \text{ with } y^\top(A - bf)x < 0 \right\}.$$

By Lemma 4.16 only the extremal points have to be checked for dissipativity. Thus we have $\mathcal{F} = \bigcap_{x\in C} \mathcal{F}_x$ and the sets $\mathcal{F}_x$ are given by finitely may inequalities of the form $y^\top(A - bf)x < 0$ for suitable $y \in C^*$. For further analysis, let us define the set

$$W^+ = \left\{ (x, y) \in C \times C^* \,\middle|\, (x, y) \text{ is a dual pair and } y^\top b > 0 \right\}.$$

Hence if $x \in V^+ \cap C$ there exists a vector $y \in C^*$ such that $(x, y) \in W^+$. Moreover, for dimension $n \geq 2$, every vertex $x \in C$ has more than one adjacent face such that for $x \in V^0 \cap C$ there exists a $y \in C^*$ with $(x, y) \in W^+$ or $(-x, y) \in W^+$ depending on the sign of $y^\top b$. Hence W^+ is located in a half-hyperspace. To this end, if both (x, y) and $(-x, -y)$ were elements of W^+ then $y^\top b > 0$ and $-y^\top b < 0$ which is a contradiction.

Lemma 7.10. *The x-components of W^+ are separable from V^- by a suitable hyperplane.*

Proof. Let $x \in (V^- \cup V^0) \cap C$. Then all dual vectors y of x satisfy $y^\top b \leq 0$. For these dual pair (x, y), $y^\top x = 1$ holds and for all $z \in V^+ \cap C$ we have $y^\top z < 1$. Hence $x \notin \operatorname{conv}(V^+ \cap C)$. Therefore the convex cone generated by $V^+ \cap C$ is separable from the convex cone generated by $V^- \cap C$ using a suitable hyperplane. □

Let us assume that (7.10) holds. As there are only finitely many points in W^+ the set of feasible feedback vectors satisfies

$$\mathcal{F} = \bigcap_{(x,y) \in W^+} \left\{ f \in \mathbb{R}^{1 \times n} \,\middle|\, fx > \tfrac{y^\top A x}{y^\top b} \right\}.$$

This set is non-empty as the points in W^+ are located in a half hyper-space and dual pairs (x_0, y_0) with $x_0 \in V^0 \cap C$, $y_0 \in \ker b^\top$ have no effect on the set $\mathcal{F}$. To this end, consider a dual pair $(x_0, y) \in W^+$ and a continuous path $y : [0, 1] \to \mathbb{R}^n$ deforming $y(0) = y$ into $y(1) = y_0$ such that $y(t)^\top x = 1$ for all $t \in [0, 1]$. Then the quotient $\frac{y(t)^\top A x}{y(t)^\top b}$ approaches $-\infty$ for $t \to 1$ since $y(t)^\top A x$ becomes negative by (7.10).

Remark 7.11. If both x and $-x$ are first components of dual pairs in W^+ then $x \in V^0$ and the feasible set is contained in a linear strip given by (7.11). As there exists a hyperplane which separates V^+ and V^-, we find a vector $h \in \mathbb{R}^n$ such that $z^\top x > 0$ for all $(x, y) \in W^+$. Note that generally, $h \neq b$.

Example 7.12. Consider $\|\cdot\| = \|\cdot\|_\infty$ on $\mathbb{R}^3$. For $b = e_3$, its kernel is given by $\ker b^\top = \operatorname{span}(e_1, e_2)$. The set $V^0 \cap C$ consists of all extremal points $(\pm 1, \pm 1 \pm 1)^\top$ of $\mathbb{S}_\infty$ and $W^+ = \{((\pm 1, \pm 1, 1)^\top, e_3)\}$ as e_3 is the only extremal point of the dual norm $\|\cdot\|_1$ with $e_3^\top b > 0$. Hence any A for which there exists a closed-loop contraction with respect to $\|\cdot\|_\infty$ must already satisfy $e_1^\top A x < 0$ for $x = (1, \pm 1, \pm 1)^\top$, and $e_2^\top A x < 0$ for $x = (\pm 1, 1, \pm 1)^\top$ Therefore the conditions on a suitable feedback vector $f \in \mathbb{R}^{1 \times n}$ are $fx > e_3^\top A x, x = (\pm 1, \pm 1, 1)^\top$. Let us consider the matrix

$$A = \begin{pmatrix} -4 & 2 & -1 \\ 3 & -5 & 1 \\ 6 & -7 & 8 \end{pmatrix}.$$

The first two rows are already diagonally dominant, and evaluating the conditions for feasible feedback vectors gives the conditions

$$f(1,1,1)^\top > 7, f(1,-1,1)^\top > 21, f(-1,1,1)^\top > -5, f(-1,-1,1)^\top > 9.$$

For example, $f = (0, 0, 24)$ is a feasible feedback vector. ■

7.3 (M, β)-Stabilizability

The questions which we are interested in this section are based on the following problem. Consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{7.1}$$

where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}$ are given. We are looking for a feedback matrix $F \in \mathbb{R}^{m \times n}$ such that the closed-loop matrix $A_F = A - BF$ is dissipative or satisfies a (M, β)-stability requirement for given constants M and β. Corollary 2.57 immediately leads to

Corollary 7.13. *The system* (7.1) *is uniformly* (M, β)*-stabilizable if and only if there exists a norm* $\nu(\cdot)$ *on* $\mathbb{R}^n$ *with eccentricity* $\operatorname{ecc} \nu \leq M$ *and a feedback matrix* $F \in \mathbb{R}^{m \times n}$ *such that* $\mu_\nu(A - BF) < \beta$ *holds.*

We now identify for a given vector norm $\nu(\cdot)$ those initial vectors which have a growth rate larger than allowed, and those initial vectors for which the initial growth rate is invariant with respect to any feedback matrix. We define the following sets

$$\begin{aligned} \mathcal{M}_\beta(\nu) &= \left\{ x \in \mathbb{R}^n \,\middle|\, \sup_{\langle y,x \rangle = \nu(y)^* \nu(x)} \tfrac{\langle y, Ax \rangle}{\langle y,x \rangle} \geq \beta \right\}, \\ \mathcal{K}(\nu) &= \left\{ x \in \mathbb{R}^n \,\middle|\, \text{for all } y \in \mathbb{R}^n, \text{ with } \langle y,x \rangle = \nu(y)^* \nu(x),\, y \in \ker B^\top \right\}. \end{aligned} \tag{7.12}$$

The set $\mathcal{M}_\beta$ contains those initial vectors x_0 for which the associated solution has an initial growth rate of at least β,

$$\mathcal{M}_\beta = \left\{ x_0 \in \mathbb{R}^n \,\middle|\, \left(\tfrac{d}{dt^+} \nu(x(t, x_0))|_{t=0} \right) \geq \beta \nu(x_0) \right\}.$$

The set $\mathcal{K}$ contains those initial x_0 vectors for which the initial growth of the associated solutions remains constant under all possible feedback matrices, since all of the dual vectors are contained in the kernel of $B^\top$. If the intersection of these two sets is non-empty, then there are solutions for which the initial growth is too large, but this growth cannot be controlled by any choice of linear feedback. Hence we have the following necessary condition.

Corollary 7.14. *Given a vector norm* ν *with eccentricity* $\operatorname{ecc} \nu \leq M$*. Suppose there exists a feedback matrix* F *such that the initial growth rate of the closed-loop system satisfies* $\mu_\nu(A - BF) < \beta$*. Then*

$$\mathcal{M}_\beta(\nu) \cap \mathcal{K}(\nu) = \{0\}. \tag{7.13}$$

We will show in the following that this condition is also sufficient when dealing with weighted Euclidean norms.

7.4 Quadratic (M, β)-Stabilizability

In this section we consider weighted Euclidean norms. Moreover, there are no difficulties when allowing for complex data. Since we consider elliptical norms $\nu(x) = \|x\|_P = \langle x, Px \rangle^{1/2}$ where $P \succ 0$ is some positive definite weight in $\mathbb{C}^{n \times n}$, dual pairs are always uniquely defined by (x, Px) because $\langle Px, x \rangle_2 = \sqrt{x^* Px} \sqrt{(Px)^* P^{-1} (Px)} = \|x\|_P \|Px\|_P^*$. These weighted Euclidean norms must be compared with the standard Euclidean norm. We have already seen in Section 3.4 that this involves the use of Hermitian matrix pencils

and quadratic Liapunov functions. The Liapunov operator which maps P to $PA+A^*P$ for a given $A \in \mathbb{C}^{n\times n}$ is denoted by $\mathcal{L}_A$.
The underlying stability concept used in this section is called *quadratic (M,β)-stability*, see Boyd et al. [22].

Definition 7.15. Given the constants $M \geq 1, \beta < 0$, a matrix $A \in \mathbb{C}^{n\times n}$ is called *quadratically (M,β)-stable* if there exists a positive definite Hermitian matrix $P \succ 0$ such that

$$\kappa_2(P) \leq M^2 \quad \text{and} \quad \mathcal{L}_A(P) = PA + A^*P \prec 2\beta P.$$

A pair $(A,B) \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times m}$ is called *quadratically (M,β)-stabilizable* if there exists a matrix $F \in \mathbb{C}^{m\times n}$ such that $A - BF$ is quadratically (M,β)-stable.

If A is quadratically (M,β)-stable then by Lemma 3.31, $\mu_P(A) < \beta$. Hence A is the generator of a uniform contraction semigroup with respect to the P-norm. Let us now turn to stabilization issues. As dual vectors are explicitly known, the sets $\mathcal{M}_\beta(P) = \mathcal{M}_\beta(\|\cdot\|_P)$ and $\mathcal{K}(P) = \mathcal{K}(\|\cdot\|_P)$ of (7.12) are now given by

$$\begin{aligned} \mathcal{M}_\beta(P) &= \{x \in \mathbb{C}^n \mid x^*PAx \geq \beta x^*Px\} = \{x \in \mathbb{C}^n \mid x^*\mathcal{L}_A(P)x \geq 2\beta x^*Px\}, \\ \mathcal{K}(P) &= \{x \in \mathbb{C}^n \mid Px \in \ker B^*\} = \ker B^*P. \end{aligned} \tag{7.14}$$

Theorem 7.16. *Consider the pair $(A,B) \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times m}$ and constants $M \geq 1, \beta < 0$. The system $\dot{x} = Ax + Bu$ is quadratically (M,β)-stabilizable if and only if there exists a matrix $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that*

$$\mathcal{M}_\beta(P) \cap \ker B^*P = \{0\}. \tag{7.15}$$

Proof. The initial growth rate for a weighted quadratic P-norm is given by (3.27). From the fact that $\left\|e^{At}\right\| \leq Me^{\beta t} \iff \left\|e^{(A-\beta I)t}\right\| \leq M,\ t \geq 0$ the following equivalences hold for any P-norm

$$\mu_P(A) < \beta \iff \forall x \in \mathbb{C}^n \backslash \{0\} \colon \langle x, (PA + A^*P)x\rangle < 2\beta\langle x, Px\rangle \iff \mathcal{L}_A(P) \prec 2\beta P, \tag{7.16}$$

where the Hermitian order relation is given with respect to the standard inner product. Let us first assume that there exists a suitable $P \succ 0$ with $\kappa_2(P) \leq M^2$ and $\mathcal{M}_P(\beta) \cap \ker B^*P = \{0\}$. We show that there exists a feedback matrix F such that $\mu(A - BF) < \beta$ by considering $\mathcal{L}_{A-BF}(P)$. Applying a QR-decomposition on B and transforming the data with the resulting unitary matrix Q, we get the following partition of the matrices, where we conveniently retain their names

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} R \\ 0 \end{pmatrix}, \qquad BF = \begin{pmatrix} G_1 & G_2 \\ 0 & 0 \end{pmatrix}, \qquad P = \begin{pmatrix} P_1 & P_{12} \\ P_{12}^* & P_2 \end{pmatrix}. \tag{7.17}$$

According to this partition, the blocks of $\mathcal{L}_{A-BF}(P) = \left(\begin{smallmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{12}^* & \mathcal{L}_{22} \end{smallmatrix}\right)$ take the form

$$\begin{aligned} \mathcal{L}_{11} &= \mathcal{L}_{A_{11}-G_1}(P_1) + P_{12}A_{21} + A_{21}^*P_{12}^*, \\ \mathcal{L}_{12} &= P_1(A_{12} - G_2) + P_{12}A_{22} + (A_{11} - G_1)^*P_{12} + A_{21}^*P_2, \\ \mathcal{L}_{22} &= \mathcal{L}_{A_{22}}(P_2) + P_{12}^*(A_{12} - G_2) + (A_{12} - G_2)^*P_{12}. \end{aligned}$$

The kernel of B^*P is the largest subspace which is invariant under changes of the feedback matrix. This can be seen as follows. Since R in (7.17) is invertible, the kernel of B^*P is spanned by the columns of $\left(\begin{smallmatrix}-P_1^{-1}P_{12}\\ I\end{smallmatrix}\right)$. Here, $P_1 \succ 0$ is invertible as it is a principal submatrix of $P \succ 0$. The term

$$\begin{aligned}
&\begin{pmatrix}-P_{12}^*P_1^{-1} & I\end{pmatrix} \mathfrak{L}_{A-BF}(P) \begin{pmatrix}-P_1^{-1}P_{12}\\ I\end{pmatrix}\\
&= P_{12}^*P_1^{-1}\mathfrak{L}_{11}P_1^{-1}P_{12}^* - P_{12}^*P_1^{-1}\mathfrak{L}_{12} - \mathfrak{L}_{12}^*P_1^{-1}P_{12} + \mathfrak{L}_{22}\\
&= P_{12}^*P_1^{-1}\mathfrak{L}_{A_{11}-G_1}(P_1)P_1^{-1}P_{12}^* + \mathfrak{L}_{A_{21}P_1^{-1}P_{12}}(P_{12}^*P_1^{-1}P_{12})\\
&\;- \Big(\mathfrak{L}_{A_{22}}(P_{12}^*P_1^{-1}P_{12}) + P_{12}^*P_1^{-1}\mathfrak{L}_{A_{11}-G_1}(P_1)P_1^{-1}P_{12} + \mathfrak{L}_{A_{21}P_1^{-1}P_{12}}(P_2)\Big) + \mathfrak{L}_{A_{22}}(P_2)\\
&= \mathfrak{L}_{A_{22}-A_{21}P_1^{-1}P_{12}}(P_2 - P_{12}^*P_1^{-1}P_{12})
\end{aligned} \tag{7.18}$$

does not depend on the choice of the feedback matrix. Hence, for every $x \in \ker B^*P$ the term $x^*\mathfrak{L}_{A-BF}(P)x$ is independent of F. But if we take a vector $x = \left(\begin{smallmatrix}x_1\\ 0\end{smallmatrix}\right) \notin \ker B^*P$ then $x \mapsto x^*\mathfrak{L}_{A-BF}(P)x = x_1^*\mathfrak{L}_{11}x_1$ depends on F, thus $\ker B^*P$ is the largest subspace such that $x \mapsto x^*\mathfrak{L}_{A-BF}(P)x = x^*\mathfrak{L}_A(P)x$ is independent of F. To achieve a growth rate of less than β with a suitable feedback matrix F, the inequality $\mathfrak{L}_{A-BF}(P) \prec 2\beta P$ has to hold, see (7.16), which transforms into

$$\begin{pmatrix}I & 0\\ -P_{12}^*P_1^{-1} & I\end{pmatrix} \mathfrak{L}_{A-BF}(P) \begin{pmatrix}I & -P_1^{-1}P_{12}\\ 0 & I\end{pmatrix} \prec 2\beta \begin{pmatrix}P_1 & 0\\ 0 & P_2 - P_{12}^*P_1^{-1}P_{12}\end{pmatrix}.$$

Hence the following matrix must be negative definite.

$$\begin{pmatrix}\mathfrak{L}_{11} - 2\beta P_1 & -\mathfrak{L}_{11}P_1^{-1}P_{12} + \mathfrak{L}_{12}\\ -P_{12}^*P_1^{-1}\mathfrak{L}_{11} + \mathfrak{L}_{12}^* & \mathfrak{L}_{A_{22}-A_{21}P_1^{-1}P_{12}-\beta I}(P_2 - P_{12}^*P_1^{-1}P_{12})\end{pmatrix} \prec 0. \tag{7.19}$$

Using a Schur complement this is equivalent to the following two conditions

$$\mathfrak{L}_{A_{22}-A_{21}P_1^{-1}P_{12}-\beta I}(P_2 - P_{12}^*P_1^{-1}P_{12}) \prec 0, \tag{7.20}$$

$$\mathfrak{L}_{A_{11}-G_1-\beta I}(P_1)+P_{12}A_{21}+A_{21}^*P_{12}^*-K^*\Big(\mathfrak{L}_{A_{22}-A_{21}P_1^{-1}P_{12}-\beta I}(P_{22}-P_{12}^*P_1^{-1}P_{12})\Big)^{-1}K \prec 0, \tag{7.21}$$

where $K := -\mathfrak{L}_{11}P_1^{-1}P_{12} + \mathfrak{L}_{12}$ is the upper right block in (7.19), which is given explicitly by

$$K = P_1(A_{12} - G_2 - (A_{11} - G_1)P_1^{-1}P_{12}) + P_{12}(A_{22} - A_{21}P_1^{-1}P_{12}) + A_{21}^*(P_2 - P_{12}^*P_1^{-1}P_{12}).$$

The kernel condition (7.15) is equivalent to the statement that for all $x \in \ker B^*P$, $x^*\mathfrak{L}_{A-\beta I}(P)x < 0$. Then the negative definiteness of the first condition (7.20) is guaranteed by (7.18). The second condition (7.21) may be satisfied by choosing F_1 in such a way that $\mathfrak{L}_{A_{11}-G_1-\beta I}(P_1) \prec -(P_{12}A_{21} + A_{21}^*P_{12}^*)$ where $G_1 = RF_1$. Therefore, if the kernel condition (7.15) is satisfied then there exists a quadratically (M,β)-stabilizing feedback F. Conversely, if the pair (A,B) is quadratically (M,β)-stabilizable there exists $P \succ 0$ with $\kappa(P) \le M^2$ and a feedback matrix F such that the Liapunov inequality of (7.16) holds. Then it also holds on $\ker B^*P$, such that $\ker B^*P \cap \mathcal{M}_\beta(P) = \{0\}$. □

From the preceding proof we have the following reformulation of the kernel condition.

Corollary 7.17. *Using the notation from Theorem 7.16 and the partition* (7.17)*, the kernel condition* (7.15) *is equivalent to the negative definiteness of a Liapunov matrix,*

$$\mathcal{M}_\beta(P) \cap \ker B^*P = \{0\} \iff \mathcal{L}_{A_{22}-A_{21}P_1^{-1}P_{12}-\beta I}(P_2 - P_{12}^*P_1^{-1}P_{12}) \prec 0.$$

This characterization gives necessary conditions on the inner product matrix P as the matrix $A|_{\ker} := A_{22} - A_{21}P_1^{-1}P_{12}$ has to be stable, $P|_{\ker} := P_2 - P_{12}^*P_1^{-1}P_{12} \succ 0$ has to hold, and $\mathcal{L}_{A|_{\ker}}(P|_{\ker})$ has to be negative definite. To select a weight P one could proceed as follows. Choose P_1 and P_{12} such that $A_{22} - A_{21}P_1^{-1}P_{12}$ is stable. Then choose P_2 in such way that $P_2 - P_{12}^*P_1^{-1}P_{12}$ is positive definite and

$$\mathcal{L}_{A|_{\ker}}(P_2) \prec \mathcal{L}_{A|_{\ker}}(P_{12}^*P_1^{-1}P_{12}). \tag{7.22}$$

Let us now show that the feedback matrix F may be chosen in a standard way.

Corollary 7.18. *Consider the pair $(A, B) \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times m}$ and constants $M \geq 1, \beta < 0$. The following statements are equivalent.*

(i) The system $\dot{x} = Ax + Bu$ is quadratically (M, β)-stabilizable.

(ii) There exist $\gamma \in \mathbb{R}$ and $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that

$$\mathcal{L}_{A-\gamma BB^*P}(P) \prec 2\beta P. \tag{7.23}$$

In this case, (7.23) *holds for all $\gamma' \geq \gamma$.*

(iii) There exists a $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that (7.15) *is satisfied.*

Proof. The equivalence of *(i)* and *(iii)* has been shown in Theorem 7.16. Clearly, if $F = \gamma B^*P$ satisfies $\mathcal{L}_{A-BF}(P) \prec 2\beta P$ then by definition, (A, B) is quadratically (M, β)-stabilizable for any $M \geq \sqrt{\kappa_2(P)}$, hence *(ii)* $\Longrightarrow$ *(i)*. For $F' = \gamma' B^*P$ with $\gamma' \geq \gamma$ we have

$$\begin{aligned}\mathcal{L}_{A-BF'}(P) = \mathcal{L}_{A-\gamma' BB^*P}(P) &= PA + A^*P - 2\gamma' PBB^*P \\ &\preceq PA + A^*P - 2\gamma PBB^*P = \mathcal{L}_{A-\gamma BB^*P}(P) = \mathcal{L}_{A-BF}(P) \preceq 2\beta P.\end{aligned}$$

Hence F' also stabilizes the pair (A, B). Let us now show that *(iii)* implies *(ii)*, that is, if (7.15) holds, there exists $\gamma \in \mathbb{R}$ such that $F = \gamma B^*P$ is a stabilizing feedback. For this let us take a look into (7.19) with the data $G_1 = \gamma RR^*P_1$ and $G_2 = \gamma RR^*P_{12}$ which correspond to $F = \gamma B^*P$. Then

$$\begin{aligned}\mathcal{L}_{11} &= \mathcal{L}_{A_{11}}(P_1) - 2\gamma P_1RR^*P_1 + P_{12}A_{21} + A_{21}^*P_{12}^*, \\ K &= P_1(A_{12} - A_{11}P_1^{-1}P_{12}) + P_{12}(A_{22} - A_{21}P_1^{-1}P_{12}) + A_{21}^*(P_2 - P_{12}^*P_1^{-1}P_{12}).\end{aligned}$$

Hence the off-diagonal block is independent of the choice of γ while in the upper left block γ is a scaling factor for the positive definite matrix $P_1RR^*P_1$. Now the lower right block is already negative definite by (7.15), and so there exists a $\gamma \in \mathbb{R}$ such that (7.19) is negative definite. Thus the pair (A, B) has been stabilized by the feedback matrix $F = \gamma B^*P$ □

Remark 7.19. Suppose $\beta < 0$ is fixed. If we choose the feedback matrix F to be of the form $F = \gamma B^*P$ the initial growth rate condition $\mu_P(A - BF) < \beta < 0$ gives rise to the following *parameterized Riccati inequality*

$$PA + A^*P - 2\gamma PBB^*P - 2\beta P \prec 0.$$

By Theorem 7.16, positive definite solutions $P \succ 0$ exist if and only if (7.15) holds.

We note two further consequences of Theorem 7.16 and Corollary 7.18 which simplify the situation for the case that minimizing the M is more important than guaranteeing a certain rate of decay $\beta < 0$. The following corollary presents conditions for quadratic (M,β)-stabilization for arbitrary $\beta < 0$.

Corollary 7.20. *Consider the pair (A,B), and let $M \geq 1$. The following statements are equivalent.*

(i) For some $\beta < 0$ the system $\dot{x} = Ax + Bu$ is quadratically (M,β)-stabilizable.

(ii) There exist $\gamma > 0$ and a matrix $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that

$$\mathfrak{L}_{A-\gamma BB^*P}(P) \prec 0. \tag{7.24}$$

(iii) There exists a matrix $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that

$$\{v \in \mathbb{C}^n \mid v^*(A^*P + PA)v \geq 0\} \cap \ker B^*P = \{0\}. \tag{7.25}$$

(iv) There exists a matrix $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that

$$x \in \ker B^* \backslash \{0\} \implies x^*(AP^{-1} + P^{-1}A^*)x < 0. \tag{7.26}$$

Proof. This is immediate from Theorem 7.16 and Corollary 7.18. □

In the previous result, condition (7.26) is remarkable as it allows for a nice geometric interpretation of the problem. Namely, given the pair (A,B) the question is if we can find a matrix $P \succ 0$ with $\kappa_2(P) \leq M^2$ such that for all $x \in \ker B^*$, $x \neq 0$ the condition $\operatorname{Re}\langle P^{-1}x, A^*x\rangle_2 < 0$ holds. In other words, the system $\dot{x} = A^*x$ is strictly dissipative in the weighted inner product $\langle\cdot,\cdot\rangle_{P^{-1}}$ on the subspace $\ker B^*$. A case of particular interest is that of feedback matrices F such that the closed loop system matrix $A - BF$ generates a strict contraction semigroup for the spectral norm, that is, if we specialize to the case $M = 1$ with $\beta < 0$ arbitrary. Then we obtain

Corollary 7.21. *Consider the pair (A,B). The following statements are equivalent.*

(i) there exists a feedback matrix F such that $A - BF$ generates a uniform contraction semigroup with respect to the spectral norm,

(ii) there exists $\gamma > 0$ such that $A - \gamma BB^*$ generates a uniform contraction semigroup with respect to the spectral norm,

(iii) it holds that

$$x \in \ker B^* \backslash \{0\} \implies x^*(A + A^*)x < 0. \tag{7.27}$$

Proof. In fact, $A - BF$ generates a uniform contraction semigroup with respect to the spectral norm, if there exists $\beta < 0$ such that for all $t \geq 0$

$$\left\| e^{(A-BF)t} \right\| \leq e^{\beta t}.$$

Therefore Corollary 7.20 is applicable with $M = 1$. In this case, the positive definite matrices $P \succ 0$ with $\kappa_2(P) \leq M^2 = 1$ occurring in the statements of Corollary 7.20 are necessarily multiples of the identity. □

7.5 Quadratic Programs for (M, β)-Stabilization

In this section we briefly discuss how the geometric characterizations for strict quadratic (M, β)-stabilizability obtained in the previous section can be reformulated in terms of quadratic programs (QPs) with constraints given by linear matrix inequalities (LMIs). We refer to [22] for an overview of applications of LMIs in control.
By Theorem 7.16 the system $\dot{x} = Ax + Bu$ is quadratically (M, β)-stabilizable if and only if the following set is non-empty,

$$\mathcal{N} := \left\{ P \in \mathcal{H}^n(\mathbb{C}) \,\middle|\, P \succ 0, \kappa_2(P) \leq M^2 \text{ and } \mathcal{M}_\beta(P) \cap \ker B^* P = \{0\} \right\}.$$

It is easy to see that $\mathcal{N}$ is a subcone of the cone of positive semidefinite matrices $\mathcal{H}^n_+(\mathbb{C})$, so that if $\mathcal{N}$ is non-empty, then there is a $P \in \mathcal{N}$ with $\sigma(P) \subset [M^{-2}, 1]$, hence $\|P\|_2 \leq 1$.
Furthermore, the set $\mathcal{N}$ is non-empty if and only if there are $P \succ 0, \kappa(P) \leq M^2$ and $F \in \mathbb{C}^{m \times n}$ such that $\mu_P(A - BF) < \beta$ is satisfied. This inequality has the disadvantage that the unknowns P and F do not appear linearly, but by setting $Q = P^{-1}$ and $F = XP$ we obtain an LMI from $\mathcal{L}_{A-BF}(P) \prec 2\beta P$ by pre- and post-multiplying with Q. So all our conditions can be summarized by the LMI

$$\begin{aligned} I \preceq Q \preceq M^2 I, \\ AQ + QA^* - (BX + X^*B^*) \prec 2\beta Q. \end{aligned} \tag{7.28}$$

where the first inequality ensures that the eigenvalues of Q are contained in the interval $[1, M^2]$ which implies that $\kappa(Q) = \kappa(P) \leq M^2$ ad the second inequality implies that XP^{-1} is a quadratically (M, β)-stabilizing feedback for the pair (A, B). By this simple reformulation we obtain another condition for quadratic (M, β)-stabilizability as an immediate corollary to Theorem 7.16.

Corollary 7.22. *Consider the pair (A, B) and constants $M \geq 1, \beta < 0$. The following statements are equivalent:*

(i) The system $\dot{x} = Ax + Bu$ is quadratically (M, β)-stabilizable.

(ii) The LMI (7.28) *is feasible, i.e., there exists a solution* $(Q,X)\in\mathbb{C}^{n\times n}\times\mathbb{C}^{m\times n}$ *of* (7.28).

(iii) There exists a solution $(Q,\varrho)\in\mathbb{C}^{n\times n}\times\mathbb{R}$ *of the LMI*

$$\begin{aligned} I \preceq Q \preceq M^2 I, \\ AQ + QA^* - 2\varrho BB^* \prec 2\beta Q. \end{aligned} \tag{7.29}$$

Proof. The equivalence of *(i)* and *(ii)* was shown in the derivation of (7.28). For the equivalence *(ii)*⇔*(iii)*, note first that (Q,ϱ) solves (7.29) if and only if $(Q,\varrho B^*)$ solves (7.28), so that *(iii)* implies *(ii)*. Furthermore, by Corollary 7.18, quadratic (M,β)-stabilizability is equivalent to the existence of a stabilizing feedback of the form $F=\varrho B^*P$. In this case $(Q,\varrho B^*)$ solves (7.28), which implies *(iii)*. □

The advantage of *(iii)* compared to *(ii)* in Corollary 7.22 is that the dimension of the parameter space is significantly reduced, depending on the dimension of B.

Remark 7.23. Using Corollary 7.22 we can add further design objectives depending on the specific problem since quadratic optimization problems may be solved on solution sets of LMIs. For example, if a feedback F of small norm is desirable, then it is advantageous to minimize $\gamma\geq 0$ under the constraints (7.28) and

$$\begin{pmatrix} \gamma I & X \\ X^* & \gamma I \end{pmatrix} \succeq 0. \tag{7.30}$$

Using the Schur complement it may be seen that (7.30) is equivalent to $\gamma^2 I - XX^* \succeq 0$, i.e., $\gamma \geq \|X\|$. As the solution set of (7.28) is not necessarily closed, there may not be an optimal solution, but at least the optimization problem yields matrices X with norm close to optimal and for the corresponding stabilizing feedback F we have $\|F\|\leq\|X\|\,\|P\|\leq\|X\|$. Similarly, (7.29) may be used to minimize ρ.

Example 7.24. Consider the system (7.1) given by

$$A=\begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -625 \\ 0 & -1 & -30 & 400 & 0 & 0 & 250 \\ -2 & 0 & -1 & 0 & 0 & 0 & 30 \\ 5 & -1 & 5 & -1 & 0 & 0 & 200 \\ 11 & 1 & 25 & -10 & -1 & 1 & -200 \\ 200 & 0 & 0 & -150 & -100 & -1 & -1000 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, B=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{7.31}$$

The transient behaviour of $t\mapsto\left\|e^{At}\right\|$ is plotted in Figure 7.3, the eigenvalues of A are $-1, -1\pm 10i, -1\pm 20i, -1\pm 25i$. The vector $x=e_7-e_6$ satisfies $x^*Ax=998$, hence the system is not a contraction with respect to the spectral norm, and as $x\in\ker B^*$, there does not exist a feedback matrix such that the closed loop system generates a contraction. The matrix B is already an upper triangular matrix[1], hence Corollary 7.17 is directly applicable.

[1] We already applied an orthogonal transformation on both A and B. The original matrix A can be found in Example 2.15.

Using partition (7.17), the submatrix A_{22} is already stable. Hence let us set $P_{12} = 0$. We therefore need to find a P_2 such that $\mathcal{L}_{A_{22}}(P_2) \prec 0$ which ensures that there exists a feedback matrix F such that the closed loop system overshoots at most $\kappa(P)^{1/2}$ where $P_1 = I, P_{12} = 0$. Using a quadratic program we find a positive definite matrix P_2 with $\kappa(P_2)^{1/2} = 315$. Hence there exists a feedback matrix such that the transient excursion of closed loop system stays below 315. And indeed choosing $F = -10B^*$ gives even an excursion below 250, as Figure 7.4 shows. ■

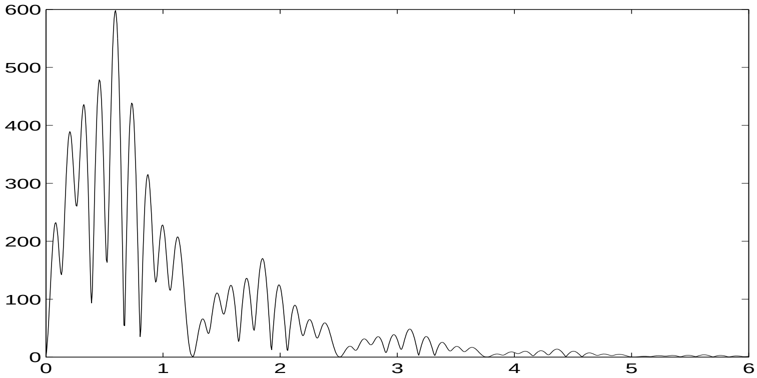

Figure 7.3: Transient excursions of an asymptotically stable linear system.

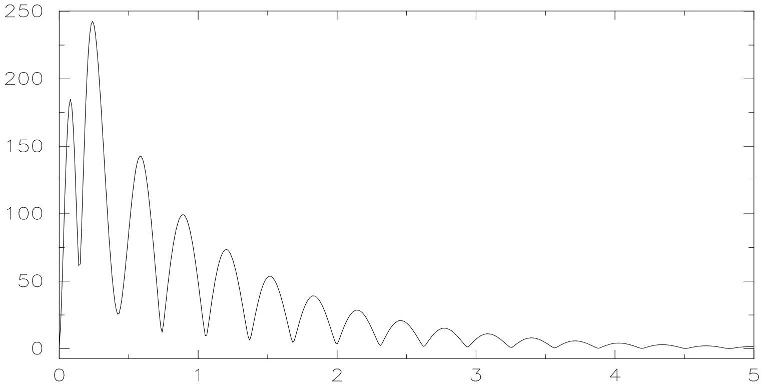

Figure 7.4: Transient humps of the closed loop system.

7.6 Notes and References

The use of time-varying linear feedback to reduce transients has been studied in Hinrichsen and Pritchard [66] and Pritchard [117]. The relation of this problem to the pole placement problem has been investigated by Izmailov [74]. For some new results in this direction see Hauksdottir [52, 53].

The material in this section is based upon the articles by Hinrichsen, Plischke and Wirth [63] and Plischke and Wirth [115]. The quadratic (M,β)-stabilization has been discussed in the former article, while the generalization to arbitrary norms was presented in the latter article. Some ideas on the influence of state feedback to the transient amplification can be found in Pritchard [117]. The link from quadratic (M, β)-stabilizability to parameterized Riccati equations is studied in Hinrichsen et al. [62], see also Hinrichsen and Pritchard [67, Section 5.5].

Contractibility is studied by Malmgren and Nordström [103, 104] for discrete-time systems. The article by Moore and Bhattacharyya [109] discusses discrete-time systems for which the overshoot is minimized via linear programming methods. A cursory discussion of LMI methods for the quadratic (M, β)-stabilization can be found in Boyd et al. [22]. Petersen [114] considers "quadratic" stabilizability which leads to optimization problems involving Riccati inequalities.

Drăgan and Halanay [35] discuss methods of finding high-gain feedback matrices which stabilize continuous-time systems with fast decay of the output and avoiding, if possible, overshoot phenomena. Scalar systems are studied in León de la Barra and Fernández [94]. Transient performance in classical output regulation is discussed in Saberi et al. [121], but here the performance measure is the integrated tracking error, and not the maximal error. Classically, the transient behaviour is analysed using the transmission zeros of the system, see Qiu and Davison [119]. In our approach, their role is reflected in condition (7.15).

List of Symbols

Sets and Norms

$\mathbb{N}$	Set of natural numbers, $\{0, 1, 2, \dots\}$
$\mathbb{Z}, \mathbb{R}, \mathbb{C}$	Ring of integers, fields of real numbers and complex numbers
$\mathbb{N}^*, \mathbb{R}^*, \mathbb{C}^*$	$\mathbb{N}$, $\mathbb{R}$, $\mathbb{C}$ without 0
$\mathbb{K}$	Number field, either $\mathbb{R}$ or $\mathbb{C}$
$\mathbb{K}^n$	Vector space over $\mathbb{K}$ of dimension n
$\mathbb{K}^{n \times m}$	Vector space of matrices over $\mathbb{K}$ with n columns and m rows
$\operatorname{Re} z, \operatorname{Im} z$	Real and imaginary parts of a complex number z
$\mathbb{C}_-, \mathbb{C}_+$	Open left half-plane, open right half-plane
$\bar{S}$, $\mathring{S}$, ∂S, S^{C}	Closure, interior, boundary and complement of a set S.
$\lVert\cdot\rVert, \nu(\cdot)$	Vector norms and the induced operator norms
$\lVert\cdot\rVert_p, \lVert\cdot\rVert_F$	p-norms ($p = 1, 2, \infty$), Frobenius norm
$\mathbb{B}$, $\mathbb{B}^*$, $\mathbb{B}_\nu$	Closed unit balls of the norm $\lVert\cdot\rVert$, of its dual norm, and of the norm $\nu(\cdot)$
$\mu(A), \mu_\nu(A)$	Initial growth rate of A with respect to the norm $\lVert\cdot\rVert$ or $\nu(\cdot)$
$\operatorname{ecc}(\nu, \lVert\cdot\rVert)$	Eccentricity of the norms $\nu(\cdot)$ and $\lVert\cdot\rVert$
$M_\beta(A)$	Transient growth of A with respect to the rate β

Operators and Matrices

$\mathcal{L}(X, Y)$	Space of bounded linear operators from X into Y
$\mathcal{L}(X)$	Space of bounded linear operators on X, $\mathcal{L}(X) = \mathcal{L}(X, X)$
$C(I, X)$	Space of continuous functions $f : I \to X$
$L^2(I, X)$	Space of square-integrable functions $f : I \to X$
$\ell^2(\mathbb{K})$	Space of square-summable sequences in $\mathbb{K}$
$\bigoplus_{k \in \mathbb{N}} X_k$	Direct sum of Hilbert spaces X_k
$D(A)$	Domain of the linear operator A
A^*	Adjoint of a matrix or operator A
I	Identity matrix or operator
$\sigma(A), \varrho(A)$	Spectrum and resolvent set of A
$\alpha(A), \rho(A)$	Spectral abscissa and spectral radius of A
$R(s, A)$	Resolvent of A, $R(s, A) = (sI - A)^{-1}$
$A^\top, A^{-1}$	Transpose of A, inverse of A
$\det A, \operatorname{trace} A$	Determinant and trace of A
$\kappa(A)$	Condition number of A

$\langle x, y\rangle_2$	Inner product on $\mathbb{K}^n$, y^*x
$A \otimes B$	Kronecker product of the matrices A and B
$\operatorname{vec} X$	Vectorization of the matrix A
$\ker A, \operatorname{im} A$	Kernel and image of a matrix A

Positivity

$\mathbb{R}_+$	Set of nonnegative real numbers
$\mathbb{R}_+^{n\times n}, \mathbb{R}_{\mathrm{M}}^{n\times n}$	Set of nonnegative matrices, set of Metzler matrices
$\operatorname{span}(S), \operatorname{conv}(S), \operatorname{cone}(S)$	Linear hull (span), convex hull and convex cone of the set S
$M(A)$	Metzler part of a matrix A
$\|A\|$	Componentwise modulus of a matrix A
$\operatorname{Diag}(A)$	Diagonal matrix with diagonal entries from A
$\operatorname{diag}(v)$	Diagonal embedding of a vector v, $\mathbb{K}^n \to \mathbb{K}^{n\times n}$
$\operatorname{diag}(A)$	Diagonal extraction from a matrix A, $\mathbb{K}^{n\times n} \to \mathbb{K}^n$
$A > B$	Componentwise comparison of the matrices A and B
$A > 0, v > 0$	Strict positivity of the matrix A or the vector v
$\mathbf{1}$	Vector of ones
$\mathcal{H}^n(\mathbb{C}), \mathcal{H}^n(\mathbb{R})$	Vector spaces of Hermitian and symmetric matrices
$P \succ 0, P \succeq 0$	Positive definite and positive semidefinite Hermitian matrix P
$\mathcal{H}_+^n(\mathbb{C}), \mathcal{H}_+^n(\mathbb{R})$	Convex cone of positive semidefinite Hermitian matrices
$\mathcal{L}_A(P)$	Liapunov operator, $P \mapsto PA + A^*P$

Bibliography

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. DuCroz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*, volume 9 of *Software - Environments - Tools.* SIAM Publications, Philadelphia, PA, 3rd edition, 1999.

[2] T. Ando. Set of matrices with common Lyapunov solution. *Arch. Math.*, 77(1):76–84, 2001.

[3] A. Aptekarev. A direct proof of Trefethen's conjecture. In A. Goncar, E. Saff, et al., editors, *Methods of Approximation Theory in Complex Analysis and Mathematical Physics*, volume 1550 of *Lecture Notes in Mathematics*, pages 147–148. Springer-Verlag, 1993.

[4] W. Arendt, C. J. Batty, M. Hieber, and F. Neubrander. *Vector-valued Laplace Transforms and Cauchy Problems*, volume 96 of *Monographs in Mathematics.* Birkhäuser, Basel, 2001.

[5] V. I. Arnold. *Ordinary Differential Equations.* MIT Press, Cambridge, MA, 1978.

[6] J.-P. Aubin and H. Frankowska. *Set Valued Analysis.* Birkhäuser, Boston, 1990.

[7] B. Aupetit and D. Drissi. Conformal transformations of dissipative operators. *Math. Proc. R. Ir. Acad.*, 99A(2):141–153, 1999.

[8] J. S. Baggett and L. N. Trefethen. Low-dimensional models of subcritical transition to turbulence. *Physics of Fluids*, 9:1043–1053, 1997.

[9] V. Balakrishnan and S. Boyd. Existence and uniqueness of optimal matrix scalings. *SIAM J. Matrix. Anal. & Appl.*, 16(1), 1995.

[10] A. Bátkai and S. Piazzera. Semigroups and linear partial differential equations with delay. *J. Math. Anal. Appl.*, 264(1):1–20, 2001.

[11] F. L. Bauer. On the field of values subordinate to a norm. *Numer. Math.*, 4:103–113, 1962.

[12] F. L. Bauer, J. Stoer, and C. Witzgall. Absolute and monotonic norms. *Numer. Math.*, 3:257–264, 1961.

[13] A. Bellen and M. Zennaro. *Numerical Methods for Delay Differential Equations.* Oxford University Press, Oxford, 2003.

[14] R. Bellman. Vector Liapunov functions. *SIAM J. Cont.*, 1:32–34, 1962.

[15] R. Bellman and K. L. Cooke. *Differential-Difference Equations.* Number 6 in Mathematics in Science and Engineering. Academic Press, New York, 1963.

[16] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. K. Mitter. *Representation and Control of Infinite Dimensional Systems. Vol. I.* Systems & Control: Foundations & Applications. Birkhäuser, Boston, MA, 1992.

[17] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, volume 9 of *Classics in Applied Mathematics.* SIAM Publications, Philadelphia, PA, 1994.

[18] R. Bhatia. *Matrix Analysis.* Springer-Verlag, New York, 1997.

[19] P.-A. Bliman. LMI characterization of the strong delay-independent stability of linear delay systems via quadratic Lyapunov-Krasovskii functionals. *Syst. Control Lett.*, 43(4):263–274, 2001.

[20] A. Böttcher. Transient behavior of powers and exponentials of large Toeplitz matrices. *Electr. Trans. Num. Anal.*, 18:1–41, 2004.

[21] A. Böttcher and B. Silbermann. *Introduction to Large Truncated Toeplitz Matrices.* Springer-Verlag, New York, 1999.

[22] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory*, volume 15 of *Studies in Applied Mathematics.* SIAM Publications, Philadelphia, PA, 1994.

[23] D. Breda, S. Maset, and R. Vermiglio. Computing the characteristic roots for delay differential equations. *IMA J. Numer. Anal.*, 24:1–19, 2004.

[24] J. V. Burke, A. S. Lewis, and M. L. Overton. Optimization and pseudospectra, with applications to robust stability. *SIAM J. Matrix. Anal. & Appl.*, 25(1):80–104 (electronic), 2003.

[25] W. B. Castelan and E. F. Infante. On a functional equation arising in the stability theory of difference-differential equations. *Q. Appl. Math.*, 35:311–319, 1977.

[26] J. Chen, G. Gu, and C. N. Nett. A new method for computing delay margins for stability of linear delay systems. *Syst. Control Lett.*, 26(2):107–117, 1995.

[27] N. Cohen and I. Lewkowicz. A pair of matrices sharing common Lyapunov solutions – A closer look. *Lin. Alg. Appl.*, 360:83–104, 2003.

[28] J. B. Conway. *Functions of One Complex Variable.* Springer-Verlag, New York, 2nd edition, 1978.

[29] R. F. Curtain and H. J. Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory.* Number 21 in Texts in Applied Mathematics. Springer-Verlag, New York, 1995.

[30] G. Dahlquist. Stability and error bounds in the numerical integration of ordinary differential equations. *Transactions Royal Inst. of Technology*, 130, 1959.

[31] J. L. Daleckiĭ and M. G. Kreĭn. *Stability of Solutions of Differential Equations in Banach Spaces.* Number 43 in Translations of Mathematical Monographs. American Mathematical Society, Providence, RI, 1974.

[32] R. Datko. Lyapunov functionals for certain linear delay differential equations in a Hilbert space. *J. Math. Anal. Appl.*, 76:37–57, 1980.

[33] O. Diekmann, S. A. van Gils, S. M. Verduyn Lunel, and H.-O. Walther. *Delay Equations*, volume 110 of *Applied Mathematical Sciences.* Springer-Verlag, New York, 1995.

[34] J. Dieudonné. *Foundations of Modern Analysis*, volume 10-I of *Pure and Applied Mathematics.* Academic Press, New York, 1969.

[35] V. Drăgan and A. Halanay. High-gain feedback stabilization of linear systems. *Int. J. Control*, 45:549–577, 1987.

[36] N. Dunford and J. T. Schwartz. *Linear Operators, Part I: General Theory.* Interscience, New York, 1958.

[37] M. Embree and L. N. Trefethen. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators.* Princeton University Press, 2005. To appear.

[38] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations.* Number 194 in Graduate Texts in Mathematics. Springer-Verlag, Berlin, 2000.

[39] D. K. Faddejew and W. N. Faddejewa. *Numerische Methoden der Linearen Algebra.* R. Oldenbourg Verlag, München-Wien, 1976.

[40] L. Farina and S. Rinaldi. *Positive Linear Systems.* John Wiley & Sons, New York, 2000.

[41] W. Feller. On the generation of unbounded semi-groups of bounded linear operators. *Ann. of Math., Ser. 2*, 58:166–174, 1953.

[42] M. Fiedler and V. Pták. On matrices with non-positive off-diagonal elements and positive principal minors. *Czechoslovak Math. J.*, 12(87):382–400, 1962.

[43] A. Fischer, D. Hinrichsen, and N. K. Son. Stability radii of Metzler operators. *Vietnam J. of Mathematics*, 26:147–163, 1998.

[44] F. R. Gantmacher. *The Theory of Matrices (Vol. I).* Chelsea, New York, 1959.

[45] F. R. Gantmacher. *The Theory of Matrices (Vol. II).* Chelsea, New York, 1959.

[46] M. I. Gil'. *Stability of Finite and Infinite Dimensional Systems.* Kluwer Academic Publishers, Boston/Dordrecht/London, 1998.

[47] I. Gohberg, M. Kaashoek, and J. Kos. The asymptotic behavior of the singular values of matrix powers and applications. *Lin. Alg. Appl.*, 245:55–76, 1996.

[48] G. H. Golub and C. F. van Loan. *Matrix Computations.* Number 3 in Johns Hopkins series in the mathematical sciences. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.

[49] S. Großmann. Wie entsteht eigentlich Turbulenz? *Phys. Bl.*, 51(7/8):641–646, 1995.

[50] S. Grossmann. The onset of shear flow turbulence. *Reviews of Modern Physics*, 72(2):603–618, 2000.

[51] J. K. Hale and S. M. Verduyn Lunel. *Introduction to Functional Differential Equations*, volume 99 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1993.

[52] A. S. Hauksdóttir. Analytic expressions of transfer function responses and choice of numerator coefficients (zeros). *IEEE Trans. Automat. Contr.*, 41(10):1482–1488, 1996.

[53] A. S. Hauksdóttir. Optimal zero locations of continuous-time systems with distinct poles tracking reference step responses. *Dyn. Contin. Discrete Impuls. Syst., Ser. B, Appl. Algorithms*, 11(3):353–361, 2004.

[54] P. Henrici. Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices. *Numer. Math.*, 4:24–40, 1962.

[55] D. Hertz, E. I. Jury, and E. Zeheb. Stability independent and dependent of delay for delay differential systems. *J. Franklin Inst.*, 318:143–150, 1984.

[56] N. J. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM Publications, Philadelphia, PA, 2nd edition, 2002.

[57] N. J. Higham. The scaling and squaring method for the matrix exponential revisted. Numerical Analysis Report 452, Manchester Center for Computational Mathematics, University of Manchester, Manchester, UK, 2004.

[58] I. Higueras and G. Söderlind. Logarithmic norms and nonlinear DAE stability. *BIT*, 42(4):823–841, 2002.

[59] E. Hille and R. S. Phillips. *Functional Analysis and Semigroups.* American Mathematical Society, Providence, RI, 1957.

[60] D. Hinrichsen, E. Gallestey, and A. J. Pritchard. Spectral value sets of closed linear operators. *Proc. R. Soc. Lond. Ser. A*, 456:1398–1418, 2000.

[61] D. Hinrichsen, M. Karow, and A. J. Pritchard. Spectral value sets, stability radii and μ-functions for composite systems. *SIAM J. Control Optim.*, 2005. submitted.

[62] D. Hinrichsen, E. Plischke, and A. J. Pritchard. Liapunov and Riccati equations for practical stability. In *Proc. 6th European Control Conf. 2001, Porto, Portugal*, pages 2883–2888, 2001. Paper no. 8485 (CDROM).

[63] D. Hinrichsen, E. Plischke, and F. Wirth. State feedback stabilization with guaranteed transient bounds. In *Proc. MTNS-2002*, Notre Dame, Indiana, 2002. Paper no. 2132 (CDROM).

[64] D. Hinrichsen and A. J. Pritchard. On the robustness of stable discrete time linear systems. In *New Trends in Systems Theory (Proc. Conf. Genova, 1990)*, pages 393–400. Birkhäuser, Basel, 1991.

[65] D. Hinrichsen and A. J. Pritchard. On spectral variations under bounded real matrix perturbations. *Numer. Math.*, 60:509–524, 1992.

[66] D. Hinrichsen and A. J. Pritchard. On the transient behaviour of stable linear systems. In *Proc. MTNS-2000*, Perpignan, France, 2000.

[67] D. Hinrichsen and A. J. Pritchard. *Mathematical Systems Theory I. Modelling, State Space Analysis, Stability and Robustness.* Springer-Verlag, Berlin, 2005.

[68] D. Hinrichsen and N. K. Son. Robust stability of positive continuous time systems. *Numer. Functional Anal. Optim.*, 17:649–659, 1996.

[69] D. Hinrichsen and N. K. Son. μ-analysis and robust stability of positive linear systems. *Appl. Math. and Comp. Sci.*, 8:253–268, 1998.

[70] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1990.

[71] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, 1991.

[72] G.-D. Hu and M. Liu. The weighted logarithmic matrix norm and bounds of the matrix exponential. *Lin. Alg. Appl.*, 390:145–154, 2004.

[73] E. F. Infante and W. B. Castelan. A Liapunov functional for a matrix difference-differential equation. *J. Diff. Eqns.*, 29:439–451, 1978.

[74] R. Izmailov. The peak effect in stationary linear systems with multivariate inputs and outputs. *Autom. Rem. Control*, 49(1):40–47, 1988.

[75] K. Jänich. *Topologie.* Springer-Verlag, Berlin, 8th edition, 2005.

[76] M. Karow. *Geometry of Spectral Value Sets.* PhD thesis, University of Bremen, Germany, 2003.

[77] T. Kato. *Perturbation Theory for Linear Operators.* Springer-Verlag, Heidelberg, 2nd edition, 1980.

[78] V. L. Kharitonov. Complete type Lyapunov-Krasovskii functionals. In S.-I. Niculescu and K. Gu, editors, *Advances in Time-Delay Systems*, volume 38 of *Lecture Notes in Computational Science and Engineering*, pages 31–42. Springer-Verlag, Berlin, 2004.

[79] V. L. Kharitonov and D. Hinrichsen. Exponential estimates for time delay systems. *Syst. Control Lett.*, 53(5):395–405, 2004.

[80] V. L. Kharitonov and E. Plischke. Lyapunov matrices for time-delay systems. *Syst. Control Lett.*, 2006. To appear.

[81] V. L. Kharitonov and A. P. Zhabko. Lyapunov-Krasovskii approach to the robust stability analysis of time-delay systems. *Automatica*, 39:15–20, 2003.

[82] D. Y. Khusainov, Y. A. Komarov, and Y. A. Yun'kova. Constructing optimal Lyapunov functions for linear differential equations. *Sov. Autom. Control*, 17 (6):80–83, 1984.

[83] H. Kiendl, J. Adamy, and P. Stelzner. Vector norms as Lyapunov functions for linear systems. *IEEE Trans. Automat. Contr.*, 37(6):839–842, 1992.

[84] V. Kolmanovskii and A. Myshkis. *Applied Theory of Functional Differential Equations*, volume 85 of *Mathematics and Its Applications (Soviet Series).* Kluwer Academic Publishers, Dordrecht, 1992.

[85] Y. Komarov, Y. Yun'kova, and D. Khusainov. On the question of the existence of extremal Lyapunov functions. *Vestn. Kiev. Univ., Model. Optimizatsiya Slozhnykh Sist.*, 3:103–106, 1984.

[86] U. Krause and T. Nesemann. *Differenzengleichungen und diskrete dynamische Systeme.* Teubner, Stuttgart, 1999.

[87] M. G. Kreĭn. *Lectures on the theory of the stability of solutions of differential equations in a Banach space.* Izdat. Akad. Nauk Ukrain. SSR, Kiev, 1964. In Russian.

[88] H.-O. Kreiss. Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differentialgleichungen approximieren. *BIT*, 2:153–181, 1962.

[89] V. Lakshmikantham, S. Leela, and A. A. Martynyuk. *Practical Stability of Nonlinear Systems.* World Scientific, Singapore, 1990.

[90] P. Lancaster and M. Tismenetsky. *The Theory of Matrices.* Academic Press, Orlando, FL, 1985.

[91] C. Lanczos. *Applied Analysis.* Prentice Hall, Englewood Cliffs, NJ, 1956. Reprint: Dover, New York, 1988.

[92] S. Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 3rd revised edition, 2002.

[93] J. P. LaSalle and S. Lefschetz. *Stability by Liapunovs's Direct Method.* Academic Press, New York, 1961.

[94] B. A. León de la Barra and M. A. Fernández. Transient properties of type m continuous time scalar systems. *Automatica*, 30(9):1495–1496, 1994.

[95] D. Liu and A. Molchanov. Criteria for robust absolute stability of time-varying nonlinear continuous-time systems. *Automatica*, 38(4):627–637, 2002.

[96] J. Louisell. A stability analysis for a class of differential-delay equations having time-varying delay. In S. Busenberg and M. Martelli, editors, *Delay Differential Equations and Dynamical Systems*, volume 1475 of *Lecture Notes in Mathematics*, pages 225–242. Springer-Verlag, Berlin, 1991.

[97] J. Louisell. Stability criteria with a symmetric operator occurring in linear and nonlinear delay-differential equations. In K. Elworthy et al., editors, *Differential Equations, Dynamical Systems, and Control Science*, volume 152 of *Lect. Notes Pure Appl. Math.*, pages 159–172. Marcel Dekker, New York, 1994.

[98] J. Louisell. A matrix method for determining the imaginary axis eigenvalues of a delay system. *IEEE Trans. Automat. Contr.*, 46(12):2008–2012, 2001.

[99] J. Louisell. Stability exponent and eigenvalue abscissa by way of the imaginary axis eigenvalues. In S.-I. Niculescu and K. Gu, editors, *Advances in Time-Delay Systems*, volume 38 of *Lecture Notes in Computational Science and Engineering*, pages 193–206. Springer-Verlag, Berlin, 2004.

[100] S. M. Lozinskiĭ. Error estimates for numerical integration of ordinary differential equations. *Izv. Vysš. Učebn. Zaved Matematika*, 5:52–90, 1958. Errata 5:222, 1959.

[101] D. G. Luenberger. *Introduction to Dynamic Systems: Theory, Models and Applications.* John Wiley & Sons, New York, 1979.

[102] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA, 2nd edition, 1984.

[103] A. Malmgren and K. Nordström. A contraction property for state feedback design of linear discrete-time systems. *Automatica*, 30(9):1485–1489, 1994.

[104] A. Malmgren and K. Nordström. Optimal state feedback control with a prescribed contraction property. *Automatica*, 30(11):1751–1756, 1994.

[105] J. E. Marshall, H. Górecki, A. Korytowski, and K. Walton. *Time-Delay Systems. Stability and Performance Criteria with Applications.* Ellis Horwood, New York, 1992.

[106] O. Mason and R. Shorten. The geometry of convex cones associated with the Lyapunov inequality and the common Lyapunov function problem. *Electron. J. Linear Algebra*, 12:42–63, 2004/2005.

[107] A. P. Molchanov and E. S. Pyatnitskij. Stability criteria for selector-linear differential inclusions. *Soviet. Math. Dokl.*, 36(3):421–424, 1988.

[108] C. B. Moler and C. F. van Loan. Nineteen dubious ways to compute the exponential of a matrix – twenty-five years later. *SIAM Review*, 45(1):3–49, 2003.

[109] K. L. Moore and S. Bhattacharyya. A technique for choosing zero locations for minimal overshoot. *IEEE Trans. Automat. Contr.*, 35(5):577–580, 1990.

[110] S.-I. Niculescu. *Delay Effects on Stability. A Robust Control Approach*, volume 269 of *Lecture Notes in Control and Information Sciences.* Springer-Verlag, London, 2001.

[111] A. Y. Obolenskiĭ. Extremal Lyapunov functions for linear systems with constant coefficients. *Mat. Fiz.*, 34:26–30, 1983.

[112] V. M. P. Benner and H. Xu. A numerically stable, structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils. *Numer. Math.*, 78(3):329–358, 1998.

[113] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer-Verlag, New York, 1983.

[114] I. R. Peterson. A stabilization algorithm for a class of uncertain linear systems. *Syst. Control Lett.*, 8:351–357, 1987.

[115] E. Plischke and F. Wirth. Stabilization of linear systems with prescribed transient bounds. In *Proc. MTNS-2004*, Leuven, Belgium, 2004. Paper no. 181 (CDROM).

[116] A. Polański. On absolute stability analysis by polyhedral Lyapunov functions. *Automatica*, 36(4):573–578, 2000.

[117] A. J. Pritchard. Transitory behaviour of uncertain systems. In F. Colonius et al., editors, *Advances in Mathematical Systems Theory*, pages 1–18, Birkhäuser, Boston, 2000.

[118] J. Prüss. On the spectrum of C_0-semigroups. *Trans. Amer. Math. Soc.*, 284(2):847–857, 1984.

[119] L. Qiu and E. J. Davison. Performance limitations of non-minimum phase systems in the servomechanism problem. *Automatica*, 29(2):337–349, 1993.

[120] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[121] A. Saberi, A. A. Stoorvogel, and P. Sannuti. *Control of Linear Systems with Regulation and Input Constraints.* Communications and Control Engineering Series. Springer-Verlag, London, 2000.

[122] D. Salamon. *Control and Observation of Neutral Systems.* Number 91 in Research Notes in Mathematics. Pitman, London, 1984.

[123] R. Sarybekov. Extremal quadratic Lyapunov functions of systems of second-order equations. *Sib. Math. J.*, 18:823–829, 1978.

[124] J.-P. Serre. *Lie Algebras and Lie Groups*, volume 1500 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2nd edition, 1992.

[125] R. N. Shorten and K. S. Narendra. On common quadratic Lyapunov functions for pairs of stable LTI systems whose system matrices are in companion form. *IEEE Trans. Automat. Contr.*, 48(4):618–621, 2003.

[126] G. V. Smirnov. *Introduction to the Theory of Differential Inclusions.* American Mathematical Society, Providence, RI, 2002.

[127] R. A. Smith. Bounds for quadratic Lyapunov functions. *J. Math. Anal. Appl.*, 12:425–435, 1965.

[128] E. D. Sontag. *Mathematical Control Theory, Deterministic Finite Dimensional Systems.* Springer-Verlag, New York, 2nd edition, 1998.

[129] M. N. Spijker. On a conjecture by LeVeque and Trefethen related to the Kreiss matrix theorem. *BIT*, 31:551–555, 1991.

[130] M. N. Spijker, S. Tracogna, and B. D. Welfert. About the sharpness of the stability estimates in the Kreiss matrix theorem. *Math. Comp.*, 72(242):697–713, 2002.

[131] G. Stépán. *Retarded Dynamical Systems: Stability and Characteristic Functions*, volume 210 of *Pitman Research Notes in Mathematics.* Longman Scientific & Technical, Harlow, UK, 1989.

[132] G. W. Stewart. *Matrix Algorithms. Vol. II: Eigensystems.* SIAM Publications, Philadelphia, PA, 2001.

[133] J. Stoer and C. Witzgall. *Convexity and Optimization in Finite Dimensions.* Springer-Verlag, Berlin, 1970.

[134] T. Ström. On logarithmic norms. *SIAM J. Numer. Anal.*, 12(5):741–753, 1975.

[135] E. C. Titchmarsh. *Introduction to the Theory of Fourier Integrals.* Chelsea, New York, 3rd edition, 1986.

[136] L. N. Trefethen. Pseudospectra of matrices. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1991*, pages 234–266, Harlow, UK, 1992. Longman Scientific & Technical.

[137] L. N. Trefethen. Pseudospectra of linear operators. *SIAM Review*, 39(3):383–406, 1997.

[138] C. van Loan. The sensitivity of the matrix exponential. *SIAM J. Numer. Anal.*, 14(6):971–981, 1977.

[139] C. van Loan. A symplectic method for approximating all the eigenvalues of a Hamiltonian matrix. *Lin. Alg. Appl.*, 61:233–251, 1984.

[140] R. S. Varga. *Matrix Iterative Analysis*, volume 27 of *Springer Series in Computational Mathematics.* Springer-Verlag, Berlin, 2nd edition, 2000.

[141] K. Veselić. Bounds for exponentially stable semigroups. *Lin. Alg. Appl.*, 358:195–217, 2003.

[142] M. Vidyasagar. On matrix measures and convex Liapunov functions. *J. Math. Anal. Appl.*, 62:90–103, 1978.

[143] M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice Hall, Englewood Cliffs, NJ, 2nd edition, 1993.

[144] R. Vinter. *Optimal control.* Systems and Control: Foundations and Applications. Birkhäuser, Boston, MA, 2000.

[145] T. Ważewski. Sur la limitation des intégrales des systèmes d'équations différentielles linéaires ordinaires. *Stud. Math.*, 10:48–59, 1948.

[146] E. Wegert and L. N. Trefethen. From the Buffon needle problem to the Kreiss matrix theorem. *Amer. Math. Monthly*, 101:132–139, 1994.

[147] D. Werner. *Funktionalanalysis*. Springer-Verlag, Berlin, 5th edition, 2005.

[148] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.

[149] J. C. Willems. Lyapunov functions for diagonally dominant systems. *Automatica*, 12:519–523, 1976.

[150] F. Wirth. *Stability theory of perturbed systems: Joint spectral radii and stability radii*. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2005. To appear.

[151] K. Yosida. *Functional Analysis*. Springer-Verlag, Berlin, 6th edition, 1980.

[152] J. Zabczyk. A note on C_0-semigroups. *Bull. Acad. Pol. Sci. Ser. Sci. Mat.*, 23(6):895–898, 1975.

Index