University of Twente

Applied Mathematics
Chair of Hybrid Systems

Master Thesis

On the Stabilization of Linear Time Invariant Systems with POSITIVE CONTROLS

Author
F.L.H. Klein Schaarsberg

Supervisor
Prof. Dr. A.A. Stoorvogel

Graduation Committee
Prof. Dr. A.A. Stoorvogel
Prof. Dr. H.J. Zwart
Dr. M. Schlottbom

April 26, 2019

Preface

During the past seven months I have been working on this project with the objective of concluding my study in Applied Mathematics at the University of Twente. It closes off my almost six year time as a student at this university which I, after a difficult start, embraced as my home.

My time as a student overall has been the most enjoyable and memorable period of my life. I believe that I have developed myself academically, but more importantly that I developed myself to become the person that I am today. As excited as I am to move on to graduate life, I can not help but to think I will miss all of the things I have done and experienced in university. I have met some amazing people, made new friends with whom I have built many good memories, and I have done many things I am proud of, looking back.

I would like to take this opportunity to express my gratitude to a couple of people. First of all I would like to thank my parents for making it possible for me to study, and for supporting me in everything I did, even though sometimes they may have had no clue what that was exactly. Special thanks go to my brother for the momorable moments we shared in the past years, to Juliët for being there for me in all times, and to Sagy, David, Sander, Rico and Mike for their mental and social support.

Needless to say I would like to thank Anton, Hans and Matthias for their time spendt reading and evaluating this work. In particular, I would like to thank Anton for supervising me throughout this lengthy project. Thank you for your time, effort and patience invested in me during our many meetings and beyond while reading my work. This project and process has not been easy on me, but most of the times our meetings brought back some energy in me. To take up on that, I would like to thank Lilian Spijker for the various meetings we had during times of doubt and frustration in the past year.

Summary

This report studies the stabilization of control systems for which the control input has a positivity constraint, that is, one or more of the control inputs is either positive or zero. Typical examples of such systems are systems for which the control can only *add* (for example energy or substances), but cannot *extract*. A classic example is that of a heating system in buildings. Radiators can only release heat into the rooms, but it cannot extract the heat when the temperature overshoots the setpoint.

This report is build up out of three main parts. The first part concerns a review on mathematical research on systems with positive control, in particular for *linear time invariant* systems. Such systems may be represented by the state-space representation

$$\dot{x} = Ax + Bu,$$

with state x and control input u. Two representative papers are studied in more detail. One approaches the positive control problem from the perspective of *optimal control*. The other investigates positive state feedback stabilization of linear time invariant systems with at most one pair of unstable complex conjugate poles. This latter paper forms the basis for this project.

The second and third part focus on linear time invariant systems in general. It extends known results to systems with more than one pair of unstable complex conjugate poles, where the positive control input is scalar. Two approaches are considered.

The first approach uses Lyapunov's stability theory as a base for a asymptotically stabilizing positive control law. Formal proofs of stability are given for stable (but not asymptotically) oscillatory systems. The feasibility of the control law for unstable oscillatory systems is investigated through simulations.

The second approach concerns techniques from singular perturbations for ordinary differential equations. The viability of the application of known techniques to the positive control problem is investigated and substantiated with various simulations.

Contents

Nomenclature

The next list describes several symbols that will be later used throughout this thesis.

$\mathbb{N}$	The set of all natural numbers: $\mathbb{N} = \{0, 1, 2, 3, \dots\}$
$\mathbb{N}_{>0}$	The set of all natural numbers except 0: $\mathbb{N} \setminus \{0\}$
$\mathbb{Z}$	The set of all integers (whether positive, negative or zero): $\mathbb{Z} = \{\dots, 2, 1, 0, 1, 2, \dots\}$
$\mathbb{Q}$	The set of all rational numbers: $\mathbb{Q} = \{\frac{a}{b} \vert a, b \in \mathbb{Z}, b \neq 0\}$
$\mathbb{R}$	The set of all real numbers
$\mathbb{R}^n$	The n-dimensional ($n \in \mathbb{N}_+$) real vector space over the field of real numbers
$\mathbb{R}^{n\times m}$	The $n \times m$-dimensional ($n, m \in \mathbb{N}_+$) real matrix space over the field of real numbers
$\mathbb{R}_{\geq 0}$, $\mathbb{R}_{>0}$	The set of non-negative real numbers $\mathbb{R}_{\geq 0} = \{\mu \in \mathbb{R} \vert \mu \geq 0\}$, and similarly $\mathbb{R}_{>0} = \{\mu \in \mathbb{R} \vert \mu > 0\}$
$\mathbb{R}^n_{\geq 0}$	The nonnegative orphant in $\mathbb{R}^n$: $\mathbb{R}^n_{\geq 0} := \{\mu \in \mathbb{R}^n \vert \mu_i \geq 0\}$
$\imath$	Imaginary unit $\imath := \sqrt{-1}$
$\mathbb{C}$	The set of all complex numbers: $\mathbb{C} = \{a + b\imath \vert a, b \in \mathbb{R}\}$
t	General variable for time
τ	Alternative variable for time
t_0	Starting time
t_f	Final time
T	Fixed time span
$\mathcal{T}$	Fixed time span (alternative)
$x(t)$	State vector in $\mathbb{R}^n$ at time t
$x_i(t)$	i^{th} entry of $x(t)$
x_0	Initial state $x_0 := x(t_0)$
x_f	Final state $x_f = x(t_f)$
$\bar{x}$	Equilibrium state in $\mathbb{R}^n$
$\dot{x}(t)$	State vector time derivative of dimension n at time t
$\theta(t)$	Angle in radians at time t
$\dot{\theta}(t)$, $\ddot{\theta}(t)$	First and second time derivative of θ at time t
$u(t)$	Control input vector of dimension p at time t
A	System matrix ($A \in \mathbb{R}^{n\times n}$)
B	Input matrix ($B \in \mathbb{R}^{n\times p}$)
$\mathcal{C}_{\{A,B\}}$	The controllability matrix of the pair (A, B), as defined in Theorem 1

$\mathcal{O}_{\{A,C\}}$	The observability matrix of the pair (A, C), as defined in Theorem 2
$V(x(t))$	Lyapunov function
$\dot{V}(x(t))$	Derivative of Lyapunov function with respect to time t
M^{T}	Transpose of matrix M, also applies to vectors
$\mathrm{rank}(M)$	The rank of matrix M.
$E_\lambda(M)$	The set of eigenvalues of matrix M
$I_{n\times n}$	The $n \times n$ identity matrix. The $n \times n$ subscript is sometimes omitted if the size is evident
P, Q	Positive definite matrices
$L_p[a, b]$	The space of functions $f(s)$ on the interval $[a, b]$ for which $\lvert\lvert f\rvert\rvert_p = \int_a^b \lvert f(s)\rvert^p \,\mathrm{d}s < \infty$. Commonly $a = t_0$ and $b = \infty$ such that one considers $L_p[t_0, \infty)$
$\lvert\lvert\cdot\rvert\rvert_p$	Standard p-norm
$\langle\cdot,\cdot\rangle$	Inner product
$\mathrm{lcm}(a, b)$	The least common multiple of numbers a and b.

1. Introduction to the positive control problem

There are many real life examples of systems for which the control input is nonnegative, also known as one-sided. A typical example most people can relate to is that of heating systems in houses or other buildings. A central heating system essentially provides warmth to the interior of a building. A conventional system circulates hot water through radiators which release heat into rooms. These radiators are often equiped with thermostatically controlled valves. These valves can either be opened (fully or to some level) or be closed. Therefore the radiator can only release energy into the room (valve open), or add no energy at all (valve closed). It cannot extract heat from the room, and hence the heating system is categorized as a system with positive control.

Examples of positive control systems that are closely related to the central heating example are systems which involve flow of fluids or gasses that are controlled by one-way valves. Such systems occur for example in mechanics and chemistry. Consider for instance a chemical plant where several substances are to be added to a reaction vessel. Once added, substances mix and/or react to produce the desired end product, and hence the raw substances cannot be extracted individually anymore. Hence, if u_i controls the addition of substance i, then this is a positive control since extraction is not possible.

These control systems also occur in medical fields. For example in pumps that intravenously administer medication to a patient's bloodstream. An artificial pancreas is an example of a positive control system since insulin can only be added to a patient's body, and cannot be extracted via the pump. Many other control systems can be listed which can be categorized as positive control. For example economic systems with nonnegative investements, or in electrical networks with diode elements (which have low resistance in one direction and high, ideally infinite, resistance in the other). Also the classical cruise control is an example of positive control since the control system only operates the throttle and not the brakes.[1]

The control systems as described in the above examples are commonly represented by a system of differential equations. Stabilization of such systems has been studied extensively in the field of control theory. Consider the control system of n coupled differential equations with p control inputs given by

$$\dot{x}(t) = f(x(t), u(t), t). \tag{1}$$

Here $x(t) \in \mathbb{R}^n$ denotes the n-dimensional *state vector* of the system at time t, its derivative with respect to time at time t is denoted by $\dot{x}(t)$, which is also an n-dimensional column vector. At time t the *control* (or input) vector is denoted by $u(t) \in \mathbb{R}^p$. If the control restraint set is denoted by $\Omega \subseteq \mathbb{R}^p$, then $u : [t_0, \infty) \to \Omega$ for some initial time t_0. Note that there may be a final time $t_f < \infty$ such that $u : [t_0, t_f] \to \Omega$. In the most general case the control input is unrestricted, in that case $\Omega = \mathbb{R}^p$. Regularity conditions should be imposed on the function $f : \mathbb{R}^n \to \mathbb{R}^n$ to ensure that (1) has a unique solution $x(t)$, $t \geq t_0$, for every initial state $x_0 \in \mathbb{R}^n$ and every control input $u(t)$.

The control problem (1) is called a *positive control problem* if one of more of the control inputs are constrained to being nonnegative. In order to formally state the positive control problem, define

$$\mathbb{R}^m_{\geq 0} := \{\mu \in \mathbb{R}^m | \mu_i \geq 0\}.$$

as the nonnegative orphant in $\mathbb{R}^m$. The control problem (1) is said to be a positive control problem if there exists a nonempty subset of indices $\mathcal{P} \subseteq \{1, 2, \ldots, p\}$ of cardinality m, $1 \leq m \leq p$, such that

$$u_{\mathcal{P}}[t_0, \infty) \to \Omega_{\geq 0},$$

[1]Assuming that the cruise controller is not configured to slow the car actively via the brakes or the gearbox.

where $\Omega_{\geq 0}$ is an m dimensional subspace of $\mathbb{R}^m_{\geq 0}$, that is $\Omega_{\geq 0} \subseteq \mathbb{R}^m_{\geq 0}$. It is (often) assumed that the m-dimensional zero vector $\vec{0}$ is an element of $\Omega_{\geq 0}$, so that the zero control is in the restraint set $\Omega_{\geq 0}$. In the least restrictive case $\Omega_{\geq 0} = \mathbb{R}^m_{\geq 0}$.

The system (1) is the most general representation of a control system of differential equations. No assumptions concerning linearity of time invariancy are made there. Control problems are more than often concerned with linear time invariant (LTI) systems. In many cases non-linear systems are approximated by linear systems, as in the example of Section 1.3. As will be apparent in Section 2, most research into positive control systems concerns LTI systems. Hence, also this report focusses on the stabilization of linear time invariant (LTI) systems of differential equations using positive control. In that case Equation (1) is of the form

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathbb{R}^p, \tag{2}$$

where matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times p}$ are called the *system matrix* and the *input matrix* respectively. In some instances the system's *output vector* $y(t) \in \mathbb{R}^q$ output is defined by

$$y(t) = Cx(t) + Du(t), \quad C \in \mathbb{R}^{q \times n}, \quad D \in \mathbb{R}^{q \times p}, \tag{3}$$

with *output matrix* C and *feedthrough matrix* D. Equations (2) and (3) are called a 'state-space representation' for an LTI system.

1.1. Structure or this report

The previous section aimed to introduce the concept of 'positive control'. The report is structured as follows: The remainder of this chapter includes some preliminary definitions in Section 1.2 followed the classic example of the simple pendulum with positive controls in Section 1.3. Section 2 concerns a study into literature of mathematical research into systems with positive controls. That section includes a more detailed review of two notable works. The main problem that is considered in this report is introduced in Section 3. It includes a motivation for the two approaches individually described in the sections to follow. Section 4 approaches the positive control problem from Lyapunov's stability theory, Section 5 approaches the problem with techniques of singular perturbations for ordinary differential equations. The report closes with a conclusion in Section 6 and a discussion and recommendations in Section 7.

1.2. Preliminary definitions

This section aims to introduce some well know concepts that will be used in one way or another throughout this report. Consider the system $\dot{x} = f(x)$, an equilibrium of such system is defined as follows:

Definition 1 (Equilibrium) $\bar{x} \in \mathbb{R}^n$ is an *equilibrium (point)* of (12) if $f(\bar{x}) = 0$. ◇

An equilibrium point is also known as a 'stationary point', a 'critical point', a 'singular point', or a 'rest state'. A basic result from linear algebra is that $Ax = 0$ has only the trivial solution $x = 0$ if and only if A is nonsingular (i.e. A has an inverse). So for LTI systems $\dot{x} = Ax$, $\bar{x} = 0$ is the only equilibrium if and only if A is nonsingular. The aim of applying control u is to stabilize the system at $\bar{x} = 0$. This means that $\dot{x} = Ax + Bu = 0$ if and only of $Ax = -Bu$. Assuming that A is non-singular one finds that $0 = \bar{x} = A^{-1}B\bar{u}$, which implies that $\bar{u} = 0$ if and only if $A^{-1}B$ is injective.

Equilibria as defined in Definition 1 can be categorized as follows:

Definition 2 (Stable and unstable equilibria) An equilibrium point $\bar{x}$ is

1. *stable* if $\forall \varepsilon > 0\ \exists \delta > 0$ such that $||x_0 - \bar{x}|| < \delta$ implies $||x(t; x_0) - \bar{x}|| < \varepsilon\ \forall t > t_0$.
2. *attractive* if $\exists \delta_1 > 0$ such that $||x_0 - \bar{x}|| < \delta_1$ implies that $\lim_{t\to\infty} x(t; x_0) = \bar{x}$.
3. *asymptotically stable* if it is stable and attractive.
4. *globally attractive* if $\lim_{t\to\infty} x(t; x_0) = \bar{x}$ for every $x_0 \in \mathbb{R}^n$.
5. *globally asymptotically stable* if it is stable and globally attractive.
6. *unstable* if $\bar{x}$ is not stable. This means that $\exists \varepsilon > 0$ such that $\forall \delta > 0$ there exist x_0 and a t_1 such that $||x_0 - \bar{x}|| < \delta$ but $||x(t_1; x_0) - \bar{x}|| \geq \varepsilon$.

◇

A similar categorization can be made for systems of the form $\dot{x} = Ax$. The poles of such system, given by the eigenvalues of the system matrix A, determine the categorization of stability of the system. If the poles of the system are given by $E_\lambda(A)$, then the system is called

- *stable* if all poles have their real part smaller or equal to zero (in the case where the real part equals zero, then the imaginary part cannot equal zero);
- *asymptotically stable* if all poles have their real part strictly smaller than zero;
- *unstable* if one or more of the poles have real part greater than zero.

Systems for which poles have nonzero imaginary part are sometimes called *oscillatory*. Note that the *poles* of the system $\dot{x} = Ax$ are equal to the *eigenvalues* of the matrix A. The terms 'poles' and 'eigenvalues' are used side by side. Note that systems with zero real part are categorized as 'stable', but explicitly not 'asymptotically'.

For state-space representation (2) and (3) representations the concepts of *controllability* and *observability* are introduced as follows.

Theorem 1 (Controllability matrix) *Consider the system $\dot{x} = Ax + Bu$, with state vector $x \in \mathbb{R}^{n\times 1}$, input vector $u \in \mathbb{R}^{r\times 1}$, state matrix $A \in \mathbb{R}^{n\times n}$ and input matrix $B \in \mathbb{R}^{n\times r}$. The $n \times nr$* controllability matrix *if given by*

$$\mathcal{C}_{\{A,B\}} = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}. \tag{4}$$

The system is controllable if the controllability matrix has full row rank, that is $\text{rank}(\mathcal{C}_{\{A,B\}}) = n$.

Roughly speaking, controllability describes the ability of an external input (the vector of control variables u) to move the internal state $x(t)$ of a system from any initial state x_0 to any other final state x_{t_f} in a finite time interval. More formally, the system Equation (2) is called controllable if, for each x_1, $x_2 \in \mathbb{R}^n$ there exists a bounded admissible control $u(t) \in \Omega$, defined in some interval $t_1 \leq t \leq t_2$, which steers $x_1 = x(t_1)$ to $x_2 = x(t_2)$.

A slightly weaker notion than controllability is that of stabilizability. A system is said to be stabilizable when all uncontrollable state variables can be made to have stable dynamics.

Theorem 2 (Observability matrix) *Consider the system $\dot{x} = Ax + Bu$ with output $y = Cx + Du$, with state vector $x \in \mathbb{R}^{n\times 1}$, input vector $u \in \mathbb{R}^{r\times 1}$, state matrix $A \in \mathbb{R}^{n\times n}$, input matrix $B \in \mathbb{R}^{n\times r}$, output matrix $C \in \mathbb{R}^{q\times n}$ and feedthrough matrix $D \in \mathbb{R}^{q\times p}$. The $qn \times n$* observability matrix *if given by*

$$\mathcal{O}_{\{A,C\}} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}. \tag{5}$$

The system is observable if the observability matrix has full column rank, that is $\mathrm{rank}(\mathcal{O}_{\{A,C\}}) = p$.

Observability means that one can determine the behavior of the entire system from the system's outputs.

Some of the literature of Section 2 use the concept of *local null-controllability*. A system is called locally null-controllable if there exists a neighbourhood $\mathcal{V}$ of the origin such that every point of $\mathcal{V}$ can be steered to $x = 0$ in finite time.

1.3. An illustrative example: the simple pendulum

To illustrate the positive control problem, consider the specific case in which the positive control input is a linear state feedback. Consider the linear system (A, B) given by Equation (2). The state feedback control input is computed as $u(t) = Fx(t)$ for some real matrix $F \in \mathbb{R}^{1\times n}$ such that the scalar positive control input is computed as $u(t) = \max\{0, Fx(t)\}$.

A classic and illustrative example is that of the so-called 'simple pendulum' as depicted in Figure 1. The rod of the pendulum has length l. At the end of the rod a weight of mass m is attached, the rod itself is assumed to be weightless. The gravitational force on the weight is equal to mg, where g is the gravitational constant. Its direction is parallel to the vertical axis. The angle between the pendulum and the vertical axis is denoted by θ radians. Positive angle is defined as the pendulum being to the right of the center. A horizontal force of magnitude u can be exerted on the weight. On the horizontal axis, denote positive forces as forces pointed to the right.

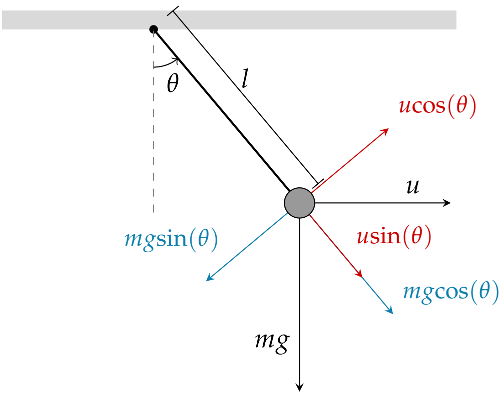

Figure 1: Simple pendulum of length l of mass m with gravitational force mg and external horizontal force u.

As in Figure 1, both forces can be decomposed into components in the direction of motion (perpendicular to the pendulum) and perpendicular to the direction of motion. The former are the forces of interest, since the latter forces are cancelled by an opposite force exerted by the rod. The netto force on the mass in the direction of movement is equal to

$u\cos(\theta) + mg\sin(\theta)$. The equation of motion for a pendulum with friction coefficient k is given by

$$ml\ddot{\theta} + kl\dot{\theta} + mg\sin(\theta) + u\cos(\theta) = 0.$$

Equivalently, one can write

$$\ddot{\theta} = -\frac{k}{m}x_2 - \frac{g}{l}\sin(x_1) + \frac{1}{ml}u\cos(x_1).$$

Now define states $x_1 := \theta$ and $x_2 := \dot{\theta}$. That way the system of differential equations is obtained described by

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{k}{m}\dot{\theta} - \frac{g}{l}\sin(\theta) - \frac{1}{ml}u\cos(\theta). \end{aligned} \tag{6}$$

This model obviously is non-linear. The equilibrium of interest is $x = \begin{bmatrix} 0 & 0 \end{bmatrix}^{\mathrm{T}}$, $u = 0$, where the pendulum hangs vertically. When the system is linearized around this equilibrium the following linear state-space matrices are acquired:

$$A = \left.\begin{bmatrix} 0 & 1 \\ -\frac{g}{l}\cos(x_1) - \frac{1}{ml}u\sin(x_1) & -\frac{k}{m} \end{bmatrix}\right|_{\substack{x=\vec{0} \\ u=0}} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{l} & -\frac{k}{m} \end{bmatrix}, \qquad B = \left.\begin{bmatrix} 0 \\ -\frac{1}{ml}\cos(x_1) \end{bmatrix}\right|_{\substack{x=\vec{0} \\ u=0}} = \begin{bmatrix} 0 \\ -\frac{1}{ml} \end{bmatrix}.$$

That way the system Equation (6) linearized around $x = \begin{bmatrix} 0 & 0 \end{bmatrix}^{\mathrm{T}}$, $u = 0$ is given by

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{l} & -\frac{k}{m} \end{bmatrix} x + \begin{bmatrix} 0 \\ -\frac{1}{ml} \end{bmatrix} u$$

For now, consider a frictionless pendulum, so $k = 0$. Let for this pendulum also $\frac{g}{l} = 1$ and $m = \frac{1}{l}$. The linearized system is then described by

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 0 \\ -1 \end{bmatrix}}_{B} u.$$

A close inspection of A reveals that its eigenvalues are equal to $\pm\imath$, which indeed illustrates the frictionless property of the pendulum. Let the control input be defined as a stabilizing state feedback control $u = Fx$, which is proportional with the direction and angular speed of motion x_2. For this example let $F = \begin{bmatrix} 0 & 1 \end{bmatrix}$. It can be verified that the eigenvalues of $(A + BF)$ are equal to $-\frac{1}{2} \pm \frac{1}{2}\sqrt{3}\imath$, and hence have strictly negative real part such that the controlled system $\dot{x}(t) = (A + BF)x(t)$ is stable. Vice versa, given the desired poles of $(A + BF)$, the feedback matrix F can for example be computed via Ackermann's pole placement formula, see Theorem B.1.

For the positive control case considered here, a force Fx is exerted on the pendulum whenever it moves against the direction of force. Now let $\tilde{u}$ be a positive state feedback control, defined as $\tilde{u}(t) = \max\{0, Fx(t)\} = \max\{0, x_2(t)\}$. This way, the system's dynamics (2) switches between $\dot{x}(t) = Ax(t)$ and $\dot{x}(t) = (A + BF)x(t)$, based on the sign of $\tilde{u}(t)$, that is based on the direction of movement of the pendulum. The expectation is that both the regular and positive control laws stabilize the system. This is trivial for the regular control input u due to the stable poles of the controlled system. For the positive control $\tilde{u}$, it tries to stabilize the system whenever $\tilde{u} > 0$. If $\tilde{u} = 0$ then the pendulum swings freely according to $\dot{x} = Ax$ (where the amplitude of swing does not increase nor decrease), until

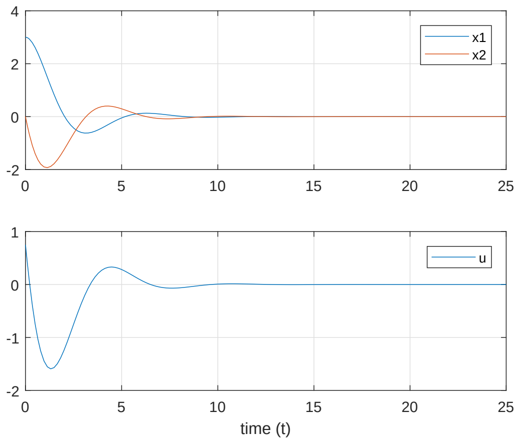

Figure 2: Simulation of a single simple pendulum with regular state feedback control $u(t)$.

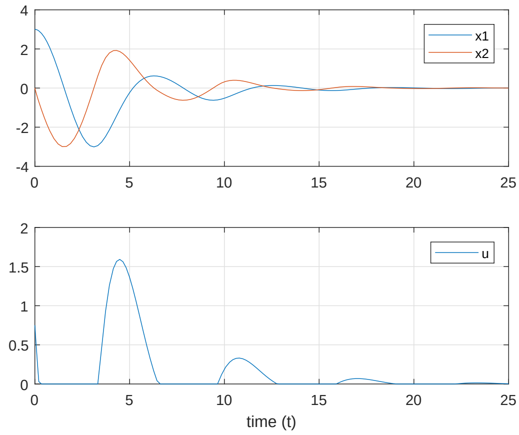

Figure 3: Simulation of single simple pendulum with positive state feedback control $\tilde{u}(t)$.

$\tilde{u}$ becomes positive again. Therefore, one would expect that also $\tilde{u}$ stabilizes the system, but slower than u does. This intuition is supported by a simulation of both systems. The results are included in Figure 2 for the regular control u, and in Figure 3 for the pendulum with positive control $\tilde{u}$.

In this example a hanging frictionless pendulum was stabilized, which is a stable system by itself (not asymptotically stable). For this system the positive state feedback control law worked in order to stabilize the system. For unstable systems, this may be a different story. Consider for example a pendulum that is supposed to be stabilized around its vertical upright position, where the positive control can only push the pendulum one way. It is a known result that (a well designed) 'regular' state feedback control can stabilize the pendulum, whereas for positive control if the pendulum is pushed too far, it needs to make a full swing in order to come back to the position where u becomes positive again, and therefore it is not stable. This illustrates the pitfalls and shortcomings of the positive control problem.

2. Overview of literature

LTI systems with positive control, such as Equation (2), have been subject of mathematical control research since the early 1970s. Early works include Saperstone and Yorke [26], Brammer [2] and Evans and Murthy [6]. The paper by Saperstone and Yorke [26] forms (as it seems) the basis of the mathematical research into the positive control problem. It concerns the n-dimensional system (2) for which the control is scalar, i.e. $p = 1$, where the control restraint set is restricted to $\Omega = [0, 1]$. The null control is an extreme point of the control restraint set Ω explicitly. It is assumed that $u(\cdot)$ belongs to the set of all bounded measurable functions $u : \mathbb{R}_+ \to \Omega$. The main result is reads the system (2) is locally controllable at the origin $x = 0$ if and only if (i) all eigenvalues of A have nonzero imaginary parts, and (ii) the controllability matrix (see Equation (4)) of the pair (A, B) has full rank. Note that if n is odd, then A must have at least one real eigenvalue. Hence according to Saperstone and Yorke [26] a system of odd dimension is not positively controllable. Saperstone and Yorke [26] extend the result to vector valued control, i.e. $p > 1$. It proves that $x = 0$ is reachable in finite time with $\Omega = \prod_{i=1}^{p}[0, 1]$ under the same assumptions on the eigenvalues of A and rank of the controllability matrix of (A, B). It should be mentioned that the paper only states conditions under which an LTI system is positively stabilizable, it does not mention any control law. A follow-up paper from the same author Saperstone [25] considers global controllability (in contrast to local controllability) of such systems. Perry and Gunderson [21] applied the results of Saperstone and Yorke [26] to systems with nearly nonnegative matrices.

Brammer [2] extends the results presented in Saperstone and Yorke [26]. It assumes that the control is not necessarily scalar, that is the control $u \in \Omega \subset \mathbb{R}^m$, without the assumption that the origin in $\mathbb{R}^m$ is interior to Ω. The main result of the paper is that if the set Ω contains a vector in the kernel of B (i.e., there exists $u \in \Omega$ such that $Bu = 0$) and the convex hull[2] Ω has a nonempty interior in $\mathbb{R}^m$, that then the conditions (i) the controllability matrix of the pair (A, B) has rank n, and (ii) there is no real eigenvector v of A^{T} satisfying $\langle v, Bu \rangle \leq 0$ for all $u \in \Omega$ are necessary and sufficient for the null-controllability of (2).

Where [2, 25, 26] are concerned with continuous time LTI systems, Perry and Gunderson [21] is the first paper to study the controllability of discrete-time linear systems with positive controls of the form $x_{k+1} = Ax_k + Bu_k$, for $k = 0, 1, \ldots$. It provides necessary and sufficient conditions for controllability for single-input systems (so $p = 1$), for the case where $u_k \in [0, \infty)$. The main result states that the system is completely controllable if and only if (i) the controllability matrix has full rank; (ii) A has no real eigenvalues $\lambda > 0$. The result for the discrete-time system is different from the continuous time system (for scalar input), as real eigenvalues of A are not allowed in the latter case. More recent work on discrete-time system with positive control is Benvenuti and Farina [1], which investigates the geometrical properties of the reachability set of states of a single input LTI system.

The positive control problem has also been approached from the field of optimal control. The problem aproached from this was introduced in Pachter [20], where the linear-quadratic optimal control problem with positive controls is considered. It provides conditions for the existende of a solution to the optimal control problem, in the explicit case where the trajectory-dependent term in the integrand of the cost function is not present. Another more recent notable example is Heemels et al. [12], a review of which is included in Section 2.1. This paper can be considered an extension of the work by Pachter [20].

A mechanical application is provided in Willems et al. [29] which applies positive state feedback to stabilize surge of a centrifugal compressor. It provides conditions on the poleplacement of $A + BF$ with the positive feedback system $u(t) = \max\{0, Fx(t)\}$ for a

[2] The convex hull (of convex closure) of a set X is the smallest convex set that contains X.

given system (A, B) for which A has at most one pair of non-stable complex conjugate eigenvalues. This paper forms the basis for this report, hence a review is included in Section 2.2.

More recent work into the controllability of linear systems with positive control includes the work in Frias et al. [8]. This paper considers a problem similar to Brammer [2], but for the case when the matrix A has only real eigenvalues. By imposing conditions on B rather than on A, and on the number of control inputs. Yoshida and Tanaka [30] provides a test for positive controllability for subsystems with real eigenvalues which involves the Jordan canonical form.

Other works include Camlibel et al. [3]; Respondek [23] on controllability with non-negative constrains; Heemels and Camlibel [11] with additional state constraints on the linear system; Leyva and Solis-Daun [16] on bounded control feedback $b_l \leq u(t) \leq b_u$ with the case included when $b_l = 0$; Grognard [9], Grognard et al. [10] with research into global stabilization of predator-prey systems[3] in biological control applications.

Closely related to positive control systems are constrained control systems, such as described in Son and Thuan [27], and control systems with saturating inputs as for example described in Corradini et al. [4]. Moreover, in the literature positive control should not be confused with

(i) Positive (linear) systems. These are classes of (linear) systems for which the state variables are nonnegative, given a positive initial state. These systems occur applications where the stares represent physical quantities with positive sign such as concentrations, levels, etc. Positive linear systems are for example described in Farina and Rinaldi [7] and de Leenheer and Aeyels [5]. Positive systems can also be linked to systems with bounded controls, such as in Rami and Tadeo [22, section 5a] where also the control input is positive.

(ii) Positive feedback processes. That is, processes that occurs in a feedback loop in which the effects of a small disturbance on a system include an increase in the magnitude of the perturbation.) Zuckerman et al. [32, page 42]. A classic example is that of a microphone that is too close to its loudspeaker. Feedback occurs when the sound from the speakers makes it back into the microphone and is re-amplified and sent through the speakers again resulting in a howling sound.

(iii) 'Positive controls' which are used to assess test validity of new biological tests compared to older test methods.

2.1. Stabilizing positive control law via optimal control

As mentioned earlier an approach for the positive control problem via optimal control theory is presented in Heemels et al. [12]. Here the existence of a stabilizing positive control is proven via the Linear Quadratic Regulator (LQR) problem. Two approaches are described: Pontryagin's maximum principle and dynamic programming. The maximum principle is extended to the infinite horizon case. At some early stage of this project, the paper was studied in extensive detail. At that time also a rather detailed summary was written. This summary is included in Appendix D.

[3] So called 'predator-prey equations' are a pair of first-order nonlinear differential equations, frequently used to describe the dynamics of biological systems in which two species interact, one as a predator and the other as prey. The populations of prey (x) and predator (y) change in time according to $\frac{dx}{dt} = \alpha x - \beta xy$, $\frac{dy}{dt} \delta xy - \gamma y$ for some parameters $\alpha, \beta, \gamma, \delta$.

2.2. Positive state feedback for surge stabilization

Positive state feedback was already introduced in the example of the frictionless pendulum in Section 1.3. Another application of positive state feedback is described by Willems et al. [29]. Here, positive feedback is used to stabilize surge in a centrifugal compressor. More specifically, the goal is to control the aerodynamic flow within the compressor. According to Willems et al. [29] the aerodynamic flow instability can lead to severe damage of the machine, and restricts its performance and efficiency. The positivity of the control is relevant as follows: The surge, which is a disruption of the flow through the compressor, is controlled by a control valve that is fully closed in the desired operating point and only opens to stabilize the system around this point. The valve regulated some flow of compressed air into the compressor, which is supposed to stabilize the surge. The positive feedback controller used by Willems et al. [29] is based on the pole placement technique. The feedback applied in this paper is simple and easily implementable. The paper considers the system (2) where $u(t) \in \mathbb{R}$ is a scalar control input, so $p = 1$. The control restraint set Ω is chosen equal to $\mathbb{R}_+ := [0, \infty)$. The state feedback control is computed as the maximum of a linear state feedback Fx and 0. That is, $u(t) = \max\{0, Fx(t)\}$. The input functions are assumed to belong to the Lebesgue space L_2 of measurable, square integrable functions on $\mathbb{R}_+$. One important aspect is that the paper considers the case where A has at most one pair of unstable, complex conjugate eigenvalues.

Note that (A, B) is positively feedback stabilizable if there exists a row vector F such that all solution trajectories of $\dot{x}(t) = Ax(t) + B\max\{0, Fx(t)\}$ are contained in L_2^n. It is clear that the closed-loop system switches between the *controlled mode* $\dot{x} = (A + BF)x$ and the *uncontrolled mode* $\dot{x} = Ax$ on the basis of the switching plane $Fx = 0$.

If the set of eigenvalues of A is denoted by $E_\lambda(A)$, the main theorem in Willems et al. [29] reads as follows:

Theorem 3 *Suppose that (A, B) has scalar input and A has at most one pair of unstable, complex conjugate eigenvalues. The problem of positive feedback stabilizability is solvable if and only if (A, B) is stabilizable, i.e. there exists a matrix F such that $A + BF$ is stable, and $E_\lambda(A) \cap \mathbb{R}_{>0} = \emptyset$.*

The proof of Theorem 3 relies on the fact that there exists a transformation (for example the Jordan normal form) which separates the system (2) into two subsystems described by

$$\dot{x}_1 = A_{11}x_1 + B_1 u \tag{7a}$$

$$\dot{x}_2 = A_{22}x_2 + B_2 u \tag{7b}$$

such that A_{11} anti-stable (i.e. $-A_{11}$ stable), A_{22} asymptotically stable and (A_{11}, B_1) controllable. The stability of A_{22} makes that the control design can be limited to finding an F_1 such that $u = \max\{0, F_1 x_1\}$ is a stabilizing input for (7a). If this u is in L_2, then also $x_2 \in L_2$ by the stability of A_{22} in Equation (7b).

Willems et al. [29] provide a simple criterion on the poles of the closed loop system $\dot{x} = (A_{11} + B_1 F_1)x$ such that the system (7) is positively stabilized. Denote the eigenvalues of A_{11} by $E_\lambda(A_{11}) = \sigma_0 \pm \imath\omega_0$, where $\omega_0 \neq 0$. The closed loop system (7) with $u = \max\{0, F_1 x_1\}$ is stable if F_1 is designed such that the eigenvalues $E_\lambda(A_{11} + B_1 F_1) = \sigma \pm \omega\imath$ are taken inside the cone

$$\left\{ \sigma + \imath\omega \in \mathbb{C} \,\middle|\, \sigma < 0 \text{ and } \left|\frac{\omega}{\sigma}\right| < \left|\frac{\omega_0}{\sigma_0}\right| \right\}, \tag{8}$$

given that the assumptions of Theorem 3 are satisfied. In other words, the poles $\sigma + \imath\omega$ of the system in controlled mode should have a 'oscillating/damping'-ratio which allows for compesation of possible divergent bahaviour of the system in uncontrolled mode. It should be mentioned that if F_1 is chosen such that $E_\lambda(A_{11} + B_1F_1) = \{\sigma_1, \sigma_2\}$, possibly $\sigma_1 = \sigma_2$, then u also yields an asymptotically stable system.

The criterion (8) can easily be visualised in the imaginary plane, as depicted in Figure 4. The blue shaded region highlites the region in which the poles of the system in controlled mode $\dot{x}(t) = (A + BF)x(t)$ may be placed to ensure stability. If $\omega_0 = 0$, then the eigenvalues $E_\lambda(A + BF)$ may be placed anywhere in the open left half plane.

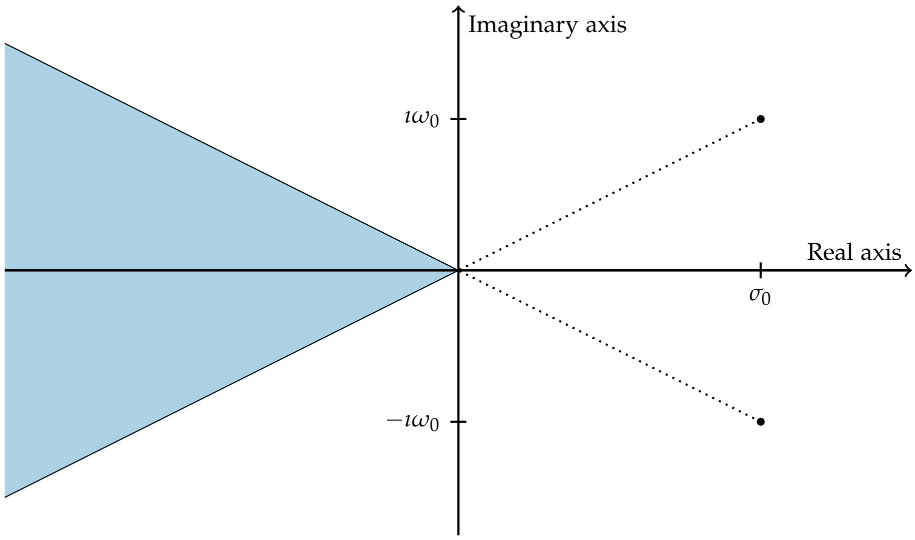

Figure 4: The blue shaded plane displays the allowable region in which the eigenvalues $E_\lambda(A + BF)$ may be placed.

The results of Willems et al. [29] can easily be supported by means of the following example.

Example 1. Consider for this example the system (2) with

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

such that the rank of $\mathcal{C}_{\{A,B\}}$ equals 2. The eigenvalues of A are equal to $1 \pm \imath$. According to (8), the system is stabilized by the positive control with eigenvalues $E_\lambda(A + BF) = -1 \pm \frac{1}{2}\imath$ in controlled mode. This is indeed confirmed by the results of a simulation of this system which is included in Figure 5. On the other hand, Willems et al. [29] yields no guarantee that $E_\lambda(A + BF) = -\frac{1}{2} \pm \imath$ results in stabilization of the system. Figure 5 shows the result of this simulation. It is obvious that the system is not asymptotically stable.

In a similar fashion Figure 4 can be 'reproduced' by placing the eigenvalues $E_\lambda(A + BF)$ on some grid. Figure 7 shows the result of a brute force series of simulations, where each simulation considers a poleplacement of $A + BF$ at some $\sigma \pm \imath\omega$ (so conjugate transpose pairs), with grids $\sigma =$`[-.1:-.1:-4.9]` and $\omega =$`[0:.1:4.9]`. The case where $\omega = 0$ was extended with an additional pole on the real axis on the same grid as was used for σ. Figure 7 plots the $E_\lambda(A + BF) = \sigma \pm \imath\omega$ which resulted in a stabilized system green, and those $E_\lambda(A + BF) = \sigma \pm \imath\omega$ that did not stabilize the system red. The figure shows a clear resemblance with (the theoretic) Figure 4. △

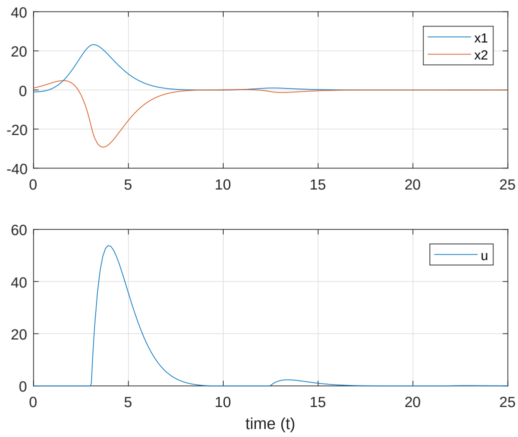

Figure 5: Results of a simulation with $E_\lambda(A+BF) = -1 \pm \frac{1}{2}\imath$ (stable).

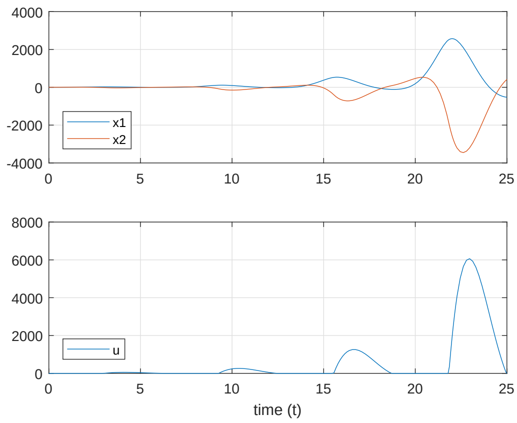

Figure 6: Results of a simulation with $E_\lambda(A+BF) = -\frac{1}{2} \pm \imath$ (not stable).

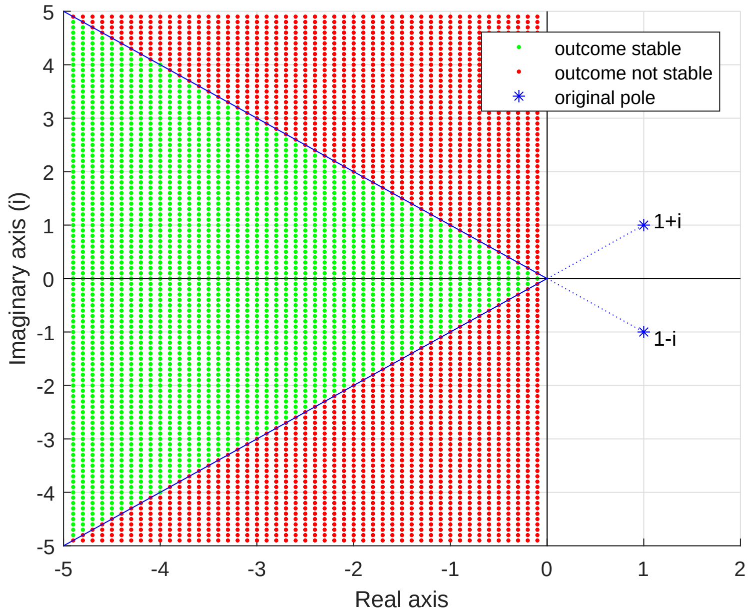

Figure 7: Results of a series of simulations of placement of eigenvalues of $A_{11} + B_1F_2$.

3. The four-dimensional problem

The approach of Section 2.2 reduces the original n-dimensional positive control problem (2) to a two dimensional problem. The problem is called 'two-dimensional' in the sense that the system has *two* non-stable poles, or in the sense that the feedback matrix $F = \begin{bmatrix} f_1 & f_2 \end{bmatrix}$ contains *two* degrees of freedom to place the *two* poles of $A + BF$. This section continues with the problem described in Section 2.2, and extends it to a system for which matrix A has at most two pairs of unstable complex conjugate eigenvalues, where the control input $u(t)$ is scalar. Thereby the n-dimensional positive control problem can in a similar fashion be reduced to a four dimensional problem. The question that rises is if such a system can be stabilized with a *scalar* positive state feedback $u(t)$.

With an extension to the four-dimensional problem, the problem is similarly extended to a six-dimensional, eight-dimensional or any higher order *even*-dimensional problem. Note that only the even dimensional problems are considered, and that for example the three dimensional problem is not considered. In case of any odd dimensional problem A has an odd number of eigenvalues. Hence at least one of the eigenvalues must be purely real, since only an even number can form complex conjugate pairs. If this real eigenvalue is negative, then it is not of interest for the positive control stabilization problem. If on the other hand this eigenvalue is positive, then there always exist initial conditions for which the system cannot be stabilized via positive control.

In terms of the illustrative pendulum problem from Section 1.3 the four-dimensional control problem with scalar input considers for example two pendula on which the exact same force $u(t)$ is excerted. Intuitively one could excert a force u on the pendula whenever *both* pendula move against the direction of force u, just as was done in the pendulum example. This could yield a state feedback matrix of the form $F = [0, a_1, 0, a_2]$, for some scalar $a_1, a_2 > 0$. Such an approach could work as long as there is a time span in which both pendula move against the direction of u. One can imagine a situation in which both pendula have the same eigenfrequency.[4] Then there exist initial conditions such that there is never a time span in which the pendula move in the same direction, and hence there is never control input. This illustrates that the positive control problem with scalar state feedback may not be as trivial for the four dimensional problem as it is in the two dimensional problem. Hence, the criterion for stability as presented in Willems et al. may not be as simple for the four dimensional problem.

Formally, consider again the system (2), where $x(t) \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times 1}$. In order for A to have (at most) two pairs of non-stable complex conjugate pole pairs, it should hold that $n \geq 4$. For the case considered, let $u(t)$ be a scalar positive state feedback of the form $u(t) = \max\{0, Fx(t)\}$, with $F \in \mathbb{R}^{1\times n}$. At this point, matrix F need not be fixed, but may for example be dependent on x, such that one could write $F(x)$. Such types of nonlinear control will be adressed later in this section.

The assumptions from Willems et al. [29] concerning the stabilizability of the pair (A, B) and $E_\lambda(A) \cap \mathbb{R}_+$ are maintained here. Since Willems et al. already considered the cases where A has zero or one stable (but not asymptotically) or unstable conjugate pole pair, these cases do not have to be reconsidered here. Only the case where A has exactly two unstable complex conjugate pole pairs are considered here.

Following the aproach of Willems et al. [29], consider system matrices A for which there exists a nonsingular transformation T and a corresponding state vector $\tilde{x} = Tx$ which

[4] It should be mentioned that if A has purely imaginary eigenvalues with same eigenfrequency, that is $E_\lambda(A) = \{\pm\omega\imath, \pm\omega\imath\}$, then the pair (A, B) is not controllable. For controllability to hold for systems with the same eigenfrequency the real part of at least one of the eigenvalues must be nonzero and if both are nonzero they must be distinct, that is $E_\lambda(A) = \{\sigma_1 \pm \omega\imath, \sigma_2 \pm \omega\imath\}$, $\sigma_1 \neq \sigma_2$.

separates the states into $\tilde{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^{\mathrm{T}}$, with $x_1, x_2 \in \mathbb{R}^2$ and $x_3 \in \mathbb{R}^{n-4}$, such that (2) can be transformed into

$$\dot{x}_1 = A_{11}x_1 + B_1 u, \quad x_1(0) = x_a \tag{9a}$$
$$\dot{x}_2 = A_{22}x_2 + B_2 u, \quad x_2(0) = x_b \tag{9b}$$
$$\dot{x}_3 = A_{33}x_3 + B_3 u, \quad x_3(0) = x_c \tag{9c}$$

with A_{11} and A_{22} anti-stable, A_{33} asymptotically stable (possibly of dimension 0), and (A_{11}, B_1), (A_{22}, B_2) stabilizable. Since A_{33} is asymptotically stable, it holds that the state vector $x_3 \in L_2$ for any input $u \in L_2$. Therefore the subsystem (9c) is not of so much interest for finding a stabilizing positive state feedback. The problem of stabilizing (9) can be reduced to finding a stabilizing input for both (9a) and (9b). In the case considered now, $u(t)$ is scalar and no two distinct controls u_1 and u_2 can be excerted.

It should be mentioned that the notation of $\tilde{x}$ will not used throughout this report. Instead the system is assumed to be of the form (9), with $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^{\mathrm{T}}$.

3.1. Brainstorm on approaches

Willems et al. [29] provide conditions under which subsystems (9a) and (9b) can be positively stabilized *individually*. That is, it is known how F_1 and F_2 should be chosen for $u_1 = \max\{0, F_1x_1\}$ and $u_2 = \max\{0, F_2x_2\}$ to be stabilizing positive feedback controls for (9a) and (9b) respectively.

It may be tempting to simply let $F = \begin{bmatrix} F_1 & F_2 \end{bmatrix}$ such that the control input is computed as $u = \max\{0, F_1x_1 + F_2x_2\}$. This however yields no guarantees for the stability of the system. First of all since if $u > 0$ the subsystems are given by

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + B_1F_1x_1 \quad + \quad B_1F_2x_2, \\ \dot{x}_2 &= \underbrace{A_{22}x_2 + B_2F_2x_2}_{\text{as in [29]}} \quad + \quad \underbrace{B_2F_1x_1}_{\text{disturbances}}, \end{aligned}$$

where clearly the 'first part' is in line with Willems et al. [29], but the cross terms act as 'disturbances' between the systems. Furthermore, if (9a) requires control $u = F_1x_1 > 0$ and at the same time (9b) requires $u = F_2x_2 > 0$, then applying the control computed as $u = F_1x_1 + F_2x_2$ is a too large input for either subsystem. One could argue to compute u as a weighted average of the two controls as $u = \max\{0, aF_1x_1 + (1-a)F_2x_2\}$, for some $a \in (0, 1)$, in order to make a trade-off between the two systems. However, no guarantee for stability follows via Willems et al. [29]. Another rather important aspect is that Willems et al. [29] pose conditions on the eigenvalues of $A_{11} + B_1F_1$, which in that paper is equivalent to the eigenvalues of the system. Here, an additional system is present and one should be careful that in general $E_\lambda(A + BF) \neq E_\lambda(A_{11} + B_1F_1) \cup E_\lambda(A_{22} + B_2F_2)$.

The suggested controls described above violate the original switching planes $F_1x_1 = 0$ and $F_2x_2 = 0$. In any case, these planes indicate when their respective subsystem should receive some positive control input and when not, according to the result of Willems et al. [29]. This suggests a control law of the form

$$\begin{aligned} u(t) &> 0 \quad \text{if } (F_1x_1(t) > 0) \wedge (F_2x_2(t) > 0) \\ u(t) &= 0 \quad \text{otherwise} \end{aligned}$$

It is debatable what the value of $u(t)$ should be whenever positive. It may be desired that the control is at least proportional to x. One could for example take $u(t) =$

$\max\{F_1x_1(t), F_2x_2(t)\}$, or compute $u(t)$ as some weighted average of F_1x_1 and F_2x_2 (whenever positive). Note that the criterion $(F_1x_1(t) > 0) \wedge (F_2x_2(t) > 0)$ may be very restrictive on the time span in which control is applied. Moreover, if the subsystems have the same eigenfrequency, i.e. $\omega_1 = \omega_2$, then there exist initial conditions such that the criterion $(F_1x_1(t) > 0) \wedge (F_2x_2(t) > 0)$ is never satisfied.

Another variation such control could be

$$u(t) = \begin{cases} \widetilde{F}_1x_1(t) + \widetilde{F}_2x_2(t) & \text{if } (F_1x_1 > 0) \wedge (F_2x_2 > 0) \\ F_1x_1(t) & \text{if } (F_1x_1 > 0) \wedge (F_2x_2 \leq 0) \\ F_2x_2(t) & \text{if } (F_1x_1 \leq 0) \wedge (F_2x_2 > 0) \\ 0 & \text{otherwise,} \end{cases}$$

for some $\widetilde{F}_1$ and $\widetilde{F}_2$ of appropriate size, possibly chosen as aF_1 and $(1-a)F_2$ respectively for some $a \in (0,1)$.

More of these types of control can be *thought up*, but none of these examples yield a guarantee for stability based on Willems et al. [29] only. In other words, the examples above illustrate that the separate results for F_1 and F_2 cannot simply be copied to the four-dimensional problem.

It may make more sense to look at the poles of the controlled system as a whole: $E_\lambda(A + BF)$. One of the conclusions from Willems et al. [29] is that pole placement on the real axis renders the positive control system stable, no matter the eigenvalues of A_{11}. Intuitively, in line with Willems et al. [29], one could presume that placing the poles of $E_\lambda(A + BF)$ on the negative real axis would suffice in stabilizing the system with $u(t) = \max\{0, Fx(t)\} = \max\{0, F_1x_1(t) + F_2x_2(t)\}$. This approach is considered in the following example.

Example 2. Consider the system given by Equations (9a) and (9b) with

$$A_{11} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix}^{\mathrm{T}}.$$

The eigenvalues of A are given by $E_\lambda(A) = \{\pm\imath, \pm 3\imath\}$, and hence are stable eigenvalues (but not asymptotically). Therefore without control input the system's states cannot explode. In this example the state feedback $u(t) = \max\{0, Fx(t)\}$ is computed such that in controlled mode $A + BF$ has its eigenvalues placed on the negative real axis. Two configurations are considered in this example, namely $E_\lambda(A + BF) = \{-2, -3, -4, -6\}$, and $E_\lambda(A + BF) = \{-3, -4, -6, -7\}$. Simulations were run with initial condition $x_0 = \begin{bmatrix} 3 & 1 & 2 & 1 \end{bmatrix}^{\mathrm{T}}$. Figures 8 and 9 show the result for the former set of eigenvalues, and Figures 10 and 11 for the latter set.

The results speak for themselves. If the poles for controlled mode are placed at $E_\lambda(A + BF) = \{-2, -3, -4, -6\}$, then the simulation shows a stabilization of the states at $x = 0$, whereas placement at $E_\lambda(A + BF) = \{-3, -4, -6, -7\}$ shows an unstable result. Note that this example considers a stable (but not asymptotically) system. That is, the eigenvalues of A are purely imaginary, and hence the 'exploding' character that is presented in Figures 10 and 11 is solely due to the control input u. △

As the previous example shows, placing all poles of $A + BF$ on the negative real axis does not guarantee an asymptotically stable system, in contrast to the result in Willems et al. [29]. There the result is based on the following observation: If the eigenvalues of $A_{11} + B_1F_1$ are real, then the system will eventually remain in controlled mode. Indeed, whenever controlled mode is entered from uncontrolled mode at time t_0 and state x_0, the dynamics are determined by $\dot{x}(t) = (A_{11} + B_1F_1)x(t)$. Then for the switching plane one

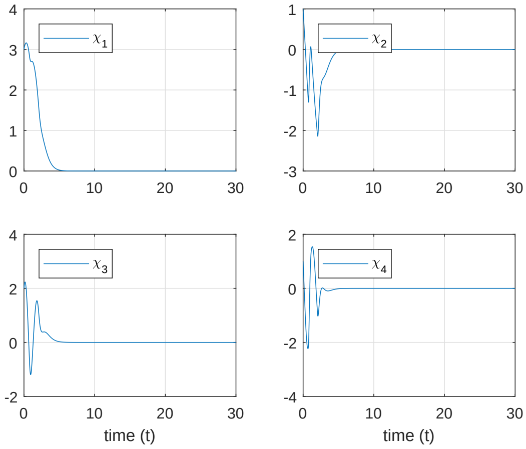

Figure 8: State trajectories for eigenvalues $E_\lambda(A+BF) = \{-2,-3,-4,-6\}$.

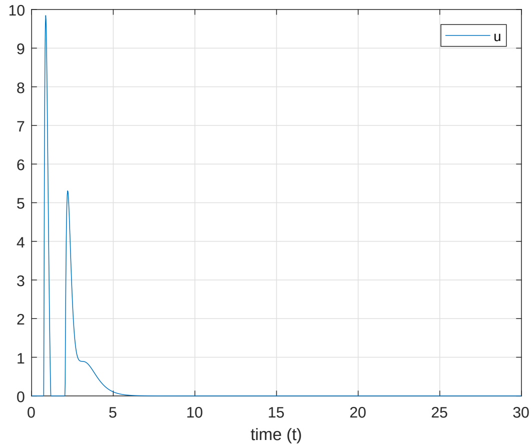

Figure 9: Control input for eigenvalues $E_\lambda(A+BF) = \{-2,-3,-4,-6\}$.

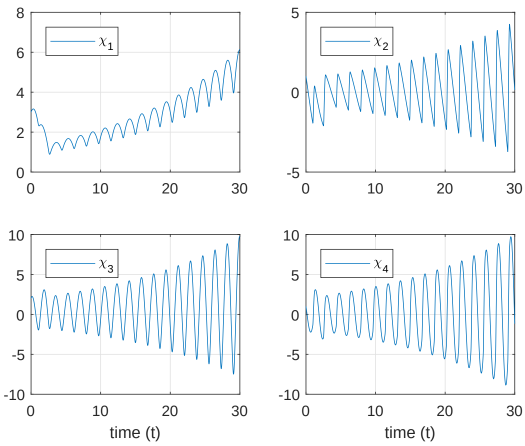

Figure 10: State trajectories for eigenvalues $E_\lambda(A+BF) = \{-3,-4,-6,-7\}$.

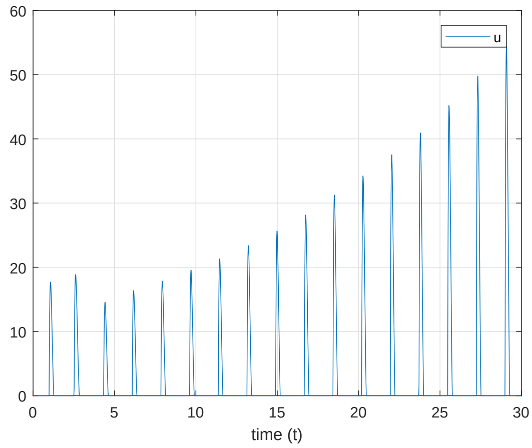

Figure 11: Control input for eigenvalues $E_\lambda(A+BF) = \{-3,-4,-6,-7\}$.

finds $F_1 x(t) = F_1 e^{(A_{11}+B_1F_1)(t-t_0)}x_0$, which can have at most one zero (since the eigenvalues of $A_{11}+B_1F_1$ are real). This zero already occured at $t = t_0$, and thus there will be no switch back to incontrolled mode. No such guarantee can be given for the four-dimensional system.

To further backup that the results from Willems et al. [29] cannot trivially be copied to the four-dimensional problem, consider Figure 12. A similar strategy as Figure 7 was tried. As it is harder to visualize the four-dimensional problem in a two-dimensional plot, two of the placed poles were fixed at -2 for this example. The others were placed at a similar grid as in Figure 7. It should be mentioned that the system used to generate this plot is a different one than in the previous example. Figure 12 was generated using a matrix A for which $E_\lambda(A) = \{\pm\imath, \pm 2\imath\}$, $B = \begin{bmatrix}1 & 0 & 1 & 0\end{bmatrix}^\mathrm{T}$ and $x_0 = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathrm{T}$. The figure is just included as an indication that a clear bound on the poleplacement is not as clearly marked out as in the two-dimensional problem.

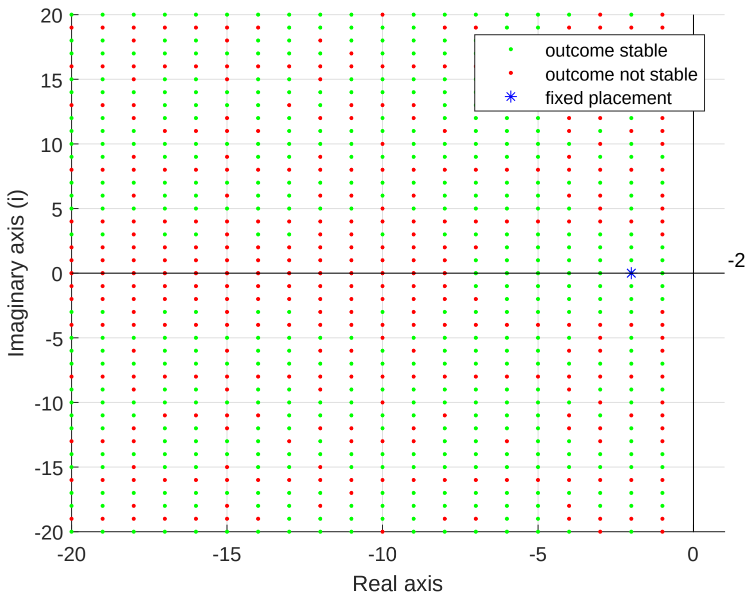

Figure 12

3.2. Motivation

The examples in the previous section show that the four-dimensional problem should be approached in a different way. It is clear that the results from the two-dimensional problem as described by Willems et al. [29] provide no direct guarantee for a stabilizing positive control for the four-dimensional problem. The goal of the remainder of this report is to investigate control laws for stabilizing the four-dimensional problem, as to form an extension to the problem considered in Willems et al. [29]. Hence, in contrast to some of the suggested control laws of the previous section, the positive state feedback control law of the form

$$u(t) = \max\{0, Fx(t)\},$$

will be considered, where the state feedback matrix F is invariant of time and state. The following section wraps up this chapter by stating a formal problem statement. This formally describes the problem that is considered throughout this report.

Section 5 and Section 4 individually propose two separate techniques to approach the four-dimensional positive control problem with the above described control law. Section 4 proposes a control law based upon Lyapunov's stability theory. Thereafter, Section 5 applies techniques of singular perturbations for ordinary differential equations to the positive control problem. Approaching the problem via singular perturbations allows the re-use of the approach considered in Willems et al. [29].

3.3. Formal problem statement

Consider only the four-dimensional non-stable subsystem extracted from (9) given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u, \quad x_1(0) = x_a, \quad x_2(0) = x_b. \tag{10a}$$

Denote the eigenvalues of A by $E_\lambda(A) = \{\sigma_1 \pm \omega_1 \imath, \sigma_2 \pm \omega_2 \imath\}$ such that

$$E_\lambda(A_{11}) = \sigma_1 \pm \omega_1 \imath \quad \text{and} \quad E_\lambda(A_{22}) = \sigma_2 \pm \omega_2 \imath. \tag{10b}$$

Note that $\sigma_1, \sigma_2 \geq 0$ and $\omega_1, \omega_2 \in \mathbb{R}$. Let the positive control state feedback be given by

$$u(t) = \max\{0, F_1 x_1(t) + F_2 x_2(t)\}. \tag{10c}$$

The question is whether such positive state feedback (10c) can be designed such that it stabilizes the system (10a) with eigenvalues (10b).

Based on the sign of $F_1x_1 + F_2x_2$ the system (10a) switches between the uncontrolled mode (11a) and the controlled mode (11b) below

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{11a}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} + B_1F_1 & B_1F_2 \\ B_2F_1 & A_{22} + B_2F_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (A_{11} + B_1F_1)x_1 + B_1F_2x_2 \\ (A_{22} + B_2F_2)x_2 + B_2F_1x_1 \end{bmatrix}. \tag{11b}$$

where (11a) holds when $F_1x_1 + F_2x_2 \leq 0$, and (11b) holds when $F_1x_1 + F_2x_2 > 0$.

Some words should be dedicated to the notation that is used. In general, capital letters are used to indicate (sub)matrices, and lower case letters for their entries. For matrices A, B and F this indicates the following general notation:

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix},$$

$$F = \begin{bmatrix} F_1 & F_2 \end{bmatrix} = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 \end{bmatrix}.$$

The reader should be aware that for example for A_{22} this yields the somewhat unfortunate notation of

$$A_{22} = \begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix}.$$

The state vector is always written as lower case, its subvectors are also written as lower case. This poses possible ambiguity for the notation of the state vector x elementwise. As a solution, the i^{th} element of x is denoted by χ_i, such that

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \chi_1 \\ \chi_2 \\ \chi_3 \\ \chi_4 \end{bmatrix}.$$

Just as for A_{22}, be aware that in this notation $x_2 = \begin{bmatrix} \chi_3 & \chi_4 \end{bmatrix}^{\mathsf{T}}$.

4. Lyapunov approach

Lyapunov proposed two methods for demonstrating stability. Lyapunov's second method, which is referred to as the *Lyapunov's stability criterion*, uses a *Lyapunov function*, commonly denoted by $V(x)$. A Lyapunov function is a scalar function defined on the state space, which can be used to prove the stability of an equilibrium point. The Lyapunov function method is applied to study the stability of various differential equations and systems.

This section uses theory of Lyapunov to design a stabilizing positive control law for stable systems in Section 4.2. The application of such control law for unstable systems is considered in Section 4.3. The following section covers some of the necessary background on Lyapunov stability theory.

4.1. Lyapunov stability theory

Consider the n-dimensional system of differential equations

$$\dot{x}(t) = f(x(t)), \quad x(t_0) = x_0 \in \mathbb{R}^n, \quad t \geq t_0, \tag{12}$$

with equilibrium $x = 0$. Suppose we have a function $V : \mathbb{R}^n \to \mathbb{R}$ that does not increase along any solution $x(t)$ of (12), i.e., that

$$V(x(t+\tau)) \leq V(x(t)), \quad \forall \tau > 0,$$

for every solution $x(t)$ of (12). Now if $V(x(t))$ is differentiable with respect to time t then it is nonincreasing if and only if its derivative with respect to time is non-positive everywhere, so

$$\dot{V}(x(t)) \leq 0 \quad \forall t.$$

Using the chain-rule one finds

$$\begin{aligned} \dot{V}(x(t)) = \frac{\mathrm{d}V(x(t))}{\mathrm{d}t} &= \frac{\partial V(x(t))}{\partial x_1}\dot{x}_1(t) + \cdots + \frac{\partial V(x(t))}{\partial x_n}\dot{x}_n(t) \\ &= \frac{\partial V(x(t))}{\partial x_1} f_1(x(t)) + \cdots + \frac{\partial V(x(t))}{\partial x_n} f_n(x(t)) =: \frac{\partial V(x)}{\partial x^{\mathrm{T}}} f(x)\bigg|_{x=x(t)}, \end{aligned} \tag{13}$$

where

$$\frac{\partial V(x)}{\partial x^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial V(x)}{\partial x_1} & \frac{\partial V(x)}{\partial x_2} & \dots & \frac{\partial V(x)}{\partial x_n} \end{bmatrix}.$$

In order to derive stability from the existence of a non-increasing function $V(x(t))$ it is additionally required that the function has a minimum at the equilibrium. It is furthermore assumed that $V(\bar{x}) = 0$. The main theorem follows.

Theorem 4 (Lyapunov's second stability theorem (as for instance presented in [17])) *Consider (12) with equilibrium $\bar{x}$. If there is a neighbourhood Ω of $\bar{x}$ and a function $V : \Omega \to \mathbb{R}$ such that on Ω*

1. *$V(x)$ is continuously differentiable,*
2. *$V(x)$ has a unique minimum on Ω in $\bar{x}$,*
3. *$\frac{\partial V(x)}{\partial x^{\mathrm{T}}} f(x) \leq 0$ for all x,*

then $\bar{x}$ is a stable equilibrium, and then $V(x)$ is called a Lyapunov function. *If in addition $\frac{\partial V(x)}{\partial x^{\mathrm{T}}} f(x) < 0$ for $x \neq \bar{x}$, then $\bar{x}$ is asymptotically stable and then $V(x)$ is referred to as a* strong Lyapunov function.

An additional condition called 'radial unboundedness' is required in order to conclude global stability. The function $V(x)$ is *radially unbounded* if $V(x) \to \infty$ as $||x|| \to \infty$. Then the theorem on global asymptotic stability reads as follows:

Theorem 5 *If $V : \mathbb{R}^n \to \mathbb{R}$ is a strong Lyapunov function for $\bar{x}$ on the entire state space $\mathbb{R}^n$ and $V(x)$ is radially unbounded, then the system is globally asymptotically stable.*

For Lyapunov stability of linear state space models, consider the linear system (10). With state feedback control, i.e. $u = Fx$, the system can be rewritten as

$$\dot{x} = (A + BF)x = \widetilde{A}x. \tag{14}$$

For these types of systems the following theorem holds.

Theorem 6 (Lyapunov equation (as presented in [17])) *Let $A \in \mathbb{R}^{n\times n}$ and consider the system $\dot{x}(t) = Ax(t)$ with equilibrium point $\bar{x} = 0 \in \mathbb{R}^n$. Suppose $Q \in \mathbb{R}^{n\times n}$ is positive definite and let*

$$V(x_0) := \int_0^\infty x^{\mathrm{T}}(t)Qx(t)\,dt \tag{15}$$

in which $x(0) = x_0$. The following three statements are equivalent.

1. *$\bar{x} = 0$ is a globally asymptotically stable equilibrium of $\dot{x}(t) = Ax(t)$.*
2. *$V(x)$ defined in (15) exists for every $x \in \mathbb{R}^n$ and it is a strong Lyapunov function for this system. In fact, $V(x)$ is then quadratic, $V(x) = x^{\mathrm{T}}Px$, with $P \in \mathbb{R}^{n\times n}$, the well defined positive definite matrix*
$$P := \int_0^\infty e^{A^{\mathrm{T}}t}Qe^{At}\,dt. \tag{16}$$
3. *The linear matrix equation*
$$A^{\mathrm{T}}P + PA = -Q \tag{17}$$
has a unique solution P, and this P is positive definite. The quadratic function $V(x) = x^{\mathrm{T}}Px$ is a Lyapunov function *for this system.*

In that case the P in (16) and (17) are the same.

If in Equation (17) P is positive definite and Q is positive *semi*-definite (see Definition B.1b), then all trajectories of $\dot{x}(t) = Ax(t)$ are bounded. This means that all eigenvalues of A have nonpositive real part. Equation (17) is called the *continuous Lyapunov equation*. The existence of a solution of which is covered in the following theorem:

Theorem 7 (Continuous Lyapunov equation) *Let $A, P, Q \in \mathbb{R}^{n\times n}$ where P and Q are symmetric. Given any Q which is positive definite, there exists a unique positive definite P such that*

$$A^{\mathrm{T}}P + PA = -Q \tag{18}$$

if and only if all eigenvalues of A have negative real part.

Consider the candidate Lyapunov function $V(x) = \frac{1}{2}x^{\mathrm{T}}Px$ for the positive control system (10a). Recall that in this case a four dimensional problem is concerned, so $n = 4$. Let $P \in \mathbb{R}^{4\times 4}$ be a symmetric positive definite matrix of the form $P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$, with $P_1, P_2 \in \mathbb{R}^{2\times 2}$ also positive definite, such that the candidate Lyapunov function can be rewritten as

$$V(x) = \tfrac{1}{2}x_1^{\mathrm{T}}P_1x_1 + \tfrac{1}{2}x_2^{\mathrm{T}}P_2x_2. \tag{19}$$

For sure, this $V(x)$ satisfies the first two conditions of Theorem 4 since P is positive definite, and $V(x)$ is quadratic in x hence positive definite relative to $\bar{x}=0$. Note also that $V(x)$ is radially unbounded. Recall that a positive control system is considered. Therefore, it is desired that $V(x)$ is a Lyapunov function for the system $\dot{x}=Ax+Bu$, for all $u>0$ as well as for $u=0$. For the derivative of $V(x(t))$ with respect to time, denoted by $\dot{V}(x(t))$, one finds

$$
\begin{aligned}
2\dot{V}(x) &= \dot{x}_1^{\mathrm{T}}P_1x_1 + x_1^{\mathrm{T}}P_1\dot{x}_1 + \dot{x}_2^{\mathrm{T}}P_2x_2 + x_2^{\mathrm{T}}P_2\dot{x}_2 \\
&= (A_{11}x_1 + B_1u)^{\mathrm{T}}P_1x_1 + x_1^{\mathrm{T}}P_1(A_{11}x_1 + B_1u) + (A_{22}x_2 + B_2u)^{\mathrm{T}}P_2x_2 + x_2^{\mathrm{T}}P_2(A_{22}x_2 + B_2u) \\
&= (x_1^{\mathrm{T}}A_{11}^{\mathrm{T}} + B_1^{\mathrm{T}}u)P_1x_1 + x_1^{\mathrm{T}}P_1(A_{11}x_1 + B_1u) + (x_2^{\mathrm{T}}A_{22}^{\mathrm{T}} + B_2^{\mathrm{T}}u)P_2x_2 + x_2^{\mathrm{T}}P_2(A_{22}x_2 + B_2u) \\
&= x_1^{\mathrm{T}}(A_{11}^{\mathrm{T}}P_1 + P_1A_{11})x_1 + x_2^{\mathrm{T}}(A_{22}^{\mathrm{T}}P_2 + P_2A_{22})x_2 + (x_1^{\mathrm{T}}P_1B_1 + B_1^{\mathrm{T}}P_1x_1 + B_2^{\mathrm{T}}P_2x_2 + x_2^{\mathrm{T}}P_2B_2)u \\
&= \underbrace{x_1^{\mathrm{T}}(A_{11}^{\mathrm{T}}P_1 + P_1A_{11})x_1 + x_2^{\mathrm{T}}(A_{22}^{\mathrm{T}}P_2 + P_2A_{22})x_2}_{=:\mathcal{V}_P(x)} + \underbrace{2(x_1^{\mathrm{T}}P_1B_1 + x_2^{\mathrm{T}}P_2B_2)u}_{=:\mathcal{V}_{u,P}(x)},
\end{aligned} \tag{20}
$$

where the time variable t is ommitted for notational purposes. Expression (20) is obtained since $B_i^{\mathrm{T}}P_ix_i = (B_i^{\mathrm{T}}P_ix_i)^{\mathrm{T}} = x_i^{\mathrm{T}}P_i^{\mathrm{T}}B_i = x_i^{\mathrm{T}}P_iB_i$, $i=1,2$, by the symmetry of P_i (i.e. $P_i^{\mathrm{T}}=P_i$) and by any scalar being equal to its transpose.

4.2. Control design for stable systems

For $V(x)$ to be a Lyapunov function, the third condition in Theorem 4 requires that $\dot{V}(x)\leq 0$ for all x and for all non-negative control values u. A closer inspection of Equation (20) reveils that one way $\dot{V}(x)\leq 0$ may be achieved is via the following conditions:

1. The positive definite matrices P_1 and P_2 are chosen such that $(A_{11}^{\mathrm{T}}P_1+P_1A_{11})$ and $(A_{22}^{\mathrm{T}}P_2+P_2A_{22})$ are either negative (semi-)definite, or equal to the zero matrix. If such matrices P_i can be found then the $x_i^{\mathrm{T}}(A_{ii}^{\mathrm{T}}P_i+P_iA_{ii})x_i$ terms, $i=1,2$, are quadratic in x_i and for sure $\mathcal{V}_P(x)\leq 0\ \forall x$.

2. The control law for u is designed such that $\mathcal{V}_{u,P}(x)\leq 0\ \forall x$.

If both conditions hold, then obviously $\dot{V}(x)\leq 0$ for all x. The existence of such P_1 and P_2 is linked to Theorem 7. It states that for systems $\dot{x}=Ax+Bu$ with A asymptotically stable, there exists positive definite matrices P_1 and P_2 such that $A_{11}^{\mathrm{T}}P_1+P_1A_{11}\leq 0$ and $A_{22}^{\mathrm{T}}P_2+P_2A_{22}\leq 0$. For the case where all eigenvalues of A are purely imaginary, as will be shown later, there do exist positive definite matrices P_1 and P_2 such that $A_{11}^{\mathrm{T}}P_1+P_1A_{11}=0$ and $A_{22}^{\mathrm{T}}P_2+P_2A_{22}=0$. This section considers systems for which the set of eigenvalues of A is equal to

$$E_\lambda(A) = \{\pm\omega_1\imath, \pm\omega_2\imath\}, \quad \omega_1,\omega_2\in\mathbb{R}_{>0} \tag{21}$$

where $\pm\omega_1\imath$ are the eigenvalues of A_{11} and $\pm\omega_2\imath$ are the eigenvalues of A_{22}. That way A may be assumed to be of the form

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad \text{where} \quad A_{11} = \begin{bmatrix} 0 & \omega_1 \\ -\omega_1 & 0 \end{bmatrix} \quad \text{and} \quad A_{22} = \begin{bmatrix} 0 & \omega_2 \\ -\omega_2 & 0 \end{bmatrix}.$$

Note that one finds $A_{ii}^{\mathrm{T}}=-A_{ii}$, $i=1,2$. In this case there exist positive definite matrices P_1 and P_2 such that

$$A_{11}^{\mathrm{T}}P_1 + P_1A_{11} = 0, \tag{22a}$$
$$A_{22}^{\mathrm{T}}P_2 + P_2A_{22} = 0. \tag{22b}$$

Conditions (22) hold for any matrices P_1 and P_2 which are positive multiplications of the identity matrix. Formally conditions (22) are satisfied for

$$P_1 = c_1 I, \quad P_2 = c_2 I, \tag{23}$$

where $c_1, c_2 > 0$ and I denotes the 2×2 identity matrix. This way Equation (20) simplifies to

$$\dot{V}(x) = (x_1^{\mathrm{T}} P_1 B_1 + x_2^{\mathrm{T}} P_2 B_2)u = (B_1^{\mathrm{T}} P_1 x_1 + B_2^{\mathrm{T}} P_2 x_2)u = (B^{\mathrm{T}} P x)u, \tag{24}$$

where the second equality holds since $(x_1^{\mathrm{T}} P_1 B_1 + x_2^{\mathrm{T}} P_2 B_2)$ is scalar, and therefore it equals its transpose.

Be aware that $B^{\mathrm{T}} P x$ may be alternatingly positive and negative. For $\dot{V}(x) \leq 0$ to hold for all x, the control input u is based upon the switching plane $B^{\mathrm{T}} P x = 0$ as follows:

$$\begin{aligned} u = 0 \quad &\text{if} \quad B^{\mathrm{T}} P x \geq 0, \\ u > 0 \quad &\text{if} \quad B^{\mathrm{T}} P x < 0. \end{aligned}$$

This way $V(x)$ as defined in Equation (19) is a Lyapunov function for the system Equation (10a) with input u, it is not a strong Lyapunov function though. A control law that suffices for stability is $u = \max\{0, -B^{\mathrm{T}} P x\}$, which if written out yields the control law specified as

$$u(t) = \max\{0, -B^{\mathrm{T}} P x(t)\} = \max\{0, -(B_1^{\mathrm{T}} P_1 x_1(t) + B_2^{\mathrm{T}} P_2 x_2(t))\}. \tag{25}$$

The positive control input (25) is of the state feedback form $u = \max\{0, F_1 x_1 + F_2 x_2\}$, with $F_1 = -B_1^{\mathrm{T}} P_1$ and $F_2 = -B_2^{\mathrm{T}} P_2$. The question is whether this control yields an *asymptotically stable system*.

Note that the control law designed in such way has two degrees of freedom $c_1, c_2 > 0$ for the matrices $P_1 = c_1 I$ and $P_2 = c_2 I$. These parameters could be used to tune the controller, or determine the relative weight the controller puts on state sets x_1 and x_2.

4.2.1. *Motivation and simulations*

This section briefly shows the result of a simulation in which the Lyapunov based control law (25) is applied. By the design of this control law, it is expected that it asymptotically stabilizes systems for which the eigenvalues of the state matrix A are purely imaginary. This expectation is supported by means of Example 3. In this example the control law (25) is applied in a simulation of a 'parallel pendulum system'.

Example 3. Consider the system (10a) with $\omega_1 = 1$ and $\omega_2 = 2$, such that

$$A_{11} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad A_{22} = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}.$$

Furthermore let $B_1 = B_2 = \begin{bmatrix} 0 & -1 \end{bmatrix}^{\mathrm{T}}$, and in Equation (25) let $P_1 = P_2 = I_{2\times 2}$. For this simulation let the initial state vector be equal to $x_0 = \begin{bmatrix} 1 & 2 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$. In terms of pendula, this example considers two pendula with different eigenfrequencies (due to different length/weight ratios), and different initial conditions. The results of the simulation are shown in Figures 13 and 14.

The simulation shows a couple of aspects. First of all, for this example the system is stabilized at $\bar{x} = 0$ by a control input u that is non-negative. Furthermore, Figure 14 shows that indeed $\dot{V}(x) \leq 0$ for all time and $V(x)$ is nonincreasing. All these observations live up to the expectations of the behaviour of the system.

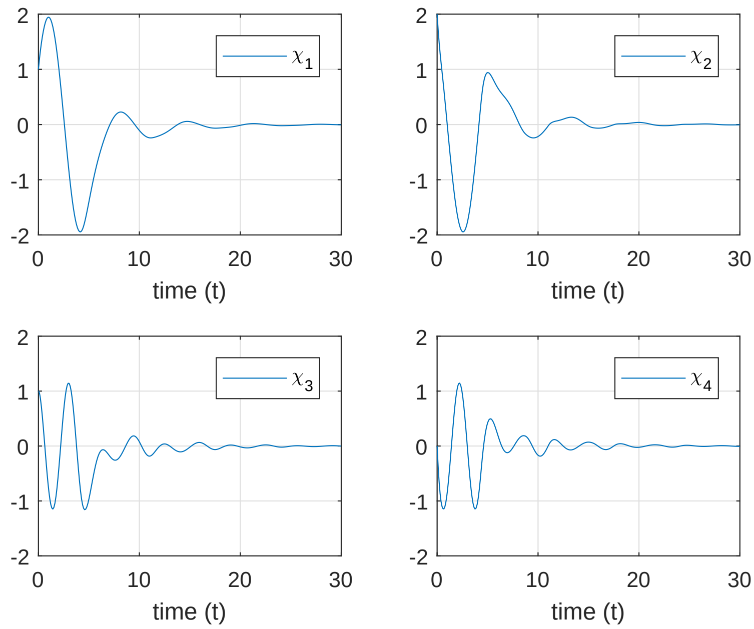

Figure 13: Simulation results for state trajectories x.

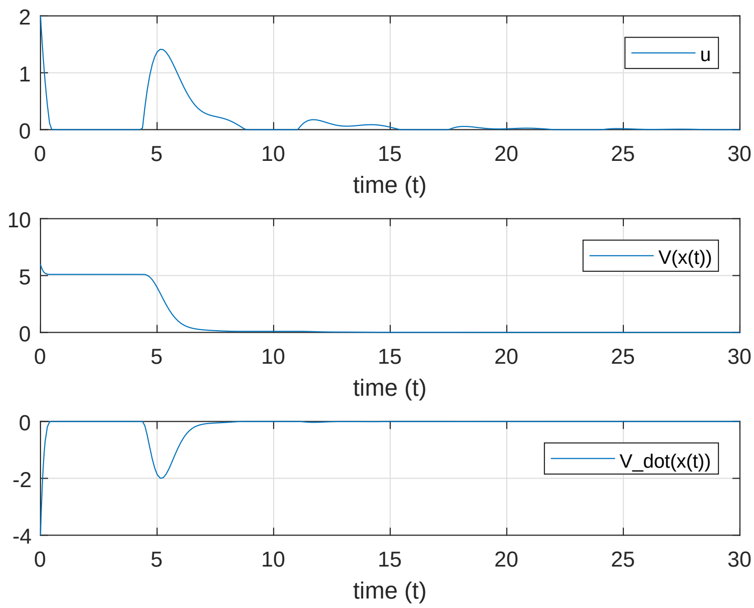

Figure 14: Simulation results for control input u and Lyapunov function $V(x)$ and its time derivative $\dot{V}(x)$.

Also for a complete illustration, the plots of a simulation of the system with $\omega_1 = \omega_2 = 2$ is included. Figures 15 and 16 show the typical result of $V(x)$ being non-increasing, but converging to a constant value instead of converging to 0, that is, state x converges to some trajectory in $\mathbb{R}^n$. $\triangle$

Example 3 motivates the step to a formal statement of stability in the coming section.

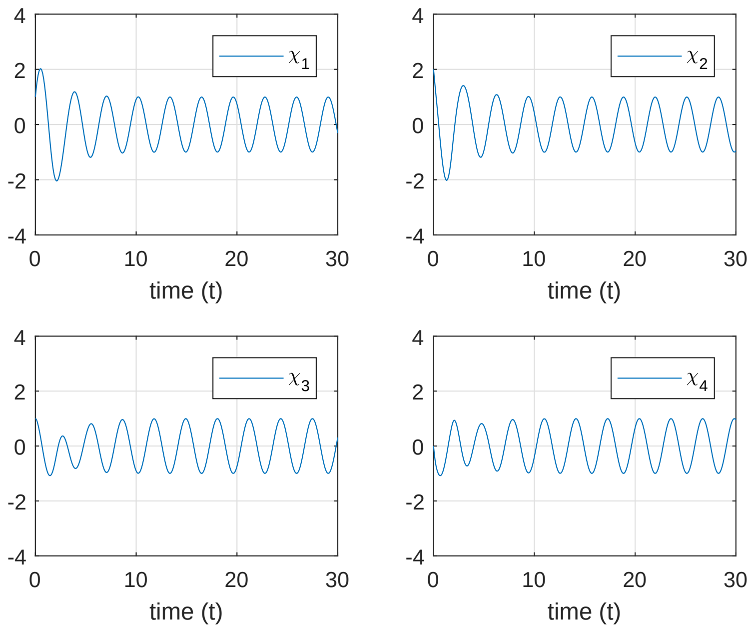

Figure 15: Simulation results for state trajectories x.

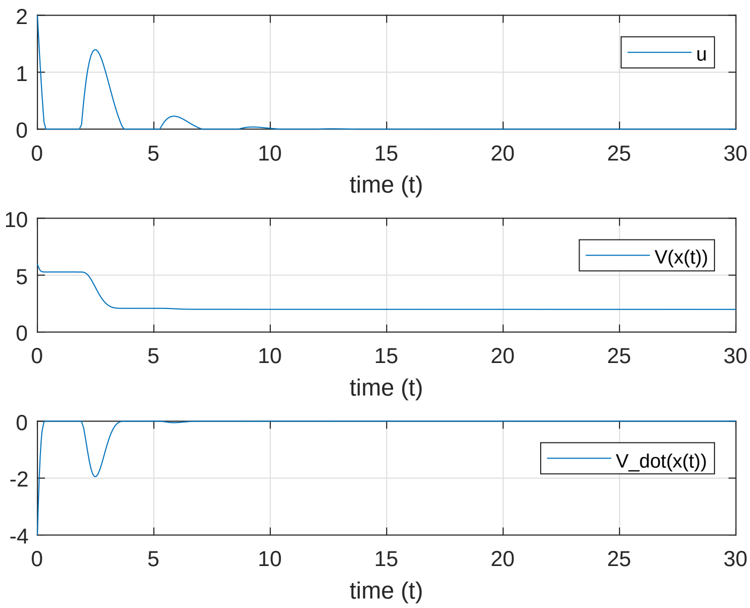

Figure 16: Simulation results for control input u and Lyapunov function $V(x)$ and its time derivative $\dot{V}(x)$.

4.2.2. Asymptotic stability

In order to analyze the stability of the positive control system (10a) with control (25), it makes sense to first investigate the behaviour of $B^{\mathsf{T}}Px(t)$ which is used in (25). Can it be the case that given some initial time t_0 and state $x(t_0) = x_0$ with $B^{\mathsf{T}}Px_0 > 0$ that $B^{\mathsf{T}}Px(t)$ remains positive for all $t \geq t_0$? In that case $u(t)$ remains 0 for all $t \geq t_0$ and the system remains 'stuck' in uncontrolled mode. The following theorem states under which conditions $B^{\mathsf{T}}Px(t)$ is ensured to change sign from positive to negative, which is needed to reach controlled mode.

Theorem 8 *$B^{\mathrm{T}}Px(t)$ switches sign from positive to negative within finite time if $\omega_1 \neq \omega_2$, or if $\omega_1 = \omega_2$ and $B_1^{\mathrm{T}}P_1x_1 \neq -B_2^{\mathrm{T}}P_2x_2$ initially.*

Proof. Let the initial conditions t_0 and $x(t_0) = x_0$ be such that $B^{\mathrm{T}}Px_0 \geq 0$. It will be proven by contradiction that $B^{\mathrm{T}}Px(t)$ switches sign to negative if $\omega_1 \neq \omega_2$, or if $\omega_1 = \omega_2$ with the additional requirement that $B^{\mathrm{T}}Px_0 \neq 0$.

Assume that $B^{\mathrm{T}}Px(t)$ does not change sign anywhere on $[t_0, \infty)$. In that case u remains equal to 0 and the value of $x(t)$ is determined by $\dot{x}(t) = Ax(t)$, with $x(t_0) = x_0$. In that case $B^{\mathrm{T}}Px(t)$ can be written as a sum of sinusoids as

$$B^{\mathrm{T}}Px(t) = \gamma_1\cos(\omega_1 t + \phi_1) + \gamma_2\cos(\omega_2 t + \phi_2), \quad \gamma_1, \gamma_2 \in \mathbb{R} \setminus 0, \quad \phi_1, \phi_2 \in [0, \pi],$$

where without loss of generality, it may be assumed that in $\gamma_1 = 1$ and $\phi_1 = 0$, such that

$$B^{\mathrm{T}}Px(t) = \cos(\omega_1 t) + \gamma\cos(\omega_2 t + \phi), \quad \gamma \in \mathbb{R} \setminus 0, \quad \phi \in [0, \pi]. \tag{26}$$

where $\phi = \phi_2$ and $\gamma = \gamma_2$, which are determined by the initial condition x_0.

Define time spans $T_1 = \frac{2\pi}{\omega_1}$ and $T_2 = \frac{2\pi}{\omega_2}$ as the periods of $\cos(\omega_1 t)$ and $\gamma\cos(\omega_2 t + \phi)$ respectively. Two cases can be distinguished, namely $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$ and $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$.

In the case where $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$ there exist $k_1, k_2 \in \mathbb{N}_{>0}$ such that $k_1T_1 = k_2T_2$. More specifically, (26) is periodic with period $T = 2m\pi$, where $m := \operatorname{lcm}(k_1, k_2)$. In this case (26) can be integrated over one period T, this integral is known to be equal to 0, i.e.,

$$0 = \int_{t_0}^{t_0+T} B^{\mathrm{T}}Px(t)\,\mathrm{d}t = \int_{t_0}^{t_0+T} \cos(\omega_1 t) + \gamma\cos(\omega_2 t + \phi)\,\mathrm{d}t. \tag{27}$$

For Equation (27) to hold, the following cases are distinguished:

1. $B^{\mathrm{T}}Px(t) = \cos(\omega_1 t) + \gamma\cos(\omega_2 t + \phi) = 0\ \forall t \in [t_0, t_0 + T]$. This condition holds if and only if $\omega_1 = \omega_2$ and $B^{\mathrm{T}}Px_0 = 0$. Here $B^{\mathrm{T}}Px_0 = 0$ implies that in (26) either $c = 1$ and $\phi = \pi$, or $c = -1$ and $\phi = 0$. This is the case if $B_1^{\mathrm{T}}P_1x_1(t) = -B_2^{\mathrm{T}}P_2x_2(t)$ for any $t \in [t_0, t_0 + T]$, and specifically for the initial time $t = t_0$. So, if $\omega_1 = \omega_2$ and $B^{\mathrm{T}}Px_0 \neq 0$ consider the following case:
2. $B^{\mathrm{T}}Px(t) = \cos(\omega_1 t) + \gamma\cos(\omega_2 t + \phi) \neq 0\ \forall t \in [t_0, t_0 + T]$. In this case $B^{\mathrm{T}}Px(t)$ must be positive as well as negative on the time interval $[t_0, t_0 + T]$, otherwise (27) could not hold. This contradicts the assumption that $B^{\mathrm{T}}Px(t)$ does not change sign. Hence there must exist a time instance $t_1 \in [t_0, t_0 + T]$ at which $B^{\mathrm{T}}Px(t)$ switches sign.

Therefore a sign switch of $B^{\mathrm{T}}Px(t)$ is guaranteed in finite time if $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$, and in the case when $\omega_1 = \omega_2$ the sign switch occurs if additionally $B_1^{\mathrm{T}}P_1x_1 \neq -B_2^{\mathrm{T}}P_2x_2$.

In the case where $\frac{\omega_1}{\omega_2}$ is irrational, that is where $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$, no $k_1, k_2 \in \mathbb{N}$ exist such that $k_1T_1 = k_2T_2$. Note also that then $\omega_1 \neq \omega_2$, which excludes the case where $B^{\mathrm{T}}Px(t) \equiv 0$. For this case let (without loss of generality) that $\omega_1 < \omega_2$, such that $T_1 > T_2$. The function $\cos(\omega_1 t)$ is strictly negative on any interval open time-interval

$$\mathcal{U}_k := \left(\tfrac{(4k+1)\pi}{2\omega_1}, \tfrac{(4k+3)\pi}{2\omega_1}\right),$$

for $k \in \mathbb{N}_{>0}$. Consider the smallest k such that $\frac{(4k+1)\pi}{2\omega_1} \geq t_0$, denote this k by $\tilde{k}$ such the interval $\mathcal{U}_{\tilde{k}}$ has length $\frac{T_1}{2}$. $\mathcal{U}_{\tilde{k}}$ contains at least one subinterval of maximum length $\frac{T_2}{2}$ on which $\gamma\cos(\omega_2 t + \phi) > 0$. Since $T_2 < T_1$, the interval $\mathcal{U}_{\tilde{k}}$ also contains a nonempty subinterval on which $\gamma\cos(\omega_2 t + \phi) \leq 0$. But then on this subinterval of it holds that $B^{\mathrm{T}}Px(t) < 0$. This implies that $B^{\mathrm{T}}Px(t)$ switched sign, which also contradicts the initial assumption that it did not. Moreover, this sign switch occurs within $\frac{(4\tilde{k}+3)\pi}{2\omega_1}$ time, which is finite. Hence, also in the case when $\frac{\omega_1}{\omega_2}$ is irrational $B^{\mathrm{T}}Px(t)$ must switch sign from positive to negative in finite time. □

Theorem 8 provides conditions which ensure that whenever the state x is in the closed half space $\{x|B^{\mathrm{T}}Px(t) \geq 0\}$, that is where $u = 0$, it leaves this space in finite time. In other words, it ensures that the control system cannot remain stuck in uncontrolled mode. Note that by construction $\dot{V}(x) = 0$ in this closed half space. Besides, $\dot{V}(x) < 0$ in the open half space $\{x|B^{\mathrm{T}}Px(t) < 0\}$. The following lemma ensures that the control input $u(t) = \max\{0, -B^{\mathrm{T}}Px(t)\}$ is an L_2 function.

Lemma 9 *For the system $\dot{x}(t) = Ax(t) + Bu(t)$, with $E_\lambda(A) = \{\pm\omega_1 \imath, \pm\omega_2 \imath\}$, the state trajectory $x(t)$ is bounded and the positive control input $u(t)$ defined by (25) is an $L_2[t_0, \infty)$ function.*

Proof. Consider the Lyapunov function $V(x(t))$ in (19) with derivative with respect to time $\dot{V}(x(t))$ given by (24). At time t the value of the Lyapunov function is equal to

$$\begin{aligned} V(x(t)) &= V(x_0) + \int_{t_0}^{t} \dot{V}(x(\tau))\,\mathrm{d}\tau \\ &= V(x_0) + \int_{t_0}^{t} (B^{\mathrm{T}}Px(\tau))u(\tau)\,\mathrm{d}\tau \\ &= V(x_0) - \int_{t_0}^{t} u^2(\tau)\,\mathrm{d}\tau \end{aligned} \tag{28}$$

where the last equality holds since either $u(\tau) = 0$, or $u(\tau) = -B^{\mathrm{T}}Px(\tau)$. Moreover, since $\dot{V}(x(t)) \leq 0$ for all $t \geq t_0$, it holds that $0 \leq V(x(t)) \leq V(x_0)$, for $t \geq t_0$. Given that $V(x_0) < \infty$ it follows that $V(x(t))$ is bounded, and hence that $x(t)$ is indeed bounded for all t (for it cannot be the case that $V(x(t))$ is bounded and $x(t)$ is unbounded). Moreover, (28) also yields that

$$\begin{aligned} &0 \leq V(x_0) - \int_{t_0}^{t} u^2(\tau)\,\mathrm{d}\tau \leq V(x_0), \text{ i.e.,} \\ &0 \leq \int_{t_0}^{t} u^2(\tau)\,\mathrm{d}\tau \leq V(x_0) < \infty, \end{aligned}$$

which holds for all $t \geq t_0$. Letting $t \to \infty$ one finds $0 \leq \int_{t_0}^{\infty} u^2(\tau)\,\mathrm{d}\tau < \infty$. Hence the control input $u(t)$ is an L_2 function on $[t_0, \infty)$. Moreover, given some finite timespan T, the sequence of integrals on the time interval $[(k-1)T, kT]$ is defined by

$$\gamma_k := \int_{(k-1)T}^{kT} u^2(\tau)\,\mathrm{d}\tau, \quad k \in \mathbb{N}. \tag{29}$$

Since $\sum_{k=1}^{\infty} \gamma_k \leq V(x_0)$, it must hold that $\gamma_k \to 0$ as $k \to \infty$. □

For any finite T it holds that

$$\int_{\tau}^{\tau+T} |u|\,\mathrm{d}t \leq \sqrt{\int_{\tau}^{\tau+T} |u|^2\,\mathrm{d}t}\sqrt{\int_{\tau}^{\tau+T} 1\,\mathrm{d}t}, \tag{30}$$

from which follows that in this case since $u|_{[\tau,\tau+T]} \xrightarrow{L_2} 0$ as $\tau \to \infty$, then also $u|_{[\tau,\tau+T]} \xrightarrow{L_1} 0$ as $\tau \to \infty$.

In order to prove the asymptotic stability of the positive control system (10a) with control (25), define the functions $\mathcal{M}_u : \mathbb{R}^n \to L_2[0,T]$ and $\mathcal{M}_0 : \mathbb{R}^n \to L_2[0,T]$ by

$$\begin{aligned} &[\mathcal{M}_u(x_0)](t) := B^{\mathrm{T}}Px(t), \quad \text{where } \dot{x}(t) = Ax(t) + B\max\{0, -B^{\mathrm{T}}Px(t)\}, \quad x(0) = x_0, \\ &[\mathcal{M}_0(x_0)](t) := B^{\mathrm{T}}Px(t), \quad \text{where } \dot{x}(t) = Ax(t), \quad x(0) = x_0, \end{aligned} \tag{31}$$

where T is some finite time span. Define similarly the positive and negative parts of the functions above separately as

$$\begin{aligned} &[\mathcal{M}_u^+(x_0)](t) := \max\{0, [M_u(x_0)](t)\}, \quad [\mathcal{M}_u^-(x_0)](t) := -\min\{0, [M_u(x_0)](t)\}, \\ &[\mathcal{M}_0^+(x_0)](t) := \max\{0, [M_0(x_0)](t)\}, \quad [\mathcal{M}_0^-(x_0)](t) := -\min\{0, [M_0(x_0)](t)\}. \end{aligned}$$

In that fashion one finds for $[\mathcal{M}_u(x_0)](t)$ that

$$\begin{aligned} [\mathcal{M}_u(x_0)](t) &= [\mathcal{M}_u^+(x_0)](t) - [\mathcal{M}_u^-(x_0)](t), \\ \Big|[\mathcal{M}_u(x_0)](t)\Big| &= [\mathcal{M}_u^+(x_0)](t) + [\mathcal{M}_u^-(x_0)](t), \\ u(t) &= [\mathcal{M}_u^-(x_0)](t), \end{aligned}$$

and similarly for $[\mathcal{M}_0(x_0)](t)$ that

$$\begin{aligned} [\mathcal{M}_0(x_0)](t) &= [\mathcal{M}_0^+(x_0)](t) - [\mathcal{M}_0^-(x_0)](t), \\ \Big|[\mathcal{M}_0(x_0)](t)\Big| &= [\mathcal{M}_0^+(x_0)](t) + [\mathcal{M}_0^-(x_0)](t). \end{aligned}$$

Serveral properties of these functions are listed below. These properties will be used in the proof of Theorem 11. Consider the following lemma:

Lemma 10 *If for the initial state $x(t_0) = x_0$ holds that $||x_0|| \geq c_l$, where $c_l > 0$ is some lower bound, then for all $T > 0$ there exist lower bounds $\mu, \mu_0 > 0$ such that*

$$\begin{aligned} \int_{t_0}^{t_0+T} \Big|[\mathcal{M}_u(x_0)](t-t_0)\Big|\, dt \geq \mu, \\ \int_{t_0}^{t_0+T} \Big|[\mathcal{M}_0(x_0)](t-t_0)\Big|\, dt \geq \mu_0. \end{aligned} \tag{32}$$

Proof. Assume without loss of generality that $t_0 = 0$. The proofs for $[\mathcal{M}_u(x_0)](t)$ and $[\mathcal{M}_0(x_0)](t)$ are similar. Hence consider the case for $[\mathcal{M}_u(x_0)](t)$.

Define 'output' $y(t) = [\mathcal{M}_u(x_0)](t)$ and recall that $[\mathcal{M}_u(x_0)](t) = B^{\mathrm{T}}Px(t)$ with $x(t)$ determined by $\dot{x}(t) = Ax(t) + Bu$ with $x(0) = x_0$. The pair (A, B) is controllable and hence the pair (A, B^{T}) is observable. From observability follows that the only state that produces output $y = 0$ is the zero state. So if then $||x_0|| \geq c_l$ then $|y(0)| = \Big|[\mathcal{M}_u(x_0)](0)\Big| \geq \tilde{c}_l$, for some lower bound $\tilde{c}_l > 0$. Since $||\dot{x}(t)||$ is bounded, it follows that for all $T > 0$ there exists a lower bound $\mu > 0$ such that

$$\int_0^T |y(t)|\,\mathrm{d}t = \int_0^T \Big|[\mathcal{M}_u(x_0)](t)\Big|\,\mathrm{d}t \geq \mu.$$

By a similar reasoning follows that for all $T > 0$ there exists a lower bound $\mu_0 > 0$ such that $\int_0^T \Big|[\mathcal{M}_0(x_0)](t)\Big|\,\mathrm{d}t \geq \mu_0$. □

Besides the property (32), the function $[\mathcal{M}_0(x_0)](t - t_0)$, $x_0 = x(t_0)$, possesses the three properties listed below, where without loss of generality $t_0 = 0$ (mainly for notational purposes) such that $x_0 = x(0)$.

1. If $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$ then $[\mathcal{M}_0(x_0)](t)$ is periodic with period T, then
$$\int_0^T [\mathcal{M}_0(x_0)](t)\,\mathrm{d}t = \int_0^T [\mathcal{M}_0^+(x_0)](t)\,\mathrm{d}t - \int_0^T [\mathcal{M}_0^-(x_0)](t)\,\mathrm{d}t = 0,$$
so it follows that
$$\int_0^T [\mathcal{M}_0^+(x_0)](t)\,\mathrm{d}t = \int_0^T [\mathcal{M}_0^-(x_0)](t)\,\mathrm{d}t. \tag{33}$$
2. In any case, also if $[\mathcal{M}_0(x_0)](t)$ is not periodic, it holds that
$$\int_0^T [\mathcal{M}_0^+(x_0)](t)\,\mathrm{d}t \leq c_u, \tag{34}$$
for all $T > 0$. That is, the integral is bounded from above by some upper bound $c_u > 0$. Moreover, it is bounded from below by $-c_u$.
3. The property (32) with $t_0 = 0$ yields
$$\int_0^T \Big|[\mathcal{M}_0(x_0)](t)\Big|\,\mathrm{d}t = \int_0^T [\mathcal{M}_0^+(x_0)](t)\,\mathrm{d}t + \int_0^T [\mathcal{M}_0^-(x_0)](t)\,\mathrm{d}t \geq \mu_0$$
for some $\mu_0 > 0$. If $[\mathcal{M}_0(x_0)](t)$ is periodic with period T, then by (33) then follows that
$$\int_0^T [\mathcal{M}_0^+(x_0)](t)\,\mathrm{d}t \geq \frac{\mu_0}{2}, \quad \int_0^T [\mathcal{M}_0^-(x_0)](t)\,\mathrm{d}t \geq \frac{\mu_0}{2}. \tag{35}$$

An important property of the functions defined in (31) is the following. Let for some initial conditions $x(t_0) = x_0$ (with initial time t_0) the state trajectory $x(t)$, $t \geq t_0$, be determined by $\dot{x}(t) = Ax(t) + Bu(t)$. Consider at time $\tau \geq t_0$ the state given by $x(\tau)$. Then for $t \in [\tau, \tau + T]$ for the functions $[\mathcal{M}_u(x(\tau))](t-\tau)$ and $[\mathcal{M}_0(x(\tau))](t-\tau)$ it holds that
$$\left| \int_\tau^{\tau+T} [\mathcal{M}_u(x(\tau))](t-\tau)\,\mathrm{d}t - \int_\tau^{\tau+T} [\mathcal{M}_0(x(\tau))](t-\tau)\,\mathrm{d}t \right| \to 0 \tag{36}$$
as $\tau \to \infty$. This follows from the following:
$$\begin{aligned}
&\left| \int_\tau^{\tau+T} [\mathcal{M}_u(x(\tau))](t-\tau) - [\mathcal{M}_0(x(\tau))](t-\tau)\,\mathrm{d}t \right| \\
&\qquad \leq \int_\tau^{\tau+T} \Big| [\mathcal{M}_u(x(\tau))](t-\tau) - [\mathcal{M}_0(x(\tau))](t-\tau) \Big|\,\mathrm{d}t \\
&\qquad = \int_\tau^{\tau+T} \left| B^{\mathrm{T}} P \int_\tau^t \mathrm{e}^{A(t-s)} Bu(s)\,\mathrm{d}s \right| \mathrm{d}t, \quad t \in [\tau, \tau+T],
\end{aligned}$$
where the latter expression defines a norm on $[\tau, \tau + T]$. For $t \in [\tau, \tau + T]$ the following inequality holds:
$$\begin{aligned}
\left|\left| B^{\mathrm{T}} P \int_\tau^t \mathrm{e}^{A(s)} Bu(s) \right|\right| &\leq ||B^{\mathrm{T}}P|| \left|\left| \int_\tau^t \mathrm{e}^{A(s)} Bu(s) \right|\right| \\
&\leq ||B^{\mathrm{T}}P|| \int_\tau^t \left|\left| \mathrm{e}^{A(t-s)} Bu(s) \right|\right|\,\mathrm{d}s \\
&\leq \underbrace{||B^{\mathrm{T}}P||}_{\text{constant}} \int_\tau^t \underbrace{\left|\left| \mathrm{e}^{A(t-s)} B \right|\right|}_{\text{constant}} ||u(s)||\,\mathrm{d}s \\
&= c \int_\tau^t u(s)\,\mathrm{d}s, \quad \text{recall } t \in [\tau, \tau+T] \\
&\leq c \int_\tau^{\tau+T} u(s)\,\mathrm{d}s \longrightarrow 0 \text{ as } \tau \to \infty
\end{aligned}$$

where $c > 0$ is some constant. Expression (36) follows.

When stability of the system $\dot{x}(t) = Ax(t) + Bu(t)$ with purely complex eigenvalues and with control (25) is concerned, two main cases can be distinguished. The case where $\omega_1 \neq \omega_2$ (which will be covered by Theorem 11) and the case when $\omega_1 = \omega_2$. The latter case is described at the end of this section, but is of lesser interest since in that case the controllability criterion of the pair (A, B) is violated.

Theorem 11 *The system $\dot{x} = Ax + Bu$ with eigenvalues $E_\lambda(A) = \{\pm\omega_1 \imath, \pm\omega_2 \imath\}$, with $\omega_1 \neq \omega_2$, and positive control input $u(t)$ defined by (25) yields an asymptotically stable equilibrium $\bar{x}(t) = 0$.*

Analogously to the proof of Theorem 8, the proof of this theorem consists of two parts: the case where $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$ and the case where $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$. For the part where $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$, the proof presented below uses *Cauchy sequences* to prove that some sequence of state vectors $x(kT)$ at time kT converges to a vector $\bar{x}$ as $k \to \infty$. An alternative proof for this statement is included in Appendix C, which uses the theorem of *Bolzano-Weierstrass*.

Proof of Theorem 11 - Part I: $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$. Fix $T > 0$ such that (33) holds, in which case Theorem 8 ensures that $B^{\mathrm{T}} P x(t)$ cannot only be positive positive on any time interval $[(k-1)T, kT]$, $k \in \mathbb{N}_{>0}$. Since $\omega_1 \neq \omega_2$, the control input $u(t)$ cannot be identically equal to 0 for $t \in [(k-1)T, kT]$ unless $x(t)$ is identically equal to 0 on that interval. The state solution at time t is given by

$$x(t) = \mathrm{e}^{A(t-t_0)} x_0 + \int_{t_0}^{t} \mathrm{e}^{A(t-\tau)} B u(\tau) \, \mathrm{d}\tau.$$

On the time interval $[(k-1)T, kT]$ the solution $x(kT)$ with initial state $x((k-1)T)$ can be expressed by

$$x(kT) = x((k-1)T) + \int_{(k-1)T}^{kT} \mathrm{e}^{A(kT-\tau)} B u(\tau) \, \mathrm{d}\tau. \tag{37}$$

Here it was used that $\mathrm{e}^{A(kT-(k-1)T)} x(kT) = \mathrm{e}^{AT} x(kT) = x(kT)$, which holds since e^{At} is T-periodic because the eigenvalues of A are purely imaginary. In this integral define $-\mathrm{e}^{A(kT-\tau)} B = -\mathrm{e}^{-A\tau} B =: z(\tau)$, which is T-periodic, such that (37) simplifies to

$$x(kT) = x((k-1)T) - \int_{(k-1)T}^{kT} z(\tau) u(\tau) \, \mathrm{d}\tau.$$

The structure of the remainder this proof is as follows. First the sequence $x(kT)$ will be shown to converge to some $\bar{x} \in \mathbb{R}^n$ as $k \to \infty$. Then it will be shown that $\bar{x} = 0$, and that not only $x(kT) \to \bar{x}$ as $k \to \infty$ but also $x(t) \to \bar{x}$ as $t \to \infty$.

It will first be shown that $x(kT)$ converges to some $\bar{x} \in \mathbb{R}^n$ as $k \to \infty$. Note that $x(kT)$ defines a sequence of vectors in $\mathbb{R}^n$ ($n = 4$ in this case), which is elementwise given by

$$x_i(kT) = x_i((k-1)T) - \int_{(k-1)T}^{kT} z_i(\tau) u(\tau) \, \mathrm{d}\tau, \tag{38}$$

where $z_i(\tau)$ denotes the i^{th} element of the column vector $z(\tau)$. Since $x(t)$ is bounded (see Lemma 9), also the sequence $x(kT)$ $(k = 1, 2, \dots)$ is a bounded sequence of vectors. Moreover, also the sequences $x_i(kT)$ are bounded for all i. It will be shown that each of these sequences $x_i(kT)$ is also *convergent*.

In (38) the expression $\langle z_i, u \rangle = \int_{(k-1)T}^{kT} z_i(\tau) u(\tau) \, \mathrm{d}\tau$ defines an inner product for the real functions z_i and u on the closed domain $[(k-1)T, kT]$. By Cauchy-Schwarz, see Theorem B.3, it holds that

$$|\langle z_i, u \rangle| \leq ||z_i||_2 \, ||u||_2 \,,$$

that is

$$\left|\int_{(k-1)T}^{kT} z_i(\tau)u(\tau)\,\mathrm{d}\tau\right| \leq \underbrace{\sqrt{\int_{(k-1)T}^{kT} |z_i(\tau)|^2\,\mathrm{d}\tau}}_{=:\zeta_{i,k}}\,\underbrace{\sqrt{\int_{(k-1)T}^{kT} u^2(\tau)\,\mathrm{d}\tau}}_{=\sqrt{\gamma_k}}\,.$$

Here it was used that $|u(\tau)|^2 = u^2(\tau)$ since $u(\tau)$ is a real valued function for all $\tau \geq t_0$. Besides, the value of $\zeta_{i,k}$ does not depend on k explicitly for any i. That is, the integral $\int_{(k-1)T}^{kT} |z_i(\tau)|^2\,\mathrm{d}\tau$ equals the same finite value for every $k \in \mathbb{N}_{>0}$. Denote this fixed value by ζ_i. For any i it must therefore hold that

$$\left|\int_{(k-1)T}^{kT} z_i(\tau)u(\tau)\,\mathrm{d}\tau\right| \leq \zeta_i\sqrt{\int_{(k-1)T}^{kT} u^2(\tau)\,\mathrm{d}\tau}.$$

Consider two elements $x_i(kT)$ and $x_i(mT)$ for some $m, k \in \mathbb{N}$ for which without loss of generality $m < k$. Then for the absolute difference between $x_i(kT)$ and $x_i(mT)$ it holds

$$\begin{aligned}|x_i(kT) - x_i(mT)| &= |x_i(kT) - x_i((k-1)T) + x_i((k-1)T) - \cdots + x_i((m+1)T) - x_i(mT)|\\ &\leq |x_i(kT) - x_i((k-1)T)| + |x_i((k-1)T) - x_i((k-2)T)| + \ldots\\ &\quad \ldots + |x_i((m+1)T) - x_i(mT)|\\ &\leq \zeta_i \int_{mT}^{kT} u^2(t)\,\mathrm{d}t\end{aligned}$$

Since $u(t)$ is an L_2 function by Lemma 9, the integral $\int_{mT}^{kT} u^2(t)\,\mathrm{d}t \to 0$ for m and k large enough. That is to say, for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that if $k, m \geq N$ the integral $\int_{mT}^{kT} u^2(t)\,\mathrm{d}t < \varepsilon$. In that case $|x_i(kT) - x_i(mT)| < \varepsilon\zeta_i$ and recall that ζ_i is a fixed and finite value. It follows that for any i the sequence $x_i(kT)$ is a Cauchy sequence of real numbers, and thus converge to some $\bar{x}_i \in \mathbb{R}$. It thereby follows that the sequence of vectors $x(kT)$ converges to some $\bar{x} \in \mathbb{R}^n$ as $k \to \infty$.

The result so far is that the sequence of vectors $x(kT) \xrightarrow{\mathbb{R}^n} \bar{x}$ as $k \to \infty$. It remains to show that $\bar{x} = 0$ and that $x(t) \to \bar{x}$ for all t. It will be proven by contradiction that $\bar{x} = 0$. Hence, assume that $\bar{x} \neq 0$. Since $||x(t)||$ is nonincreasing, there exists a lower bound $c_l > 0$ such that $||x(kT)|| \geq ||\bar{x}|| \geq c_l$ for any k. From (32) then follows that there exists some $\rho > 0$ such that

$$\begin{aligned}&\int_{kT}^{(k+1)T} \Big|[\mathcal{M}_u(kT)](t-kT)\Big|\,\mathrm{d}t =\\ &\qquad \int_{kT}^{(k+1)T} [\mathcal{M}_u^+(kT)](t-kT)\,\mathrm{d}t + \int_{kT}^{(k+1)T} [\mathcal{M}_u^-(kT)](t-kT)\,\mathrm{d}t \geq 3\rho \end{aligned} \tag{39}$$

for all k. Besides, property (36) holds in particular for $\tau = kT$, such that one finds

$$\left|\int_{kT}^{(k+1)T} [\mathcal{M}_u(kT)](t-kT)\,\mathrm{d}t - \int_{kT}^{(k+1)T} [\mathcal{M}_0(kT)](t-kT)\,\mathrm{d}t\right| \to 0 \tag{40}$$

as $k \to \infty$. Since by (33) the integral $\int_{kT}^{(k+1)T} [\mathcal{M}_0(kT)](t-kT)\,\mathrm{d}t = 0$ for all k, expression

(40) reduces to

$$\left|\int_{kT}^{(k+1)T}[\mathcal{M}_u(kT)](t-kT)\,\mathrm{d}t\right| = \left|\int_{kT}^{(k+1)T}[\mathcal{M}_u^+(kT)](t-kT)\,\mathrm{d}t - \int_{kT}^{(k+1)T}[\mathcal{M}_u^-(kT)](t-kT)\,\mathrm{d}t\right| \to 0.$$

Hence for the same $\rho > 0$ that satisfies (39) there exists a $K \in \mathbb{N}$ large enough such that for all $k \geq K$

$$\int_{kT}^{(k+1)T}[\mathcal{M}_u(kT)](t-kT)\,\mathrm{d}t = \int_{kT}^{(k+1)T}[\mathcal{M}_u^+(kT)](t-kT)\,\mathrm{d}t - \int_{kT}^{(k+1)T}[\mathcal{M}_u^-(kT)](t-kT)\,\mathrm{d}t \leq \rho. \quad (41)$$

Combining the results from (39) with (41) yields the inequality

$$\begin{aligned} 3\rho \leq \int_{kT}^{(k+1)T}[\mathcal{M}_u^+(kT)](t-kT)\,\mathrm{d}t + \int_{kT}^{(k+1)T}[\mathcal{M}_u^-(kT)](t-kT)\,\mathrm{d}t \\ \leq \rho + 2\int_{kT}^{(k+1)T}[\mathcal{M}_u^-(kT)](t-kT)\,\mathrm{d}t, \end{aligned}$$

which can be simplified to

$$\int_{kT}^{(k+1)T}[\mathcal{M}_u^-(kT)](t-kT)\,\mathrm{d}t = \int_{kT}^{(k+1)T} u(t)\,\mathrm{d}t \geq \rho$$

for all $k \geq K$. This is a contradiction with $\int_{kT}^{(k+1)T} u(t)\,\mathrm{d}t \to 0$ as $k \to \infty$. And hence we have proven that $\bar{x} = 0$.

It remains to be shown that not only $x(kT) \to 0$ in $\mathbb{R}^n$ as $k \to \infty$, but that also $x(t) \to 0$ in $\mathbb{R}^n$ for all t as $t \to \infty$. In order to do so, note that for any $t \in [kT, (k+1)T]$ it holds that

$$x(t) = \mathrm{e}^{A(t-kT)}x(kT) + \int_{kT}^{t} \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau. \quad (42)$$

It follows from $x(kT) \xrightarrow{\mathbb{R}^n} 0$ as $k \to \infty$ that for all $t \in [kT, (k+1)T]$

$$x(t) \xrightarrow{\mathbb{R}^n} \int_{kT}^{t} \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau$$

as $k \to \infty$. Here for any $t \in [kT, (k+1)T]$ it holds by Cauchy-Schwarz that

$$\left\|\int_{kT}^{t} \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau\right\| \leq \sqrt{\left|\int_{kT}^{t} \mathrm{e}^{A(t-\tau)}B\,\mathrm{d}\tau\right|^2}\sqrt{\int_{kT}^{t} u^2(\tau)\,\mathrm{d}\tau}.$$

Since $t \in [kT, (k+1)T]$ it also holds for any such t that

$$\int_{kT}^{t} u^2(\tau)\,\mathrm{d}\tau \leq \int_{kT}^{(k+1)T} u^2(\tau)\,\mathrm{d}\tau,$$

in which $\int_{kT}^{(k+1)T} u^2(t)\,\mathrm{d}t \to 0$ as $k \to \infty$. Moreover, $\sqrt{\left|\int_{kT}^{t} \mathrm{e}^{A(t-\tau)}B\,\mathrm{d}\tau\right|^2}$ cannot explode for any $t \in [kT, (k+1)T]$, for any k. It follows that

$$\left\|\int_{kT}^{t} \mathrm{e}^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau\right\| \xrightarrow{\mathbb{R}} 0$$

for all $t \in [kT, (k+1)T]$ as $k \to \infty$. But then from (42) follows that $x(t) \xrightarrow{\mathbb{R}^n} 0$ for all $t \in [kT, (k+1)T]$ as $k \to \infty$. Or equivalently $x(t) \to 0$ as $t \to \infty$.

So for the case where $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$ the positive control system is asymptotically stable. □

Proof of Theorem 11 - Part II: $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$. The proof for $\frac{\omega_1}{\omega_2} \in \mathbb{Q}$ uses the fact that in that case $[\mathcal{M}_0(x_0)](t)$ is T-periodic. This is not true if $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$. By Theorem 8 there does exists a finite time-span $\mathcal{T}$ (use $\mathcal{T}$ here instead of T to prevent confusion with T-periodicy) which is large enough such that $B^{\mathrm{T}}Px(t)$ cannot only be positive for $t \in [\tau, \tau + \mathcal{T}]$, for any $\tau \geq t_0$. It will be proven by contradiction that $x(t) \to 0$ as $t \to \infty$.

Assume that $x(t) \not\to 0$ as $t \to \infty$. Since $||x(t)||$ is nonincreasing there exists a lower bound c_l such that $||x(t_0)|| \geq ||x(t)|| \geq c_l$ for all $t \geq t_0$. By (34) there exists an upper bound $\rho > 0$ such that all $\tau \geq t_0$ it holds that

$$\int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_0(x(\tau))](t-\tau)\,\mathrm{d}t \leq \rho.$$

for all $\mathcal{T} > 0$. Then (if $\mathcal{T}$ is fixed and finite) it follows from (36) that

$$\left| \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u(x(\tau))](t-\tau)\,\mathrm{d}t - \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_0(x(\tau))](t-\tau)\,\mathrm{d}t \right| \to 0$$

as $\tau \to \infty$. In that case, given $\mathcal{T}$, exists a $\tau_0 \geq t_0$ large enough such that for $\tau \geq \tau_0$ it hold that

$$\begin{aligned} &\int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u(x(\tau))](t-\tau)\,\mathrm{d}t \\ &\qquad = \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^+(x(\tau))](t-\tau)\,\mathrm{d}t - \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^-(x(\tau))](t-\tau)\,\mathrm{d}t \leq 2\rho. \end{aligned} \tag{43}$$

Fix $\mathcal{T}$ large enough such that $B^{\mathrm{T}}Px(t)$ cannot be only positive on any interval $[\tau, \tau + \mathcal{T}]$ and also large enough such that

$$\begin{aligned} &\int_{\tau}^{\tau+\mathcal{T}} \left| [\mathcal{M}_u(x(\tau))](t-\tau) \right| \mathrm{d}t \\ &\qquad = \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^+(x(\tau))](t-\tau)\,\mathrm{d}t + \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^-(x(\tau))](t-\tau)\,\mathrm{d}t \geq 4\rho \end{aligned} \tag{44}$$

for all τ. Such $\mathcal{T}$ exists according to Theorem 8 and (32). Then combining (43) and (44) yields the inequality

$$\begin{aligned} 4\rho &\leq \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^+(x(\tau))](t-\tau)\,\mathrm{d}t + \int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^-(x(\tau))](t-\tau)\,\mathrm{d}t \\ &\leq 2\rho + 2\int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^-(x(\tau))](t-\tau)\,\mathrm{d}t, \end{aligned}$$

which can be simplified to

$$\int_{\tau}^{\tau+\mathcal{T}} [\mathcal{M}_u^-(x(\tau))](t-\tau)\,\mathrm{d}t = \int_{\tau}^{\tau+\mathcal{T}} u(t)\,\mathrm{d}t \geq \rho$$

for all $\tau \geq \tau_0$. This is a contradiction with $\int_{\tau}^{\tau+\mathcal{T}} u(t)\,\mathrm{d}t \to 0$ as $\tau \to \infty$. So $x(t) \to 0$ as $t \to \infty$ must hold, which implies that also in the case $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$ the positive control system is asymptotically stable. □

Theorem 11 proves that $u(t) = \max\{0, -B^{\mathrm{T}}Px(t)\}$ stabilizes the system $\dot{x}(t) = Ax(t) + Bu(t)$ at $\bar{x} = 0$ where A has purely imaginary eigenvalues, as long as $\omega_1 \neq \omega_2$. As Example 3 shows, this is not generally true if $\omega_1 = \omega_2$. In this case the example shows that $u(t)$ still converges to 0, but $x(t)$ does not, and instead converges to some equilibrium trajectory. As mentioned earlier, the case where $\omega_1 = \omega_2$ is of lesser interest since then $A^2 = -\omega_1^2 I$ and for the controllability matrix (see Theorem 1) it holds that

$$\mathcal{C}_{\{A,B\}} := \begin{bmatrix} B & AB & A^2B & A^3B \end{bmatrix} = \begin{bmatrix} B & AB & -\omega_1^2 B & -\omega_1^2 AB \end{bmatrix},$$

and hence $\mathcal{C}_{\{A,B\}}$ has (at most) rank 2.

The 'controlled' open half space $\{x | B^{\mathrm{T}}Px > 0\}$ hence contains a 2-dimensional controllable subspace Ω_c spanned by the linearly independent columns of $\mathcal{C}_{\{A,B\}}$, i.e.

$$\Omega_c := \textsc{span}\{B, AB\} = \textsc{span}\left\{ \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}, \begin{bmatrix} b_2 \\ -b_1 \\ b_4 \\ -b_3 \end{bmatrix} \right\}.$$

Since $\mathcal{C}$ is not full rank, there also exists a 2-dimensional uncontrollable subspace Ω_{nc} in $\mathbb{R}^4$ spanned by two vectors v_{nc_1} and v_{nc_2}. Since $\Omega_{nc} = \Omega_c^{\perp}$, it holds that for any vector $w \in \Omega_c$ it holds $v_{nc_1} \perp w$ and $v_{nc_2} \perp w$. Both are perpendicular to $B = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 \end{bmatrix}^{\mathrm{T}}$, i.e. $\langle v_{nc_i}, B \rangle = 0$ for $i = 1, 2$. Moreover, for any $v \in \Omega_{nc}$ it holds that $B^{\mathrm{T}}Pv = 0$, and hence the space Ω_{nc} is fully contained in the space $\{x | B^{\mathrm{T}}Px = 0\}$, i.e. $\Omega_{nc} \subset \{x | B^{\mathrm{T}}Px = 0\}$.

The state $x(t)$ in controlled mode can only be stabilized in two of the four directions, movements in the plane Ω_{nc} cannot surpressed. It depends on the initial conditions whether the system is stabilized or not. In general the positive control will stabilize the system in the controllable directions, hence the state is expected to converge to some equilibrium *trajectory* $\bar{x}(t) \in \Omega_{nc}$. So some trajectory for which holds that $B^{\mathrm{T}}P\bar{x}(t) = 0$, that is $B_1^{\mathrm{T}}P_1\bar{x}_1(t) = -B_2^{\mathrm{T}}P_2\bar{x}_2(t)$ for all $t \geq t_0$. Two special cases can be distinguished. Firstly the case where the initial conditions are such that $x_0 \in \Omega_c$, then the system is already stable in its uncontrollable directions, and hence the system is stabilized at 0. Secondly the case where the initial conditions are such that $B_1^{\mathrm{T}}P_1x_1(t_0) = -B_2^{\mathrm{T}}P_2x_2(t_0)$, that is if $B^{\mathrm{T}}Px(t_0) = 0$. Due to the fact that $\omega_1 = \omega_2$ it holds that $B^{\mathrm{T}}Px(t_0) = 0$ for all $t \geq t_0$, and hence also $u(t) = 0$ for all $t \geq t_0$. In this case the initial conditions are in the uncontrollable subspace, $x_0 \in \Omega_{nc}$.

4.2.3. *Generilization for even-dimensional problems*

The control law as described so far for the four-dimensional problem can be extended to $2n$-dimensional systems with n pairs of stable (but not asymptotically) pairs of complex conjugate eigenvalues. In that case

$$E_\lambda(A) = \{\pm\omega_1 \imath, \pm\omega_2 \imath, \ldots, \pm\omega_n \imath\}.$$

In that case the positive definite symmetric matrix P in $V(x) = \frac{1}{2}x^{\mathrm{T}}Px$ extends to an $2n \times 2n$ matrix, composed of n 2×2 positive definite symmetric matrices as

$$P = \begin{bmatrix} P_1 & 0 & \ldots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & P_n \end{bmatrix},$$

such that the Lyapunov function (19) becomes

$$V(x) = \sum_{i=1}^{n} \tfrac{1}{2} x_i^{\mathrm{T}} P_i x_i. \tag{45}$$

Equivalently to Equation (20) for the time derivative of Equation (45) one finds

$$\begin{aligned} 2\dot{V}(x) &= \sum_{i=1}^{n} \dot{x}_i^{\mathrm{T}} P_i x_i + x_i^{\mathrm{T}} P_i \dot{x}_i \\ &= \sum_{i=1}^{n} x_i^{\mathrm{T}} (A_{ii}^{\mathrm{T}} P_i + P_i A_{ii}) x_i + 2 \left(\sum_{i=1}^{n} B_i^{\mathrm{T}} P_i x_i \right) u. \end{aligned}$$

Then there exists a positive definite P_i such that each $A_{ii}^{\mathrm{T}} P_i + P_i A_{ii}$ is equal to the zero matrix. Moreover, for any i any $P_i = c_i I$ for which $c_i > 0$ makes sure that $x_i^{\mathrm{T}}(A_{ii}^{\mathrm{T}} P_i + P_i A_{ii}) x_i = 0$. In that case, analogously to (24) it follows that

$$\dot{V}(x) = (B^{\mathrm{T}} P x) u = \left(\sum_{i=1}^{n} B_i^{\mathrm{T}} P_i x_i \right) u,$$

and yields the similar control law as in (25), namely

$$u(t) = \max\{0, B^{\mathrm{T}} P x\} = \max \left\{ 0, -\sum_{i=1}^{n} B_i^{\mathrm{T}} P_i x_i \right\}.$$

When stability of this control law is concerned the theorems from Section 4.2.2 can be generalized to $2n$-dimensional systems, that is to systems with $E_\lambda(A) = \{\pm\omega_1 \imath, \pm\omega_2 \imath, \ldots, \pm\omega_n \imath\}$.

In general the theorems from the previous section concerning stability make the distinction between $\omega_1 \neq \omega_2$, or $\omega_1 = \omega_2$. The extension to higher dimensions distinguishes $\omega_i \neq \omega_j$ for all $i \neq j$, and the existence of at least one pair i, j $(i \neq j)$ for which $\omega_i = \omega_j$. A similar distinction is made in the proofs of these theorems, where the cases in which $\frac{\omega_1}{\omega_2}$ is rational and irrational are distinguished. If higher dimensions are concerned, this distinction extends to the case where *all* ratios $\frac{\omega_i}{\omega_j} \in \mathbb{Q}$ for all $i \neq j$, and to the case where *at least one* pair i, j $(i \neq j)$ for which $\frac{\omega_1}{\omega_2} \notin \mathbb{Q}$.

Theorem 8 ensures the sign switch of $B^{\mathrm{T}} P x(t)$ from positive to negative within finite time, and hence making sure that the system switches from uncontrolled to controlled mode. The generalization of which is given by the following theorem.

Theorem (Extension of Theorem 8) *$B^{\mathrm{T}} P x(t)$ switches sign from positive to negative within finite time if $\omega_i \neq \omega_j$ $\forall i \neq j$. If for any $i \neq j$ it holds $\omega_i = \omega_j$ then the sign switch is ensured with the additional condition that $B_i^{\mathrm{T}} P_i x_i \neq -B_j^{\mathrm{T}} P_j x_j$ initially.*

Assuming that the initial state x_0 is such that $B^{\mathrm{T}} P x_0 > 0$, then if all ratios $\frac{\omega_i}{\omega_j}$ are rational then the proof is analogous to that of Theorem 8. Indeed, if one defines $T_i = \frac{2\pi}{\omega_i}$, there exist $k_1, k_2, \ldots, k_n \in \mathbb{N}_{>0}$ such that

$$k_1 T_1 = k_2 T_2 = \cdots = k_n T_n.$$

Then, since $x(t)$ is always determined by $\dot{x}(t) = Ax(t)$, the $B^{\mathrm{T}} P x(t)$ is T-periodic with period $T = 2m\pi$, where $m = \mathrm{lcm}(k_1, k_2, \ldots, k_n)$. For the case where at least one of the ratios $\frac{\omega_i}{\omega_j} \notin \mathbb{Q}$ one should be more careful in the proof. Consider the following case with ω_1, ω_2 and ω_3. Then, again assuming that $B^\mathrm{T}Px(t)$ does not switch sign, one can write

$$B^\mathrm{T}Px(t) = \cos(\omega_1 t) + \gamma_2\cos(\omega_2 t + \phi_2) + \gamma_3\cos(\omega_3 t + \phi_3),$$

where without loss of generality assume that for the periods of the individual cosine terms it holds that $T_3 < T_2 < T_1$. Then a known result is that every interval of length $\frac{1}{2}T_2$ in which $\gamma_2\cos(\omega_2 t + \phi_2) < 0$ contains one or more subintervals on which also $\gamma_3\cos(\omega_3 t + \phi_3) < 0$. Denote each of these intervals on which $\gamma_2\cos(\omega_2 t + \phi_2) < 0$ by

$$\mathcal{U}_{2_n} := \left(\frac{(4n+1)\pi}{2\omega_2} - \phi_2, \frac{(4n+3)\pi}{2\omega_2} - \phi_2\right), \quad n \geq 0.$$

Similarly, $\cos(\omega_1 t) < 0$ on any $\mathcal{U}_{1_k} := \left(\frac{(4k+1)\pi}{2\omega_1}, \frac{(4k+3)\pi}{2\omega_1}\right)$, $k \geq 0$. Now since $\omega_1 \neq \omega_2 \neq \omega_3$, there exist finite $\tilde{k}, \tilde{n}$ such that $\mathcal{U}_{2_{\tilde{n}}} \subset \mathcal{U}_{1_{\tilde{k}}}$. Hence, within this $\mathcal{U}_{1_{\tilde{k}}}$ there is a nonempty subinterval in which all cosine terms of $B^\mathrm{T}Px(t)$ are negative, and hence $B^\mathrm{T}Px(t)$ itself must be negative. So in that case a sign switch is guaranteed within $\frac{(4\tilde{k}+3)\pi}{2\omega_1}$ time. A similar reasoning goes for an additional fourth ω_4, and a fifth, etcetera.

Lemma 9 does not assume dimension, and hence holds similarly for $E_\lambda(A) = \{\pm\omega_1\imath, \pm\omega_2\imath, \ldots, \pm\omega_n\imath\}$, the same goes for Theorem 11 but with the condition that $\omega_i \neq \omega_j$ for all indices $i \neq j$. In the proof one should distinguish the same cases as in the extension of Theorem 8. The theorem then reads as follows:

Theorem (Extension of Theorem 11) *The system $\dot{x} = Ax + Bu$ with eigenvalues $E_\lambda(A) = \{\pm\omega_1\imath, \pm\omega_2\imath, \ldots, \pm\omega_n\imath\}$, with $\omega_i \neq \omega_j$ for all $i \neq j$, is stabilized by the positive control input $u(t)$ defined by (25) at equilibrium $\bar{x}(t) = 0$.*

The proof is similar to that of Theorem 11 since the statements hold irrespectively of the dimension n.

4.3. Unstable systems

So far only systems were considered for which the eigenvalues have their real part equal to zero, that is systems for which $E_\lambda(A) = \{\pm\omega_1\imath, \pm\omega_2\imath\}$. In this small section is dedicated to systems with one or more unstable poles. For these systems the eigenvalues of the system matrix are given by

$$E_\lambda(A) = \{\sigma_1 \pm \omega_1\imath, \sigma_2 \pm \omega_2\imath\},$$

where $\sigma_1, \sigma_2 \geq 0$, but not $\sigma_1 = \sigma_2 = 0$. Such matrices are assumed to be of the form

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad \text{where} \quad A_{11} = \begin{bmatrix} \sigma_1 & \omega_1 \\ -\omega_1 & \sigma_1 \end{bmatrix} \quad \text{and} \quad A_{22} = \begin{bmatrix} \sigma_2 & \omega_2 \\ -\omega_2 & \sigma_2 \end{bmatrix}. \tag{46}$$

The main conclusion of this section is that the control law from Section 4.2 for stable systems does in general not stabilize systems with unstable poles. Whenever the system is in uncontrolled mode (11a) it cannot be prevented that $||x||$ increases. This is a direct consequence of the eigenvalues of A having a positive real part. Anyhow, the increase of $||x||$ during uncontrolled mode should be compensated by the control (11b) in controlled mode.

The approach of Section 4.2 is based upon finding a Lyapunov function for the entire system, that is for the controlled mode and uncontrolled mode. For $V(x)$ to be a Lyapunov function it should be nonincreasing. Now, with increase of $||x||$, also $V(x)$ will increase. This forebodes that the approach of Lyapunov for unstable systems and the stability theorems from Section 4.2.2 are not feasible in general for systems with unstable eigenvalues. Consider the Lyapunov function $V(x) = \frac{1}{2}x^{\mathrm{T}}Px$ with time derivative copied from (20):

$$2\dot{V}(x) = \underbrace{x_1^{\mathrm{T}}(A_{11}^{\mathrm{T}}P_1 + P_1A_{11})x_1 + x_2^{\mathrm{T}}(A_{22}^{\mathrm{T}}P_2 + P_2A_{22})x_2}_{=:\mathcal{V}_P(x)} + \underbrace{2(x_1^{\mathrm{T}}P_1B_1 + x_2^{\mathrm{T}}P_2B_2)u}_{=:\mathcal{V}_{u,P}(x)},$$

It is easily seen that for the conditions (22) there exist no symmetric positive definite matrices P_1 and P_2 that yield positive definite Q_1 and Q_2. This is a direct result of Theorem 7. Furthermore, there neither exist symmetric positive definite matrices P_1 and P_2 such that $A_{ii}^{\mathrm{T}}P_i + P_iA_{ii} = 0$, $i = 1, 2$. An important conclusion is that the function $\mathcal{V}_P(x)$ cannot be made equal to 0 or negative for all x. Moreover, its value is always nonnegative since $\mathcal{V}_P(x)$ is quadratic in x. In terms of this Lyapunov based control, the 'best that can be done' is at least have $u = 0$ if $B^{\mathrm{T}}Px > 0$. This way, by keeping the switching plane unaltered, at least the increase of V is minimized. The problem for unstable systems is anyhow not trivial in the Lyapunov control design.

Consider the control input $u = \max\{0, -B^{\mathrm{T}}Px\}$. Once the system enters controlled mode, $u(t)$ should be large enough to first of all compensate the increase of $V(x)$ that occured during uncontrolled mode. In addition to that, whenever $u > 0$ it should at least make sure that $\mathcal{V}_{u,P}(x) > \mathcal{V}_P(x)$. It is expected that the original control law has some additional strength, so that it can cope with small instabilities, that is small $\sigma_i > 0$. For larger rates of instability the strenght of u may be increased by increasing c_1 and/or c_2 in Equation (23). Note that this changes the matrices P_1 and/or P_2, and hence also changes the switching plane $B^{\mathrm{T}}Px = 0$. Another option is to amplify the control via some $\gamma > 1$, and apply the control

$$u(t) = \gamma \max\{0, -B^{\mathrm{T}}Px(t)\}. \tag{47}$$

One of the pitfalls of increasing the magnitude of the control via γ or via c_1, c_2 as described is that it may make the system leave controlled mode very fastly. Hence the 'strenghthening' of u is in that case counteracted by the short time span in which the control is applied. Moreover, in contrast to increasing γ, increasing c_1 and c_2 also works through in the $A_{ii}^{\mathrm{T}}P_i + P_iA_{ii}$ terms in $\mathcal{V}_P(x)$ such that a similar counteraction occurs between $\mathcal{V}_{P,u}$ and $\mathcal{V}_P$ in (20).

4.3.1. *Application in simulations*

The presumptions described so far are supported by simulations in the examples that follow in this section.

Example 4. This first example considers a system with only small instability. Let the system matrix be given by Equation (46), with $\sigma_1 = 0.1$, $\omega_1 = 1$, $\sigma_2 = 0.2$ and $\omega_2 = 3$, such that $E_\lambda(A) = \{0.1 \pm \imath, 0.2 \pm 3\imath\}$. For this example let $B = \begin{bmatrix} 0 & -1 & 0 & -1 \end{bmatrix}^{\mathrm{T}}$ and initial condition $x_0 = \begin{bmatrix} 1 & 2 & 1 & 0 \end{bmatrix}$. For this example take $P_1 = P_2 = I$. Figures 17 and 18 show that the control manages to bring the system to rest. Figure 18 clearly shows that V shows ascending and increasing behaviour, but eventually goes to zero as time progresses.

Also the result of a simulation of the same system is included, with the exeption that $\sigma_1 = \sigma_2 = 0$, in the plots indicated by '(st)'. This shows that the control u needs to work a lot harder to in the case when $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$. △

The result of the simulation of the previous example nicely shows that for unstable systems the function $V(x)$ is not (necessarily) non-increasing. The following example is similar to the previous, but with a system that is slightly more unstable.

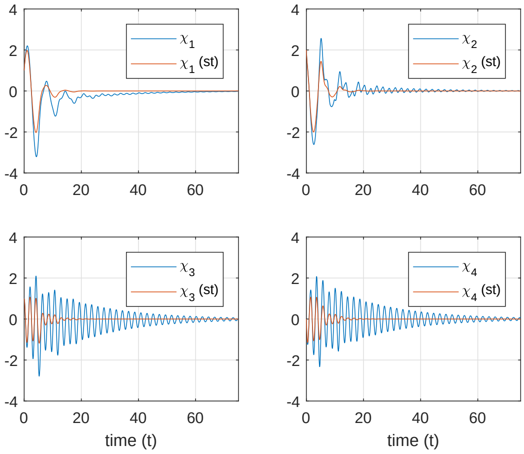

Figure 17: Simulation results for state trajectories x.

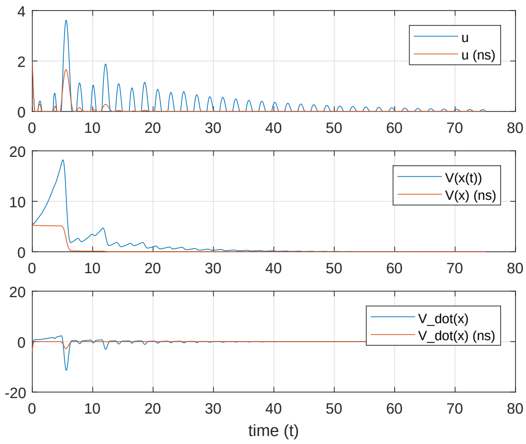

Figure 18: Simulation results for u and $V(x)$ and $\dot{V}(x)$.

Example 5. Consider the system from the previous example with $\sigma_1 = 0.2$ and $\sigma_2 = 0.3$, so the system is slightly more 'unstable'. As Figures 19 and 20 show, the standard control with $P_1 = P_2 = I$ cannot overpower these slightly increased σ_1 and σ_2.

Setting $\gamma = 2$ in Equation (47), and hence doubling the control without altering the switching plane yields a stable result, see Figures 22 and 22. Now for $\gamma = 4$ the results of the simulation are included in Figures 23 and 24. Apparently this additional increase of γ counterproductive. $\triangle$

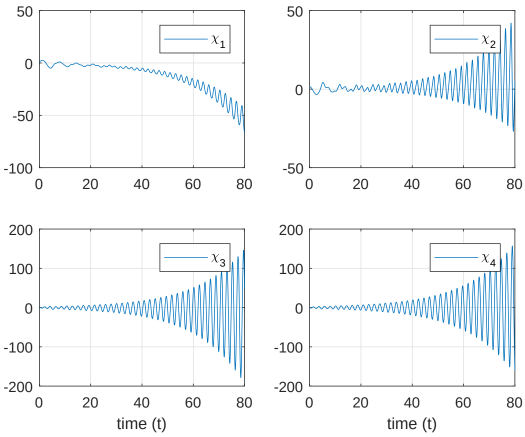

Figure 19: Simulation results for state trajectories x.

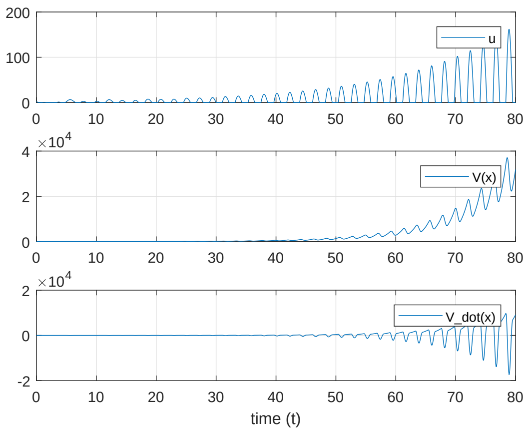

Figure 20: Simulation results for u and $V(x)$ and $\dot{V}(x)$.

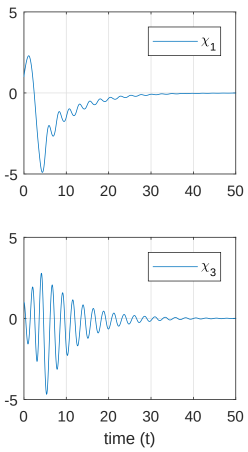

Figure 21: Simulation results for state trajectories x.

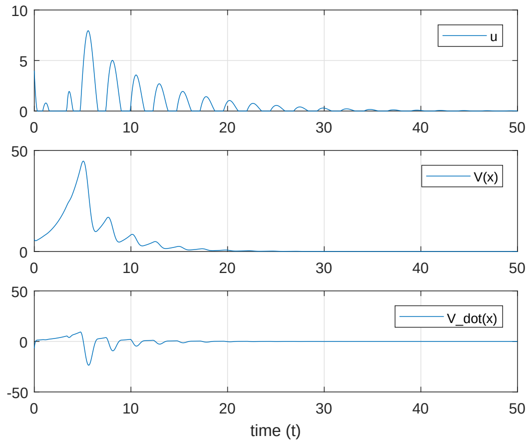

Figure 22: Simulation results for u and $V(x)$ and $\dot{V}(x)$.

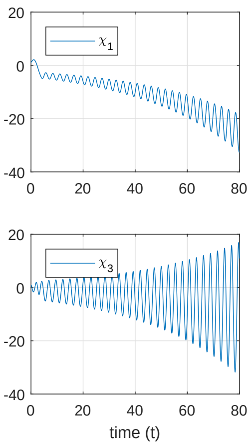

Figure 23: Simulation results for state trajectories x.

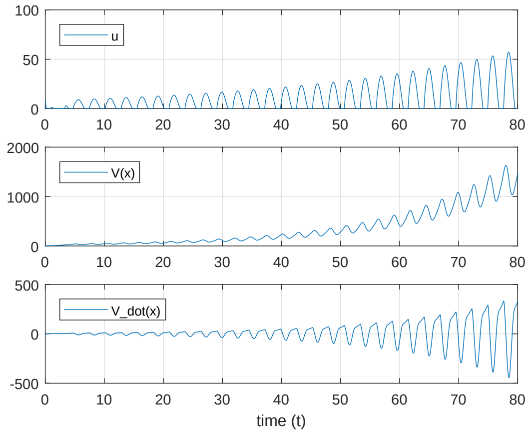

Figure 24: Simulation results for u and $V(x)$ and $\dot{V}(x)$.

Example 5 indicates that increasing u through γ may work. It should be noted that several simulations have shown that this does not work in general.

4.3.2. *Remarks concerning unstable systems*

The section on unstable systems is concluded with a few remarks. First of all the examples and the reasoning as previously described indicate that the results achieved for stable systems (but not asymptotically) in Section 4.2 do not hold for systems with unstable poles in general. If the control framework (47) is to be applied to unstable systems, then more research is needed. For example the role of γ, c_1 and c_2 on stability of the control system could be analyzed. Also more research into the behaviour of $V(x)$ and $\dot{V}(x)$ could be done in general. For example can conditions on P be conjured up such that $\mathcal{V}_P(x(t))$ is in some sense (a lot) smaller than $\mathcal{V}_{u,P}(x(t))$.

Also different classifications of instability could be considered. For example systems for which only one subsystem has unstable poles. In that case $\sigma_1 = 0$ and $\sigma_2 > 0$ for example. One could also distinguish between different levels of instability. For example systems for which the subsystems show a similar level of instability, that is $\sigma_1 \approx \sigma_2$. Or systems with different levels of instability, $\sigma_1 \not\approx \sigma_2$.

5. State feedback via singular perturbations

This part of the report approaches the main problem using the techniques of singular perturbations for ordinary differential equations. Section 5.1 introduces this theory in general. A detailed description of how the theory would apply to four-dimensional problem is given in Section 5.2. Section 5.3 shows the results of a series of simulations in which the control law is based upon the theory of singular perturbations. Comments on these simulations and on the applicability of the techniques described are given in Section 5.4.

The general idea of singular perturbations is best illustrated as follows. Consider the controlled system equation (11b). As mentioned earlier in Section 3.1, the subsystems are individually 'disturbed' by cross terms as is clear from the following notation:

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + B_1F_1x_1 \quad + \quad B_1F_2x_2, \\ \dot{x}_2 &= A_{22}x_2 + B_2F_2x_2 \quad + \quad \underbrace{B_2F_1x_1}_{\text{disturbances}} \ . \end{aligned}$$

Recall that if the systems for x_1 and x_2 could be separated or decoupled fully, then the subsystems could be stabilized separately with conditions provided by Willems et al. [29]. Since $u(t)$ is scalar, the subsystems can never be decoupled fully, after all the cross terms $B_1F_2x_2$ and $B_2F_1x_1$ cannot be eliminated. Singular perturbations aim to separate the systems *in time* by creating subsystems that operate on different time scales. This way, the systems can be transformed into a *slow* and a *fast* system. In the four-dimensional problem the scalar control $u(t)$ could, for example, first be focused on stabilizing the state-set x_1 and thereafter on stabilizing x_2.

5.1. Theory of singular perturbations

An illustrative example of the use of singular perturbation techniques is that of an electrically driven robot which has slower mechanical dynamics and faster electrical dynamics. It therefore exhibits two time scales. In such cases the system can be divided into two subsystems, one describing the faster dynamics, the other the slower dynamics. A controller can then be designed separately for both subsystems. Through a singular perturbation technique the two systems can be made independent of each other, thereby simplifying the control problem.

The theory of singular pertubations has been studied extensively during the nineteen-seventies and -eighties. Literature can be found in the form of books such as Naidu [18] or O'Malley Jr. et al. [19], and even books with clear educational purposes such as Kokotovic et al. [15]. Extensive surveys and overviews of research concerning singular perturbations were published by Kokotovic et al. [14] and Saksena et al. [24]. The latter paper states that singular perturbation theory can be of use to lessen the model order by first neglecting the fast dynamics. It then modifies the considered approximation by reintroducing its effects as boundary layer corrections computed in separate time scales. In general, singular perturbation techniques is a well established field of research which has a broad scope of applications, for example in aerospace, mechanical systems, electrical and electronic circuits and systems, chemical reactions and biology, see Zhang et al. [31]

Consider the class of subsystems described by

$$\dot{x} = f_1(x,z,t) + \varepsilon g_1(x,z,\varepsilon,t), \tag{48a}$$
$$\varepsilon\dot{z} = f_2(x,z,t) + \varepsilon g_2(x,z,\varepsilon,t), \tag{48b}$$

with $x(0) = x_0$ and $z(0) = z_0$ and $0 < \varepsilon \ll 1$. By ε being small, the system (48) indicates that the dynamics of x is much faster than that of z. By 'slow' and 'fast' is meant that $||\dot{z}|| \gg ||\dot{x}||$, such that $||\varepsilon\dot{z}|| \approx ||\dot{x}||$. There is a theorem by Tikhonov that states that, with the correct conditions on the system, it will initially and very quickly approximate the solution to the equations

$$\begin{aligned} \dot{x} &= f_1(x,z,t), \quad x(0) = x_0, \\ 0 &= f_2(x,z,t) \end{aligned}$$

on some interval of time. As ε decreases toward 0, the system will approach the solution more closely in that same interval.

5.2. Application to the four-dimensional problem

In the case at hand, an LTI system is considered which switches between controlled mode $\dot{x} = (A + BF)x$ and uncontrolled mode $\dot{x} = Ax$. Since only the controlled mode yields an opportunity to influence the system through the control input, this will be the main focus for now. According to Equation (11b) the system in controlled mode can be represented by

$$\dot{x}_1 = (A_{11} + B_1F_1)x_1 + B_1F_2x_2 \tag{49a}$$
$$\dot{x}_2 = (A_{22} + B_2F_2)x_2 + B_2F_1x_1 \tag{49b}$$

Here, matrices A and B are given and the feedback matrix F is free to choose at this point. If a singular perturbed system of the form (48) is to be created, it should be achieved through the feedback F, resulting in the decoupling of the state-sets x_1 and x_2. If then the separate systems (48a) and (48b) can also be made stable by a suitable choice of F, then the stabilization of the system with the positive control $u(t) = \max\{0, Fx(t)\}$ may be succesful. Be aware that such a decoupling can only be achieved and possibly also be maintained as long as the system is in controlled mode.

Consider (49a) and consider the state feedback matrix $F = \begin{bmatrix} F_1 & F_2 \end{bmatrix}$ with the property that $F_1 = O(\frac{1}{\varepsilon})$ and $F_2 = O(1)$, for some $\varepsilon > 0$. In that case the matrices $\widetilde{F}_1 = \varepsilon F_1$ and F_2 are of the same order.

The subsystem (49a) can be rewritten as

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + \tfrac{1}{\varepsilon}B_1\widetilde{F}_1x_1 + B_1F_2x_2 \\ \varepsilon\dot{x}_1 &= \varepsilon A_{11}x_1 + B_1\widetilde{F}_1x_1 + \varepsilon B_1F_2x_2 \\ \varepsilon\dot{x}_1 &= B_1\widetilde{F}_1x_1 + \varepsilon(A_{11}x_1 + B_1F_2x_2) \end{aligned} \tag{50}$$

where the last equation is of the form (48b). If $F_1 = \frac{1}{\varepsilon}\widetilde{F}_1$ is substituted into (49b) one gets

$$\begin{aligned} \dot{x}_2 &= A_{22}x_2 + B_2F_2x_2 + \tfrac{1}{\varepsilon}B_2\widetilde{F}_1x_1 \\ &= (A_{22} + B_2F_2)x_2 + \varepsilon(\tfrac{1}{\varepsilon^2}B_2\widetilde{F}_1x_1) \end{aligned} \tag{51}$$

which is of the form (48a). It is important to make a comment on what is meant by '$F_1 = O(\frac{1}{\varepsilon})$', and more importantly what it is supposed to achieve. The aim of $F = \begin{bmatrix} F_1 & F_2 \end{bmatrix}$ is to create a slow and a fast subsystem. Fastness of a system is indicated by the real parts of its poles. The goal of F_1 and F_2 is to make sure that poles of the fast system for x_1 are indeed a lot faster than the poles of the slow system for x_2. The arguments of '$F_1 = O(\frac{1}{\varepsilon})$' and '$F_2 = O(1)$' are used to come to the forms (50) and (51).

Note that the norm of F_1 is a way of quantifying the 'order of F_1'. Say we wish to place both poles of the controlled system at $\frac{1}{\varepsilon}$, such that the system can be made faster by making $\varepsilon > 0$ smaller. Then the characteristic polynomial for $A_{11} + B_1F_1$ equals

$$p(s) = (s + \tfrac{1}{\varepsilon})^2 = s^2 + 2\tfrac{1}{\varepsilon}s + \tfrac{1}{\varepsilon^2}.$$

If for example the system is described by

$$A_{11} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{and} \quad F_1 = \begin{bmatrix} f_1 & f_2 \end{bmatrix},$$

then

$$p_{A_{11}+B_1F_1}(s) = \det\left(\begin{bmatrix} s & -1 \\ 1 - f_1 & s - f_2 \end{bmatrix}\right) = s^2 - f_2 s + 1 - f_1.$$

So for this case choose $f_1 = -\frac{2}{\varepsilon}$ and $f_2 = 1 - \frac{1}{\varepsilon^2}$. Obviously, making ε small, directly increases $||F_1||$.

5.2.1. *Closer inspection of slow and fast system*

Consider again the system given by Equation (49) with F_1 and F_2 with $F_1 = O(\frac{1}{\varepsilon})$ and $F_2 = O(1)$ with $\varepsilon > 0$ small enough such that time sepatation into a slow and a fast system is obtained to some level. This way if x_1 and x_2 are of similar magnitude, the contribution of F_1x_1 in Equation (49a) is much more significant than the contribution of F_2x_2. In other words, the control $u(t)$ puts a much larger weight on x_1 than on x_2. The subsystem for x_1 can be seen as being of the following form

$$\dot{x}_1 = O(\tfrac{1}{\varepsilon})x_1 + O(1)x_2. \tag{52}$$

Assuming that x_1 and x_2 are of similar magnitude, the dynamics of x_1 are initially dominant in (52). Therefore it is argued that in terms of control design (i.e. the stabilization of the fast system) the dynamics for x_1 are approximated as

$$\dot{x}_1 = (A_{11} + B_1F_1)x_1 + B_1F_2x_2 \approx (A_{11} + B_1F_1)x_1. \tag{53}$$

This way, the state feedback F_1 for the fast system be chosen such that the fast system (49a) stabilizes very quickly, that is so that $\dot{x}_1 \to 0$ quickly. Therefore it follows for the fast state-set that

$$x_1 \to -(A_{11} + B_1F_1)^{-1}B_1F_2x_2 =: Mx_2 \tag{54}$$

so that it is expected to converge to a function of the states x_2 only. Since the dynamics of x_1 and x_2 operate in different time-scales, that is x_1 has much faster dynamics than x_2, the function Mx_2 is approximately constant on the time scale of x_1, that is to say that $\dot{x}_1 \approx 0$. Note that since F_1 is $O(\frac{1}{\varepsilon})$, we have that $(A_{11} + B_1F_1)^{-1}$ is of order ε. Therefore the matrix M defined in (54) is of order ε too. Since x_1 converges to an $O(\varepsilon)$ term, it should be mentioned that the dominancy of x_1 in (52) diminishes as $x_1 \to Mx_2$.

An important aspect to realize is that if $x_1 = -(A_{11} + B_1F_1)^{-1}B_1F_2x_2$, then $x_2 \to 0$ implies that $x_1 \to 0$ if M is nonsingular. Therefore, the remaining degrees of freedom in F_2 should be focussed on stabilizing x_2, at $x_2 = 0$. Substituting $x_1 = -(A_{11} + B_1F_1)^{-1}B_1F_2x_2$ into the slow system Equation (49b) yields

$$\begin{aligned} \dot{x}_2 &= A_{22}x_2 + B_2F_2x_2 + B_2F_1x_1, \\ &= A_{22}x_2 + B_2F_2x_2 - B_2F_1(A_{11} + B_1F_1)^{-1}B_1F_2x_2, \\ &= A_{22}x_2 + \left(B_2 - B_2F_1(A_{11} + B_1F_1)^{-1}B_1\right)F_2x_2, \\ &= A_{22}x_2 + \widetilde{B}_2\widetilde{u}. \end{aligned} \tag{55}$$

In the last expression (55) define $\widetilde{B}_2 := \left(B_2 - B_2F_1(A_{11}+B_1F_1)^{-1}B_1\right)$ and a control input $\widetilde{u} = F_2x_2$. Hence a new system

$$\dot{x}_2 = (A_{22} + \widetilde{B}_2F_2)x_2, \tag{56}$$

is obtained where A_{22} and $\widetilde{B}_2$ are explicitly given. The feedback matrix F_2 should be chosen such that it stabilizes the system. Stablilizing (56) at $x_2 = 0$ should make sure that $x_1 \to 0$ as well, as was mentioned before.

Since $x_1 \to -(A_{11}+B_1F_1)^{-1}B_1F_2x_2$, the control input $u = F_1x_1 + F_2x_2$ is expected to converge too. If the converged expression for x_1 is substituted one finds that

$$\begin{aligned}
u &\to -F_1(A_{11}+B_1F_1)^{-1}B_1F_2x_2 + F_2x_2,\\
&= \left(F_2 - F_1(A_{11}+B_1F_1)^{-1}B_1F_2\right)x_2,\\
&= \left(I - F_1(A_{11}+B_1F_1)^{-1}B_1\right)F_2x_2,\\
&=: \widetilde{F}_2x_2.
\end{aligned}$$

Note that the systems $\dot{x}_2 = A_{22}x_2 + \widetilde{B}_2F_2x_2$ from Equation (55) and $\dot{x}_2 = A_{22}x_2 + B_2\widetilde{F}_2x_2$ are equal expressions.

Concluding: if F_1 is chosen such that the system $\dot{x_1} = (A_{11}+B_1F_1)x_1$ has fast stable poles, and F_2 makes sure that $\dot{x}_2 = (A_{22}+\widetilde{B}_2F_2)x_2$ has stable poles (but 'a lot slower'), then the dynamics (49) are expected to converge to

$$\begin{aligned}
\dot{x}_2 &= A_{22}x_2 + \widetilde{B}_2F_2x_2, \quad \widetilde{B}_2 = B_2\left(I - F_1(A_{11}+B_1F_1)^{-1}B_1\right),\\
x_1 &= -(A_{11}+B_1F_1)^{-1}B_1F_2x_2.
\end{aligned}$$

5.2.2. *Back to positive control*

So far the focus was solely on the controlled mode. The control systems that are considered here are systems which switch between controlled mode and uncontrolled mode. Therefore the convergence as described above may only take place once the system enters the controlled mode, and only holds as long as the system remains in controlled mode. Convergence of $x_1 \to Mx_2$ may very well be undone once the system enters the non-controlled mode again.

Note that $F_1x_1 + F_2x_2$ switches sign from negative to positive according to a similar reasoning as in the proof of Theorem 8 in the part on Lyapunov. In this case the similar expression to (26) is that $F_1x(t) + F_2x(t)$ can be written as

$$\gamma_1 e^{\sigma_1 t}\cos(\omega_1 t + \phi_1) + \gamma_2 e^{\sigma_2 t}\cos(\omega_2 t + \phi_2).$$

where $\gamma_1, \gamma_2 \in \mathbb{R}$ (both nonzero) and $\phi_1, \phi_2 \in [0, \pi]$ follow from initial conditions. Anyhow, the dynamics cannot remain stuck in controlled mode.

Due to the possible switching behaviour of the positive control system, the system's dynamics (with the control input designed as described) may be described by either of the following classes of dynamics, denoted by i, ii and iii:

i. *No control:* $F_1x_1 + F_2x_2 \le 0$ and hence $u = 0$.
   In this case the system dynamics are described by

$$\begin{aligned}
\dot{x}_1 &= A_{11}x_1, \quad x_1(t_i) = x_{1_i},\\
\dot{x}_2 &= A_{22}x_2. \quad x_2(t_i) = x_{2_i}.
\end{aligned}$$

   Here t_i denotes the time instance at which this class is entered. The system's dynamics remain described by these equations until $F_1x_1 + F_2x_2$ becomes positive.

ii. *Fast system is dominant:* $u = F_1x_1 + F_2x_2 > 0$.
While the fast system is dominant, the systems dynamics are given by

$$\dot{x}_1 = (A_{11} + B_1F_1)x_1 + B_1F_2x_2 \approx (A_{11} + B_1F_1)x_1, \quad x_1(t_{ii}) = x_{1_{ii}},$$
$$\dot{x}_2 = A_{22}x_2 + B_2F_1x_1 + B_2F_2x_2. \quad x_2(t_{ii}) = x_{2_{ii}}.$$

Here t_{ii} denotes the time instance at which this class is entered. Note that here the control input is dominated by F_1x_1, so that one may say that $u = F_1x_1 + F_2x_2 \approx F_1x_1$. During this state the control F_1 should make sure that $x_1 \to -(A_{11} + B_1F_1)^{-1}B_1F_2x_2$, while the controller does not focus on stabilizing the dynamics for x_2.

iii. *Slow system dynamics:* $u = F_1x_1 + F_2x_2 > 0$.
Here it is assumed that the dynamics have converged to the following system:

$$\dot{x}_2 = A_{22}x_2 + \widetilde{B}_2F_2x_2, \quad \widetilde{B}_2 = B_2\left(I - F_1(A_{11} + B_1F_1)^{-1}B_1\right), \quad x_2(t_{iii}) = x_{2_{iii}},$$
$$x_1 = -(A_{11} + B_1F_1)^{-1}B_1F_2x_2 =: Mx_2, \quad x_1(t_{iii}) = x_{1_{iii}}.$$

Here t_{iii} denotes the time instance at which this class is entered.

In any case, the application of the techniques of singular perturbations reduces the four-dimensional control problem to the following subproblems:

1. Choose F_1 such that system $\dot{x}_1 = (A_{11} + B_1F_1)x_1$ is stable and fast. This can be achieved by choosing F_1 such that the eigenvalues of $A_{11} + B_1F_1$ have a large negative real part. Notice that the choice of F_1 is a two-dimentional control problem with positive controls.

2. Choose F_2 such that the system $\dot{x}_2 = (A_{22} + \widetilde{B}_2F_2)x_2$ is stable, where $\widetilde{B}_2 = B_2\left(I - F_1(A_{11} + B_1F_1)^{-1}B_1\right)$. Note that this is also a two dimensional problem with positive controls.

3. Make sure that the fast subsystem is 'fast enough' compared to the slow system.

Applying singular perturbations poses an opportunity to apply the techniques as described in Willems et al. [29] to the two separated two-dimensional subproblems. On the one hand F_1 is chosen to place the poles of $A_{11} + B_1F_1$, but the subtle difference is that F_2 should be chosen to place the poles of $A_{22} + B_2\left(I - F_1(A_{11} + B_1F_1)^{-1}B_1\right)F_2$, and not those of $A_{22} + B_2F_2$ as was hinted upon in Section 3.1.

The system's dynamics may change between the classes described earlier. For example, the system may start in class i, then switch to ii once control is applied, and then (quickly) switch to iii. This is just one of the examples of possible transitions. In general, the system's dynamics may evolve according to the diagram in Figure 25.

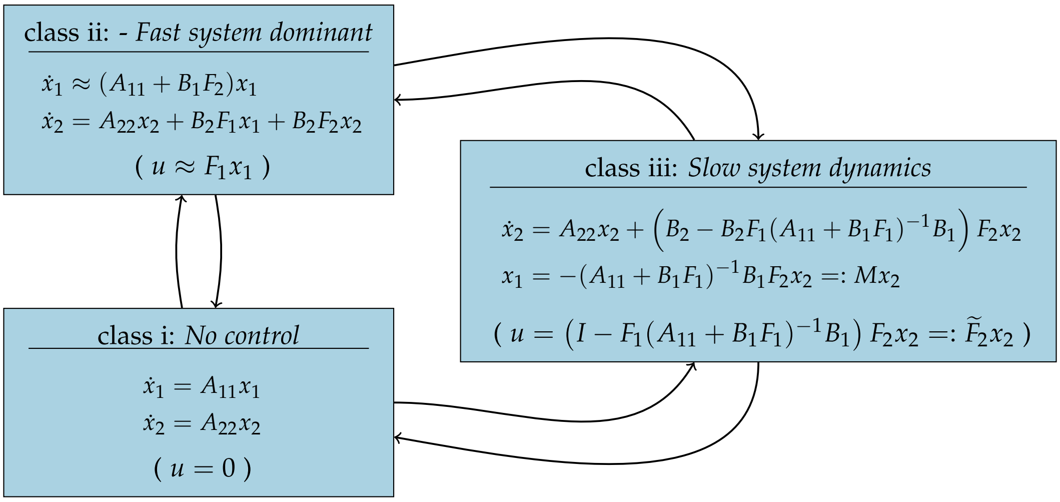

Figure 25: Diagram of the switching between the three classes of dynamics.

What follows is a brief description of all transitions between the classes of dynamics as depicted in Figure 25.

(i $\longrightarrow$ ii) The system is currently in uncontrolled mode, and once u becomes positive, it makes the transition to controlled mode where the fast system is dominant. This is most likely to happen.

(ii $\longrightarrow$ iii) It is desired that the dynamics quickly makes the transition to the singularly perturbed dynamics, this is mainly due to a suitable choice of F_1.

(iii $\longrightarrow$ ii) This is not desired. If such a transition occurs, then the choices of F_1 and F_2 should be reconsidered.

(ii $\longrightarrow$ i) It may happen that the dynamics make the transition back to i before it could make the transition to iii. Such a transition is undesirable.

(iii $\longrightarrow$ i) This transition is likely to happen if for example if $A_{22} + \widetilde{B}_2 F_2$ has oscillatory poles.

(i $\longrightarrow$ iii) This transition would be alright, but probably will not occur. More likely is that such a transition from i to iii will occur via ii.

It is important to realize that in class ii the control mainly focusses on bringing the system to the representation of class iii. By doing so, $||x_2||$ will most likely be increased (destabilizing x_2) before the control focusses on stabilizing x_2. While the system's dynamics are described by iii, F_1 focusses on maintaining $x_1 = -(A_{11} + B_1 F_1)^{-1} B_1 F_2 x_2$, while F_2 focusses on stabilizing x_2. Wrong choice of F_1 and F_2 may lead to cycling through the diagram of Figure 25 endlessly. Moreover, during this cycling any increase of $||x_2||$ in class ii should at least be compensated in class iii before the end of the 'cycle'.

5.3. Application in simulations

The aim of this section is to substantiate the applicability of techniques of singular perturbations to systems with positive control. It consideres various examples to illustrate its effectiveness. It should be stressed that this section does not aim to formally prove stability of positive control systems with techniques from singular perturbations.

Example 6. This first example is included to illustrate how x_1 is made to converge quickly to Mx_2 due to a suitable choice of F_1. The results of a simulation are included for

$$A_{11} = \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix}, \quad B_1 = B_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}^{\mathrm{T}},$$

hence $E_\lambda(A_{11}) = \pm 3\imath$. For this simulation the states x_2 do not receive control input, and x_3 is remained kept at a standard sinusoid of frequency $\frac{1}{4}$, and x_4 is computed as its derivative. Poles of the fast system $A_{11} + B_1 F_1$ are placed at placed at -10 and -10. The results are included in Figure 26. Here, the black line at the horizontal axis indicates when control is applied to the fast system. Figure 26 clearly shows that whenever control is applied, x_1 quickly starts following the path of Mx_2. This path is denoted by 'conv' in the legend of the plots. Another aspect that follows from Figure 26 is that as soon as control is cut-off, x_1 evolves according to the dynamics $\dot{x}_1 = A_{11}x_1$ (with the initial conditions at the time the control was cut off). Be aware of the different time scaling of the vertical axes for x_1 and x_2 versus that of x_3 and x_4. △

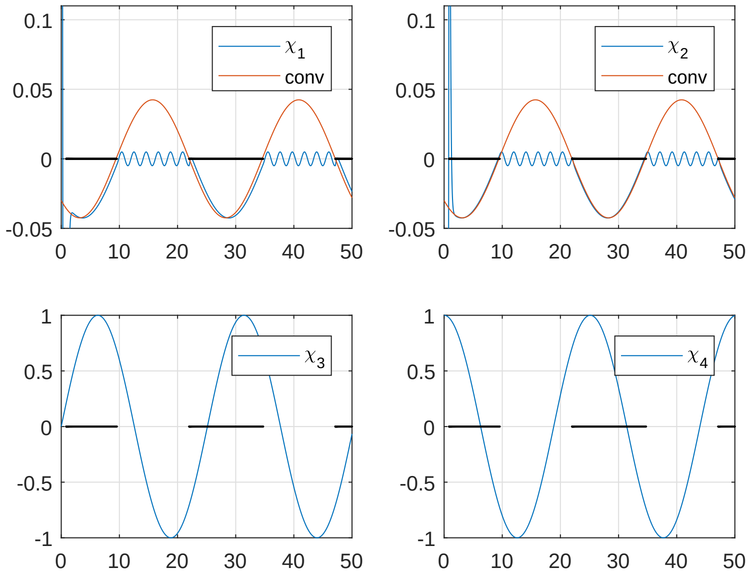

Figure 26: Illustration of the converging behaviour of $x_1 \to Mx_2$. The black line on the horizontal axis indicate the intervals on which $u > 0$.

The example above is just a 'toy example' to indicate how x_1 quickly converges to Mx_2, and returns to its original dynamics whenever $u = 0$. In the following example the control input is also applied to the dynamics of x_2.

Example 7. The previous example is extended to the case there x_2 also receives control input. For this example consider the simulation for

$$A_{11} = \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad B_1 = B_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}^{\mathrm{T}}. \tag{57}$$

The eigenvalues of A are given by $E_\lambda(A) = \{\pm 3\imath, \pm\imath\}$ in this case. Two typical results of simulations are included here. Figure 27 includes the results where F_1 and F_2 are chosen such that $E_\lambda(A_{11} + B_1 F_1) = \{-10, -10\}$ and $E_\lambda(A_{22} + \widetilde{B}_2 F_2) = \{-0.1, -0.1\}$. Figure 28 shows the result for a choice of F_2 such that $A_{22} + \widetilde{B}_2 F_2$ has oscilatory poles, in this case $E_\lambda(A_{22} + \widetilde{B}_2 F_2) = -0.1 \pm \imath$.

The simulations included in this example are typical results of a succesful application of singular perturbations theory to the positive control problem, for systems which have poles with zero real part. To illustrate that the approach also works for systems with eigenvalues with a positive real part, consider Figure 29. Here

$$A_{22} = \begin{bmatrix} 0.5 & 1 \\ -1 & 0.5 \end{bmatrix}, \quad E_\lambda(A_{22}) = 0.5 \pm \imath,$$

and F_2 is chosen such that $E_\lambda(A_{22} + \widetilde{B}_2 F_2) = -0.2 \pm 0.3\imath$, such that it suffices the condition from Willems et al. [29]. Again the control steers the states to $x = 0$. △

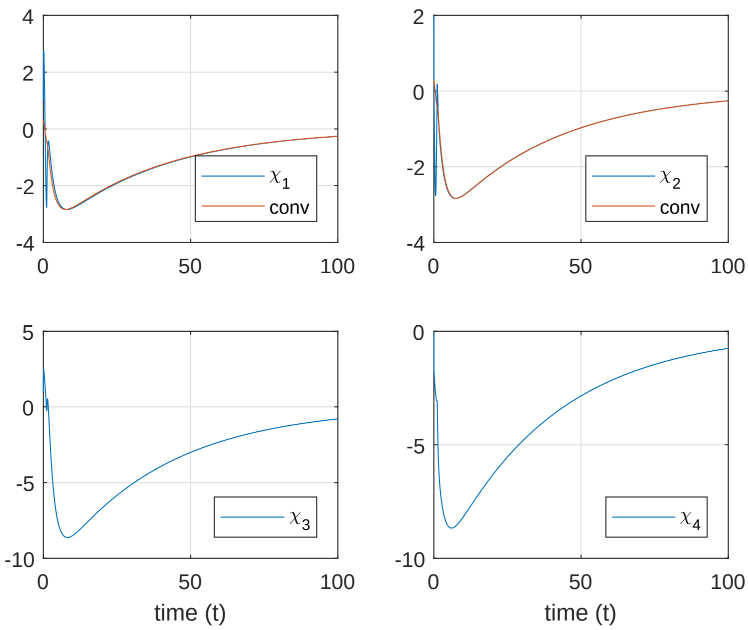

Figure 27: Simulation with F_1 and F_2 such that $E_\lambda(A_{11} + B_1 F_1) = \{-10, -10\}$ and $E_\lambda(A_{22} + \widetilde{B}_2 F_2) = \{-0.1, -0.1\}$.

The examples so far illustrate the feasibility of a positive control law based on singular perturbations theory. The example to follow shows controls which yield unstable systems.

Example 8. Consider the system (57) from the previous example with the same choice of F_1. In this case choose F_2 such that $E_\lambda(A_{22} + \widetilde{B}_2 F_2) = -0.1 \pm 2\imath$. The results are shown in Figure 30. This indicates that even though the choice of F_2 suffices the conditions from Willems et al. [29], it does not yield a stable system.

Another interesting example is that of Figure 31. Here the same system and poleplacement were used as in Example 7, with the subtle difference that the roles of A_{11} and A_{22} have switched, that is

$$A_{11} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix}.$$

Where Figure 27 shows stabilization, Figure 31 clearly does not. Apparently, it does matter which subsystem we choose to be the fast system, and which to be the slow. △

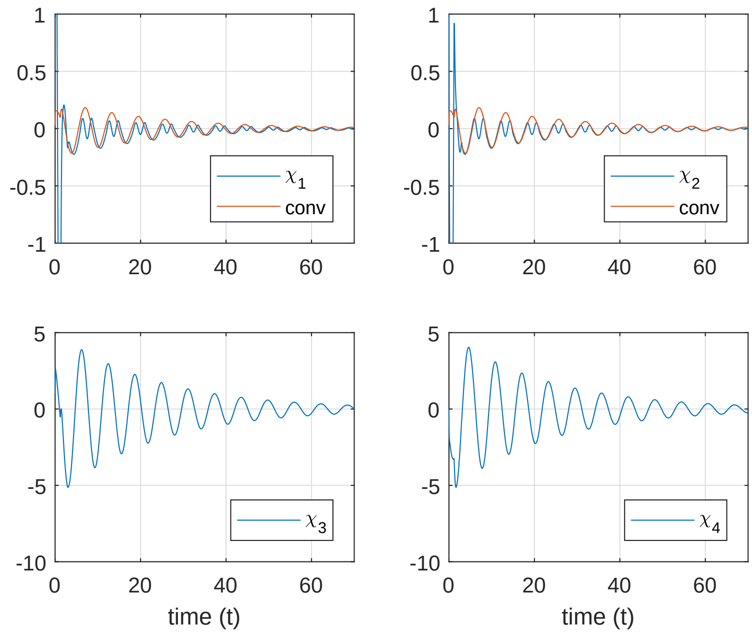

Figure 28: Simulation with F_1 and F_2 such that $E_\lambda(A_{11} + B_1F_1) = \{-10, -10\}$ and $E_\lambda(A_{22} + \widetilde{B}_2F_2) = -0.1 \pm \imath$.

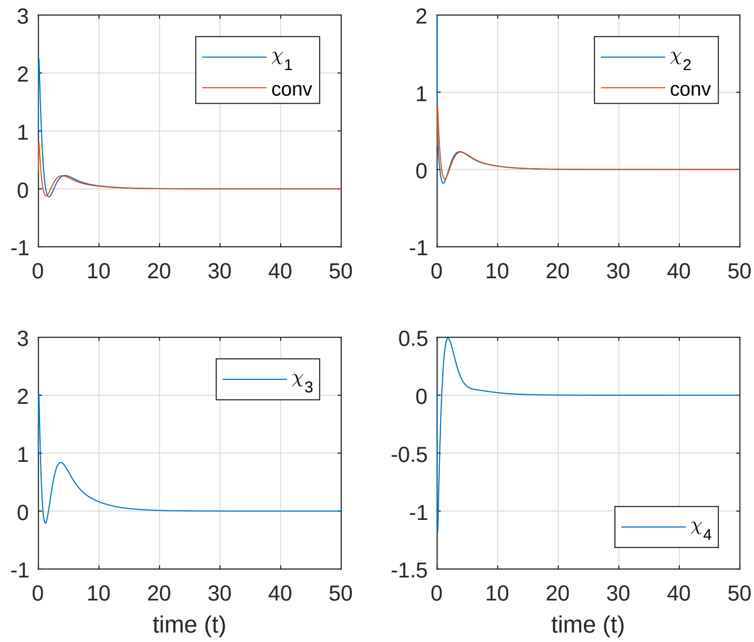

Figure 29: Simulation of Figure 28 redone for an unstable system.

5.4. Concluding remarks

In this chapter the techniques of singular perturbations for systems with 'regular' control were shown to be feasible for systems with positive control. However, they are not *proven* to be feasible and no conditions for feasibility are given. It is only suggested by means of various typical examples that the techniques may be of help for solving the four-dimensional positive control problem and thereby making direct use of the result of

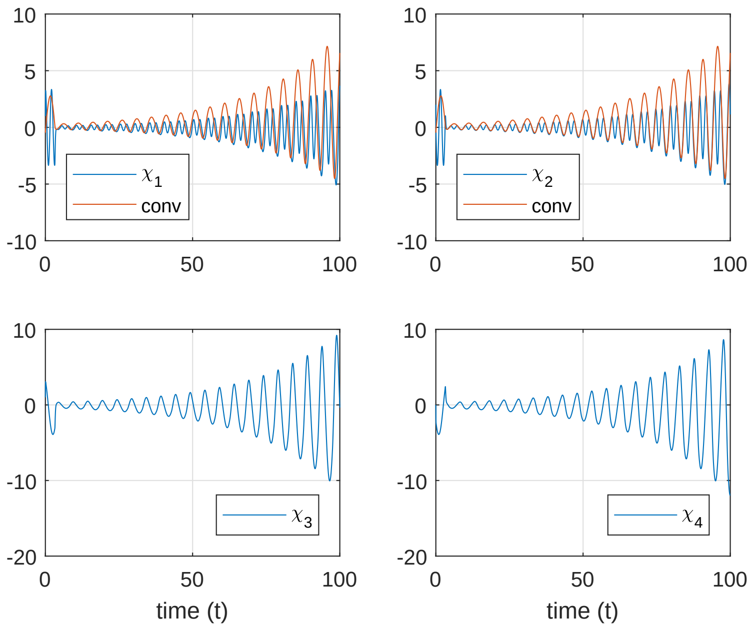

Figure 30: Example of an unstable positive control system where F_2 does suffice the conditions from Willems et al. [29].

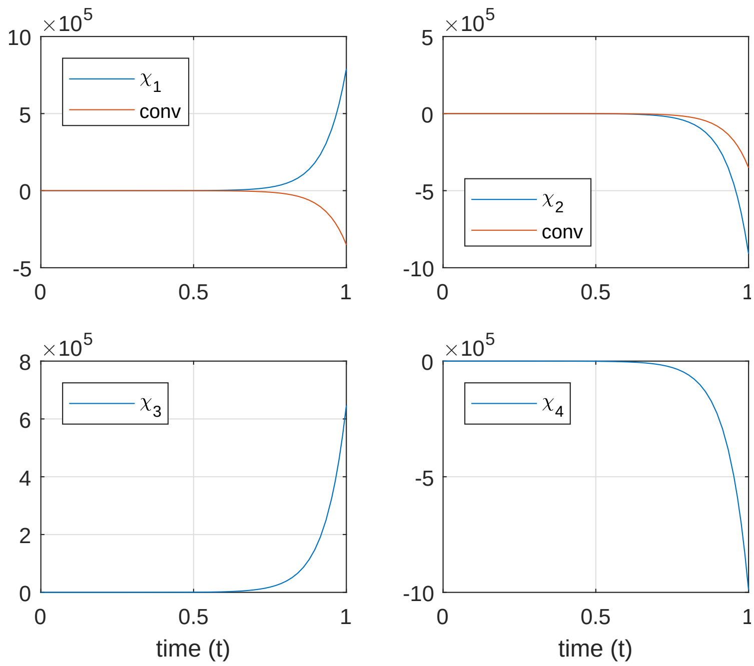

Figure 31: Unstable behaviour for switched A_{11} and A_{22}.

Willems et al. [29] concerning the two-dimensional problem. However carefully selected, examples are always specfic cases. Other simulations (with other system matrices, other input matrices, other initial conditions) were investigated though, which showed similar results to the simulations which were included in the previous examples. It should be mentioned that simulations always show some kind of 'bias', as it is the result of a rather specific choice of system parameters.

Another important aspect is that it was mentioned multiple times that the time separation between the slow and fast system should be 'large enough'. However, it was not investigated how large then exactly this separation should be, and how this separation should be measured. Moreover, any results on this topic for systems with 'regular' control yield no guarantee for the stability of positive control problems. Say it is known how large such a separation should be, then the four-dimensional problem can similarly be extended to a six-dimensional (and higher even-dimensional) problem as long as all three subsystems are mutually separated in time.

A remark should be made in response of the last example. As it shows, it apparently does matter which system is selected to be the fast system, and which to be the slow system. The reason for this has not been uncovered at this point. Hence, more research should be done into the behaviour of singular perturbed systems for positive control, specifically on this topic.

The remarks above suggest several topics for further research. Another topic of interest is the switching bahaviour in Figure 25, in relation with the choices of F_1 and F_2. As mentioned before, an unfortunate choice of controls F_1 and F_2 may result in cycling through Figure 25 endlessly. It would for example be interestig to look at

- whether certain switchings between classes can be prevented. If for example by a suitable choice of F_1 and F_2 the control system can be made to walk through the classes as follows: i $\rightarrow$ ii $\rightarrow$ iii, then a conclusion on $u = \max\{0, Fx\}$ being a stabilizing control may be drawn rather easily.
- the state solutions of x_1 and x_2 specifically. If one allows cycling through Figure 25 one could look at conditions on F_1 and F_2 which make sure that $||x_1||$ and $||x_2||$ at least decrease every cycle.

In the end the application of singular perturbations to this problem yields opportunities to extend to four-dimensional problem using the results from Willems et al. [29]. However, as is mentioned here before, the problem should be investigated in greater detail.

6. Conclusion

The main goal of this report was to study linear time invariant dynamical systems with positive controls. The research was initiated with a study of literature in general which led a more detailed study of two papers. The paper by Willems et al. [29] brought forth the main problem considered in this report, namely the stabilitation of linear time invariant dynamical systems with a positive state feedback control $u = \max\{0, Fx\}$ for systems with two non-asymptotically stable complex conjugate pole pairs. This problem was commonly referred to as 'the four-dimensional' problem throughout this report. The report contains various words on the extension of the results for the four-dimensional problem to any *even*-dimensional problem.

The difficulties of the four-dimensional problems were extensively described and illustrated in Section 3. The main conclusion of which was that known results for the two-dimensional problem could not straightforwardly be taken over to the four-dimensional problem. This motivated the two main approaches studied in this report.

First of all the design of a stabilizing positive state feedback control law based upon the stability theory of Lyapunov. The control law developed in Section 4.2 was proven to be asymptotically stabilizing for systems with at most two stable (but not asymptotically) poles. The result was extended to systems with an even number of stable (but not asymptotically) poles. At the end of the section the control law was applied to systems with unstable poles. For these systems it was shown and motivated that the Lyapunov based control law does not in general stabilize such systems. Suggestions for further research were made.

In contrast to the Lyapunov based control law, the second approach through techniques of singular perturbations attempted to stay closer to the known results of Willems et al. [29] for the two-dimensional problem. The goal of applying techniques of singular perturbations is to separate the four-dimensional problem into two two-dimensional positive control problems for which the stabilization problem was solved. A study into techniques of singular perturbations for ordinary differential equations was initiated, as the theories were unknown to the author. In Section 5.2 the general techniques were applied to the four-dimensional problem at hand. The expected behaviour of a singularly perturbed system with positive controls was described and visualized in Figure 25. The viability of the application of these techniques was supported by a series of simulations. No formal proof was stated here, and hence the section concludes with a series of remarks and suggestions for further research.

7. Recommendations and discussion

As always there are many remarks to be made concerning the approaches, results and contents of this report.

At the beginning of the report a summary of Heemels et al. [12] is inlcuded (rather extensive as appendix). This paper, together with Willems et al. [29], marked the start of this final project and was hence studied in great detail in the very first days. Also during that time the summary of Appendix D was written. The course of development of this project was to extend the result of Willems et al. [29], and leave the approach from the theories of optimal control for what it was. In the end the extension of Willems et al. [29] took up all the remaining time, and some form of follow up of Heemels et al. [12] was never established. Nevertheless the summary was still included.

Section 4 contains a lengthy proof for the stability of systems with purely imaginary poles. This proof was rectified and improved many times to come to the final version. At some point it contained two approaches which showed that the sequence $x(kT)$ converges to some $\bar{x} \in \mathbb{R}^n$: one showed that the sequence is a Cauchy sequence, another used the theorem of Bolzano-Weierstrass. Since both approached were investigated and worked out, they were also both included in the report.

The focus for the Lyapunov based control was mostly on systems with stable (but not asymptotically) poles. An attempt to extend the result to systems with unstable poles failed. This is mainly due to the fact that no Lyapunov function exists for unstable systems. The section concludes with many suggestions for further research into conditions for making the Lyapunov based control law feasible for systems with unstable poles. The question whether this approach is viable at all for these systems remains open. Suggestions for further research were listed in Section 4.3.2. However, it remains questionable whether a Lyapunov based control is even viable for unstable systems.

Another point for discussion is the choice of Lyapunov function. The control law was based upon the specific choice for a candidate Lyapunov function which is quadratic in x, namely $V(x) = \frac{1}{2}x^{\mathrm{T}}Px$. Other Lyapunov functions may be considered, for the proof on stability in Section 4.2.2 should not matter. Other Lyapunov functions may add more freedom to the control design, and may leave more options open for the case of systems with unstable poles.

The approach of singular perturbations was only investigated superficially. As has been mentioned before, no proofs of feasibility were included. Examples suggest the usability of singular perturbations for systems with positive controls. The nice thing of this approach is that the results from Willems et al. [29] can to some extent directly be applied to the four-dimensional problem. Naturally, more research is needed in order to pose conditions on the applicability of these results in combination with the techniques of singular perturbations. Other important remarks, as well as suggestions for further research were given in Section 5.4.

This report contains quite some examples. The main reason is that the author gained a lot of insight into the positive control system and other new theories via multiple simulations. Some of the examples were included as ‘typical results’ of simulations, especially for the section on singular perturbations. As was mentioned in Section 5.4, these simulations always show some bias and are based upon specific settings. They do give an indication whether the techniques that are tested in the simulation are usable or not.

8. Bibliography

[1] L. Benvenuti and L. Farina. The geometry of the reachability set for linear discrete-time systems with positive controls. *SIAM Journal on Matrix Analysis and Applications*, 28(2): 306–325, 2006. doi: 10.1137/040612531.

[2] R.F. Brammer. Controllability in linear autonomous systems with positive controllers. *SIAM Journal on Control*, 10(2):339–353, 1972. doi: 10.1137/0310026.

[3] M.K. Camlibel, W.P.M.H. Heemels, and J.M. Schumacher. Algebraic necessary and sufficient conditions for the controllability of conewise linear systems. *IEEE Transactions on Automatic Control*, 53(3):762–774, 2008. doi: 10.1109/TAC.2008.916660.

[4] M.L. Corradini, A. Cristofaro, F. Giannoni, and G. Orlando. *Control systems with saturating inputs: Analysis tools and advanced design*, volume 424. Springer Science & Business Media, 2012.

[5] P. de Leenheer and D. Aeyels. Stabilization of positive linear systems. *Systems & control letters*, 44(4):259–271, 2001. doi: 10.1016/S0167-6911(01)00146-3.

[6] M. Evans and D. Murthy. Controllability of discrete-time systems with positive controls. *IEEE Transactions on Automatic Control*, 22(6):942–945, 1977. doi: 10.1109/TAC.1977.1101648.

[7] L. Farina and S. Rinaldi. *Positive linear systems: theory and applications*, volume 50. John Wiley & Sons, 2011.

[8] M.E. Frias, F. Verduzco, H. Leyva, and F.A. Carrillo. On controllability of linear systems with positive control. *IFAC Proceedings Volumes*, 38(1):73–77, 2005. doi: 10.3182/20050703-6-CZ-1902.00582.

[9] F. Grognard. Feedback stabilization of predator-prey systems for impulsive biological control. *IFAC Proceedings Volumes*, 47(3):5264–5269, 2014. doi: 10.3182/20140824-6-ZA-1003.00666.

[10] F. Grognard, J. Rault, and J.L. Gouzé. Positive control for global stabilization of predator-prey systems. *IFAC Proceedings Volumes*, 46(23):265–270, 2013. doi: 10.3182/20130904-3-FR-2041.00136.

[11] W.P.M.H. Heemels and M.K. Camlibel. Controllability of linear systems with input and state constraints. In *CDC*, pages 536–541, 2007. doi: 10.1109/CDC.2007.4434588.

[12] W.P.M.H. Heemels, S.J.L. Van Eijndhoven, and A.A. Stoorvogel. Linear quadratic regulator problem with positive controls. *International Journal of Control*, 70(4):551–578, 1998. doi: 10.1080/002071798222208.

[13] G. Jeurnink. *Applied Functional Analysis*. University of Twente, reader 514, 2013.

[14] P.V. Kokotovic, R.E. O'Malley Jr., and P. Sannuti. Singular perturbations and order reduction in control theory - an overview. *Automatica*, 12(2):123–132, 1976. doi: 10.1016/0005-1098(76)90076-5.

[15] P.V. Kokotovic, H.K. Khali, and J. O'reilly. *Singular perturbation methods in control: analysis and design*, volume 25. Siam, 1999.

[16] H. Leyva and J. Solis-Daun. Synthesis of positive controls for the global clf stabilization of systems. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 656–660. IEEE, 2009. doi: 10.1109/CDC.2009.5400363.

[17] G. Meinsma. *Optimal Control*. University of Twente, reader 537, 2017.

[18] D.S. Naidu. *Singular perturbation methodology in control systems*. Number 34. IET, 1988.

[19] R.E. O'Malley Jr. et al. *Singular perturbation methods for ordinary differential equations*, volume 89. Springer Science & Business Media, 1991.

[20] M. Pachter. The linear-quadratic optimal control problem with positive controllers. *International Journal of Control*, 32(4):589–608, 1980. doi: 10.1080/00207178008922875.

[21] T. Perry and R. Gunderson. Controllability of nearly nonnegative linear systems using positive controls. *IEEE Transactions on Automatic Control*, 22(3):491–491, 1977. doi: 10.1109/TAC.1977.1101528.

[22] M.A. Rami and F. Tadeo. Controller synthesis for positive linear systems with bounded controls. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 54(2):151–155, 2007. doi: 10.1109/TCSII.2006.886888.

[23] J. Respondek. Controllability of dynamical systems with constraints. *Systems & Control Letters*, 54(4):293–314, 2005. doi: 10.1016/j.sysconle.2004.09.001.

[24] V.R. Saksena, J. O'reilly, and P.V. Kokotovic. Singular perturbations and time-scale methods in control theory: survey 1976–1983. *Automatica*, 20(3):273–293, 1984. doi: 10.1016/0005-1098(84)90044-X.

[25] S.H. Saperstone. Global controllability of linear systems with positive controls. *SIAM Journal on Control*, 11(3):417–423, 1973. doi: 10.1137/0311034.

[26] S.H. Saperstone and J.A. Yorke. Controllability of linear oscillatory systems using positive controls. *SIAM Journal on Control*, 9(2):253–262, 1971. doi: 10.1137/0309019.

[27] N.K. Son and D.D. Thuan. Controllability radii of linear systems with constrained controls under structured perturbations. *SIAM Journal on Control and Optimization*, 54 (5):2820–2843, 2016. doi: 10.1137/15M1006702.

[28] W.R. Wade. *Introduction to Analysis: Pearson New International Edition*. Always learning. Pearson Education Limited, 2013. ISBN 9781292039329.

[29] F. Willems, W.P.M.H. Heemels, B. De Jager, and A.A. Stoorvogel. Positive feedback stabilization of centrifugal compressor surge. *Automatica*, 38(2):311–318, 2002. doi: 10.1016/S0005-1098(01)00202-3.

[30] H. Yoshida and T. Tanaka. Positive controllability test for continuous-time linear systems. *IEEE Transactions on Automatic Control*, 52(9):1685–1689, 2007. doi: 10.1109/TAC.2007.904278.

[31] Y. Zhang, D. Subbaram Naidu, Chenxiao C., and Y. Zou. Singular perturbations and time scales in control theories and applications: an overview 2002–2012. *International Journal of Information and Systems Sciences*, 9(1):1–36, 2014.

[32] B. Zuckerman, D. Jefferson, et al. *Human population and the environmental crisis*. Jones & Bartlett Learning, 1996.

A. Simulations in Matlab

All simulations that were included in this report were made using MATLAB. The scripts are rather plain, so that it suffices to include a few comments here.

Positive control system

The positive state feedback system $\dot{x} = Ax + B\max\{0, Fx\}$ is defined in a separate MATLAB function. The dynamics are computed according to the following code:

```
function    dx = positive_state_feedback(t,x,F,A,B)
    u=max(0,F*x);
    dx = A*x + B*u;
end
```

the function requires matrices `A`, `B` and `F` as parameters, and uses the state `x` as input. The input `t` is required for the ODE solver.

ODE solver

Generally, the simulations use MATLAB 's ordinary differential equation solver `ode45`.

```
[t,y]=ode45(odefun,tspan,x0);
```

It requires the `odefun` f from the system $\dot{x} = f(t, x)$, the time span `tspan` for the simulation with initial condition x_0. For simulations with positive control the syntax is as follows.

```
[t,y]=ode45(@(t,x) positive_state_feedback(t,x,F,A,B),tspan,x0);
```

Kalman poleplacement formula

For a given system the command `K=place(A,B,p)` computes the state feedback matrix $F = -K$ which yield the desired poles of the system $\dot{x} = (A + BF)x$. However, for systems where B has dimension $n \times 1$, as in the case of scalar control, the command `place` cannot place poles with multiplicity larger than 1. For that reason Kalman's poleplacement formula was used instead of MATLAB's command `place`.

B. Auxiliary Theorems and Definitions

Theorem B.1 (Ackermann's pole placement formula) *Suppose that* (A, B) *is controllable with* $B \in \mathbb{R}^{n\times 1}$. *Denote the characteristic polynomial of* A *by* $p_A(s) = s^n + p_{n-1}s^{n-1} + \cdots + p_0$. *Given a monic polynomial*

$$p(s) = s^n + r_{n-1}s^{n-1} + \cdots + r_0$$

there exists a unique F *for which* $p_{A+BF}(s) = p(s)$. *This* F *can, among other methods, be determined as*

$$F = -\begin{bmatrix} 0 & \ldots & 0 & 1 \end{bmatrix} \mathcal{C}^{-1}_{\{A,B\}} p(A)$$

where $\mathcal{C}_{\{A,B\}}$ *is the controllability matrix of the pair* (A, B) *and* $p(A)$ *is the desired characteristic polynomial* $p(s)$ *is evaluated for* $s = A$.

Definition B.1a (Positive and negative (semi) definite functions) A function $V : \Omega \to \mathbb{R}$ is *positive definite* on Ω relative to $\bar{x} \in \Omega$ if

$$V(\bar{x}) = 0 \text{ while } V(x) > 0 \text{ for all } x \in \Omega\bar{x}.$$

It is *positive demi-definite* if $V(\bar{x}) = 0$ and $V(x) \geq 0$ for all other x. Likewise $V(x)$ is *negative (semi) definite* if $-V$ is positive (semi) definite.

Definition B.1b (Positive and negative (semi) definite matrix) Let $M \in \mathbb{R}^{n\times n}$ be symmetric. M is said to be *positive definite* if $x^\mathsf{T} Mx > 0$ for all $x \in \mathbb{R}^n \backslash \vec{0}$. M is said to be *positive semi-definite* if $x^\mathsf{T} Mx \geq 0$ for all $x \in \mathbb{R}^n \backslash \vec{0}$. Likewise M is is *negative (semi) definite* if $-M$ is positive (semi) definite.

Theorem B.2 (Bolzano Weierstrass - as presented in [28]) *Every bounded sequence in* $\mathbb{R}^n$ *has a convergent subsequenct.*

Definition B.2 (Inner product space - as presented in [13]) The linear space H is called an *inner product space* if to any two elements f and g in H there is associated a real number $\langle f, g\rangle$, called the *inner product* between f and g with the properties

(i) $\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$

(ii) $\langle \lambda f, g\rangle = \lambda \langle f, g\rangle$ for all $\lambda \in \mathbb{R}$

(iii) $\langle g, f\rangle = \langle f, g\rangle$

(iv) $\langle f, f\rangle > 0$ if $f \neq 0$, while $\langle f, f\rangle = 0$ for $f = 0$

◇

Theorem B.3 (Cauchy-Schwarz inequality - as presented in [13]) *For any two elements* f *and* g *in an inner product space there holds*

$$|\langle f, g\rangle| \leq ||f||\,||g||\,.$$

Moreover, the equality holds when f *and* g *are linearly independent.*

C. Alternative approach concerning Theorem 11

Proof. Consider the sequence of vectors $x(kT)$ defined in (37)

$$x(kT) = x((k-1)T) + \int_{(k-1)T}^{kT} \mathrm{e}^{A(kT-\tau)} Bu(\tau)\,\mathrm{d}\tau.$$

The sequence is bounded in the sense that $||x(kT)|| \leq ||x_0||$. This is a result of the property that the Lyaponov function $V(x)$ is nonincreasing. By Theorem B.2, $x(kT)$ contains a convergent subsequence $x(k_iT)$, $i = 1, 2, \ldots$. The elements of this subsequence can be defined in two ways:

$$x(k_iT) = x((k_i-1)T) - \int_{(k_i-1)T}^{k_iT} z(\tau,k)u(\tau)\,\mathrm{d}\tau \tag{C.1a}$$

$$x(k_iT) = x(k_{i-1}T) - \int_{k_{i-1}T}^{k_iT} z(\tau,k)u(\tau)\,\mathrm{d}\tau \tag{C.1b}$$

The definition used in (C.1a) is in line with the original sequence whereas the definition in (C.1b) only uses elements from the convergent subsequence. Denote the accumulation vector of the subsequence by $\bar{x}$ such that $x(k_iT) \to \bar{x}$ as $i \to \infty$, or equivalently $|x(k_iT) - x(k_{i-1}T)| \to 0$ as $i \to \infty$. Therefore in (C.1b) for the integral it holds that $\int_{k_{i-1}T}^{k_iT} z(\tau,k)u(\tau)\,\mathrm{d}\tau \to 0$ as $i \to \infty$. □

D. Positive control problem approached from optimal control

This appendix includes a summary of Heemels et al. [12]. The following section contains some general theorems that are used in the summary in D.2.

D.1. Preliminary theorems

Consider the problem of minimizing a cost

$$J(x_0, u(\cdot)) = S(x_0) + \int_{t_0}^{t_f} L(x(t), u(t))\,\mathrm{d}t$$

over all inputs $u : [t_0, t_f] \to \Omega$, subject to $\dot{x}(t) = f(x(t), u(t))$, $x(0) = x_0$. The function

$$H(x, p, u) := p^{\mathrm{T}} f(x, u) + L(x, u) \tag{D.1}$$

is called the *Hamiltonian*. The stationary solutions $(x(\cdot), p(\cdot), u(\cdot))$ of the minimization problem with $x(0) = x_0$ satisfy

$$\begin{aligned}
\dot{x}(t) &= \frac{\partial H(x(t), p(t), u(t))}{\partial p}, \quad x(0) = x_0 \\
\dot{p}(t) &= -\frac{H(x(t), p(t), u(t))}{\partial x}, \quad p(t_f) = \frac{\partial S(x(t))}{\partial x} \\
0 &= \frac{H(x(t), p(t), u(t))}{\partial u}
\end{aligned}$$

Theorem D.1 (Pontryagin's Minimum Principle) *Consider the optimal control problem of minimizing a cost*

$$J(x_0, u(\cdot)) = S(x(t_f)) + \int_{t_0}^{t_f} L(x(t), u(t))\,dt$$

over all inputs $u : [t_0, t_f] \to \mathbb{U}$, subject to $\dot{x}(t) = f(x(t), u(t)), x(0) = x_0$.

Suppose $u^* : [t_0, t_f] \to \mathbb{U}$ is a solution of the optimal control problem, and $x^*(\cdot)$ the resulting optimal state trajectory. Then there exists a function $p^* : [t_0, t_f] \to \mathbb{R}^n$ such that

$$\dot{x}^*(t) = \frac{\partial H(x^*(t), p^*(t), u^*(t))}{\partial p}, \quad x^*(0) = x_0, \tag{D.2}$$

$$\dot{p}^*(t) = -\frac{\partial H(x^*(t), p^*(t), u^*(t))}{\partial x}, \quad p^*(t_f) = \frac{\partial S(x^*(t_f))}{\partial x} \tag{D.3}$$

and at the solution $x^*(t), p^*(t)$ the input $u^*(t)$ at each moment in time minimizes the Hamiltonian*:*

$$H(x^*(t), p^*(t), u^*(t)) = \min_{u \in \mathbb{U}} H(x^*(t), p^*(t), u(t))$$

for every $t \in [t_0, t_f]$.

In general, i.e. for general input ($u^* = u$), state ($x^* = x$) and costate ($p^* = p$), (D.2) and (D.3) are called the Hamiltonian equations.

Theorem D.2 (Hamilton-Jacobi-Bellman) *Consider the minimization problem of Theorem D.1. Suppose a real-valued fuction $V : \mathbb{R}^n \times [t_0, t_f] \to \mathbb{R}$ exists that is continuously differentiable in $\mathbb{R}^n \times (t_0, t_f)$ and that satisfies the partial differential equation*

$$\frac{\partial V(x,t)}{\partial t} + \min_{u\in\mathbb{U}} \left[\frac{\partial V(x,t)}{\partial x^{\mathrm{T}}} f(x,u) + L(x,u) \right] = 0 \tag{D.4}$$

for all $x \in \mathbb{R}^n$ and all $t \in (t_0, t_f)$ and with final condition $V(x,t_f) = S(x(t_f))$ for all $x \in \mathbb{R}^n$. Then the following holds:

1. *For any admissible input $u : [t_0, t_f] \to \mathbb{U}$, the function $V(\tau, x(\tau))$ $(t_0 \leq \tau < T)$ is a lower bound for the cost over $[\tau, t_f]$,*

$$J_{[\tau,t_f]}(x(\tau), u(\cdot)) \geq V(\tau, x(\tau)).$$

2. *If there exists a function $\tilde{u} : \mathbb{R}^n \times [t_0, t_f] \to \mathbb{U}$ such that*

$$\tilde{u}(x,t) \in \arg\min_{u\in\mathbb{U}} \left[\frac{\partial V(x,t)}{\partial x^{\mathrm{T}}} f(x,u) + L(x,u) \right]$$

for every $t \in (t_0, t_f)$ and every $x \in \mathbb{R}^n$ and such that the differential equation

$$\dot{x}(t) = f(x(t), \tilde{u}(t, x(t))), \quad x(0) = x_0$$

has a solution for $t \in [t_0, t_f]$, then $u^*(t) := \tilde{u}(t, x(t))$ is a solution to the optimal control problem and

$$J_{[t_0,t_f]}(x_0, u^*(\cdot)) = \min_{u:[t_0,t_f]\to\mathbb{U}} J_{[t_0,t_f]}(x_0, u(\cdot)) = V(t_0, x_0)$$

Furthermore, if the differential equation

$$\dot{x}(t) = f(x(t), \tilde{u}(t, x(t))), \quad x(\tau) = z_\tau$$

has a solution for every $\tau \in [t_0, t_f]$ and for every $z_\tau \in \mathbb{R}^n$, then $V(x,t)$ is the value function.

In this theorem, Equation (D.4) is called the *Hamiltonian-Jacobi-Bellman equation*, or abbreviated HJB equation.

Definition D.1 (Lipschitz continuity) A function f is *Lipschitz continuous* on $\Omega \subset \mathbb{R}^m$ if a *Lipschitz constant* $K \geq 0$ exists such that

$$||f(x) - f(z)|| \leq K\,||x - z||$$

for all $x, z \in \Omega$. It is Lipschitz continuous *at* $x_0 \in \mathbb{R}^m$ if it is Lipschitz continuous on some neighbourhood Ω of x_0. Consider the following theorem:

> **Theorem** (Lipschitz condition) *Let $t_0 \in \mathbb{R}$ and $x_0 \in \mathbb{R}^n$. If f is Lipschitz continuous at x_0 then, for some $\delta > 0$, the differential equation $\dot{x} = Ax + Bu$ has a unique solution $x(t, x_0)$ on the time interval $[t_0, t_0 + \delta)$. Furthermore for any fixed $t \in [t_0, t_0 + \delta)$ the solution $x(t; x_0)$ then depends continuously on x_0.*

A function is *locally Lipschitz continuous* if it is Lipschitz continuous at every x_0. It is *uniformly Lipschitz continuous* if we can find a single $K \geq 0$ such that the Lipschitz condition is satisfied. ◇

Definition D.2 (Hilbert space) A closed inner product space is called a *Hilbert space*. ◇

D.2. Summary

Heemels et al. [12] is an extension to the findings in Pachter [20] which addresses the linear quadratic regulator (LQR) problem with positive controls in the case where the integrand in the cost functional is independent of the trajectory x. This restriction is not made in Heemels et al. [12], where the existence and uniqueness of an optimal control is proven for the more general case. In general, the LQR problem considers a minimization of the quadratic cost

$$J_{[t_0,t_f]}(t_0, x_0, u(\cdot)) = x(t_f)^{\mathrm{T}} S x(t_f) + \int_{t_0}^{t_f} x(t)^{\mathrm{T}} Q x(t) + u(t)^{\mathrm{T}} R u(t)\, \mathrm{d}t$$

over all inputs $u : [t_0, t_f] \to \mathbb{R}^p$ and states $x : [t_0, t_f] \to \mathbb{R}^n$ governed by a linear time invariant system $\dot{x}(t) = Ax(t) + Bu(t)$, $x(t_0) = x_0$.

The *maximum principle* and *dynamic programming*, both classic approaches in optimal control theory, are addressed for the linear quadratic regulator problem with positive controls. The maximum principle involves the Hamiltonian and the costate equation. Conditions for an optimum involve conditions on the Hamiltonian, see Equation (D.1). Dynamic programming on the other hand leads to a partial differential equation, the so-called *Hamilton-Jacobi-Bellman* (HJB) equation, see Equation (D.4). The maximum principle yields a sufficient condition for an optimum, moreover it leads to an open-loop control. For dynamic programming on the other hand, the HJB equation provides a necessary and sufficient condition for an optimum. Also, it yields a closed-loop control law.

Both approaches are treated by Heemels et al. [12]. Firstly, the finite horizon cases are consdered. At the end, the paper poses conditions under which the finite horizon case extends to the infinite horizon case.

D.2.1. Problem formulation

Section 2 of Heemels et al. [12] formulates the main problem that is considered. They consider the linear system with input $u : [t_0, t_f] \to \mathbb{R}^p$, state $x : [t_0, t_f] \to \mathbb{R}^n$ and output $z : [t_0, t_f] \to \mathbb{R}^q$, given by

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{D.5}$$

$$z(t) = Cx(t) + Du(t) \tag{D.6}$$

where t denoted time, A, B, C and D are matrices of appropriate dimension, t_0 is the starting time and t_f is some fixed end time ($t_f > t_0$). The paper restricts to the regular case, by which is meant that matrix D has full column rank.

Heemels et al. consider inputs in the Lebesgue space of square integrable, measurable functions on $[t_0, t_f]$ taking values in $\mathbb{R}^p$, denoted by $L_2[t_0, t_f]^p$. For every $u \in L_2[t_0, t_f]^p$ and initial condition (t_0, x_0), that is $x(t_0) = x_0$, the solution of Equation (D.5) is an absolutely continuous state trajectory denoted by $x_{t_0,x_0,u}(t)$. The corresponding output $z_{t_0,x_0,u}(t)$ can be written as

$$z_{t_0,x_0,u}(t) = \underbrace{C e^{A(t-t_0)} x_0}_{(\mathcal{M}_{t_0,t_f} x_0)(t)} + \underbrace{\int_{t_0}^{t_f} C e^{A(t-\tau)} B u(\tau)\, \mathrm{d}\tau + Du(t)}_{(\mathcal{L}_{t_0,t_f} u)(t)} \tag{D.7}$$

for $t \in [t_0, t_f]$. Note that both $\mathcal{M}_{t_0,t_f} : \mathbb{R}^n \to L_2[t_0, t_f]^m$ and $\mathcal{L}_{t_0,t_f} : L_2[t_0, t_f]^p \to L_2[t_0, t_f]^q$ are bounded linear operators. Heemels et al. [12] only consider positive controls. Therefore they define the closed convex cone of positive functions in $L_2[t_0, t_f]^p$ by

$$P[t_0, t_f] := \left\{ u \in L_2[t_0, t_f]^p \mid u(t) \in \Omega \text{ almost everywhere on } [t_0, t_f] \right\}.$$

Here Ω is the control restraint set, which is equal to

$$\mathbb{R}^p_+ := \{\mu \in \mathbb{R}^p \mid \mu_i \geq 0 \text{ for } i = 1, \ldots, p\},$$

the non-negative orphant in $\mathbb{R}^p$.

The paper treats the following problem for the finite horizon case:

Problem 1 (Finite horizon optimal control problem) *The objective is to determine for every initial condition $(t_0, x_0) \in [0, t_f] \times \mathbb{R}^n$ a control input $u \in P[t_0, t_f]$, an optimal control, such that*

$$\begin{aligned} J(t_0, x_0, u) &:= ||z_{t_0,x_0,u}||_2^2 = \left|\left|\mathcal{M}_{t_0,t_f} x_0 + \mathcal{L}_{t_0,t_f} u\right|\right|_2^2 \\ &= \int_{t_0}^{t_f} z^{\mathrm{T}}_{t_0,x_0,u}(t) z_{t_0,x_0,u}(t)\, dt \\ &= \int_{t_0}^{t_f} \left[x^{\mathrm{T}}_{t_0,x_0,u}(t) C^{\mathrm{T}} C x_{t_0,x_0,u}(t) + 2x^{\mathrm{T}}_{t_0,x_0,u}(t) C^{\mathrm{T}} D u(t) + u^{\mathrm{T}}(t) D^{\mathrm{T}} D u(t)\right] dt \end{aligned} \tag{D.8}$$

is minimal.

Later on, this problem is extended to the infinite horizon case. The infinite horizon problem is stated as follows (with the assumption that $t_0 = 0$):

Problem 2 (Infinite horizon optimal control problem) *The objective is to determine for all initial states x_0 a control input $u \in P[0, \infty)$ minimizing*

$$J_\infty(x_0, u) = \int_0^\infty z^{\mathrm{T}}_{x_0,u}(t) z_{x_0,u}(t)\, dt \tag{D.9}$$

subject to (D.5), (D.6) and $x(0) = x_0$ and the additional constraint that the corresponding state trajectory x is contained in $L_2[0, \infty)^n$

Heemels et al. dedicate the greater part of their paper to the finite horizon problem. The last chapter of the paper is dedicated to extend the finite-horizon results to results for the infinite-horizon case. For this reason, let us first focus on the results achieved for the finite horizon problem.

In the finite horison problem the control functions minimizing (D.8) for given initial conditions and horizon are called *optimal controls*. Heemels et al. consider questions concerning the existence of an optimal control. And if it does, is it unique and how is it characterized? They also consider whether the optimal control can be explicitly computed or whether there is a numerical method to approximate the exact optimal control.

The optimal value of J for all considered initial conditions is described by the value function, defined as follows:

Definition D.3 (Value function) The value function V is a function from $[0, t_f] \times \mathbb{R}^n$ to $\mathbb{R}$ and is defined for every $(t_0, x_0) \in [0, t_f] \times \mathbb{R}^n$ by

$$V(t_0, x_0) := \inf_{u \in P[t_0, t_f]} J(t_0, x_0, u) \tag{D.10}$$

◇

D.2.2. *Projections on closed convex sets*

The existence of the optimal control is based upon projections on closed convex sets. The most important definitions and theorems listed by Heemels et al. are included in this section.

Theorem D.3 (Minimum distance to a convex set) *Let x be a vector in a Hilbert space H with inner product $\langle \cdot, \cdot \rangle$ and let K be a closed convex subset of H. Then there exists exactly one $k_0 \in K$ such that $||x = k_0|| \leq ||x - k||$ for all $k \in K$. Furthermore, a necessary and sufficient condition that k_0 is the unique minimizing vector is that $\langle x - k_0, k - k_0 \rangle \leq 0$ for all $k \in K$.*

Definition D.4 Let K be a closed convex set of the Hilbert space H. The projection P_K onto K is defined as the operator that assigns for each $x \in H$ the vector contained in K that is closest to x in the norm induced by the inner product. Formally,

$$P_K x = k_0 \iff ||x - k_0|| \leq ||x - k|| \quad \forall k \in K$$

for $x \in H$. ◇

In accordance with [12] P_K is globally Lipschitz continuous, where for all $x, y \in H$ it holds that

$$||P_k x - P_k y|| \leq ||x - y|| \,.$$

Besides, assume K is a cone (besides being closed and convex). Then

$$P_k(\alpha x) = \alpha P_k(x) \quad \forall \alpha \geq 0, x \in H.$$

Note that the definition of Lipschitz continuity is given in Definition D.1.

D.2.3. Existence and uniqueness

A standing assumption throughout Heemels et al. [12] is that D in Equation (D.6) has full column rank. In that case, $\mathcal{L}_{t_0,t_f}$ in (D.7) has a bounded left inverse $\widetilde{\mathcal{L}}_{t_0,t_f} : L_2[t_0, t_f]^q \to L_2[t_0, t_f]^p$. $\widetilde{\mathcal{L}}_{t_0,t_f}$ is a bounded left inverse of $\mathcal{L}_{t_0,t_f}$ if $\mathcal{L}_{t_0,t_f} u = z$ implies $\widetilde{\mathcal{L}}_{t_0,t_f} z = u$. In agreement with Heemels et al. $\widetilde{\mathcal{L}}_{t_0,t_f}$ is described by the state-space representation

$$\begin{aligned} \dot{x}(t) &= \left(A - B(D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}C\right) x(t) + B(D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}z(t), \quad x(t_0) = 0 \\ u(t) &= (D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}} \left\{z(t) - Cx(t)\right\} \end{aligned}$$

Using $\widetilde{\mathcal{L}}_{t_0,t_f}$ the original minimization problem (see Problem 1) can be reformulated as the minimization of $\left|\left|-v - \mathcal{M}_{t_0,t_f} x_0\right|\right|$ over $v \in \mathcal{L}_{t_0,t_f}(P[t_0, t_f])$. Compare this minimization problem to Theorem D.3 with $K = \mathcal{L}_{t_0,t_f}(P[t_0, t_f])$ and $k_0 = \mathcal{L}_{t_0,t_f}(u^*)$. The linearity of $\mathcal{L}_{t_0,t_f}$ in u shows that $\mathcal{L}_{t_0,t_f}(P[t_0, t_f])$ is a convex cone. Moreover, since K is the inverse image of a closed set under $\mathcal{L}_{t_0,t_f}$, it is closed as well. This proves the following theorem:

Theorem D.4 (Existence and uniqueness of the optimal control – *theorem 2 of [12]*) *Let $t_0, t_f \in \mathbb{R}$ with $t_0 < t_f$ and $x_0 \in \mathbb{R}^n$. There exists a unique control $u_{t_0,t_f,x_0} \in P[t_0, t_f]$ such that*

$$\left|\left|\mathcal{M}_{t_0,t_f} x_0 + \mathcal{L}_{t_0,t_f} u_{t_0,t_f,x_0}\right|\right|_2 \leq \left|\left|\mathcal{M}_{t_0,t_f} x_0 + \mathcal{L}_{t_0,t_f} u\right|\right|_2$$

for all $u \in P[t_0, t_f]$. A necessary and sufficient condition for $u^* \in P[t_0, t_f]$ to be the unique minimizing control is that

$$\left\langle \mathcal{M}_{t_0,t_f} x_0 + \mathcal{L}_{t_0,t_f} u^*, \mathcal{L}_{t_0,t_f} u - \mathcal{L}_{t_0,t_f} u^* \right\rangle \geq 0$$

for all $u \in P[t_0, t_f]$.

The optimal control, the optimal state trajectory and the optimal output with initial conditions (t_0, x_0) and final time t_f exist and are denoted by u_{t_0,t_f,x_0}, x_{t_0,t_f,x_0} and z_{t_0,t_f,x_0} respectively.

We can write

$$u_{t_0,t_f,x_0} = \widetilde{\mathcal{L}}_{t_0,t_f}\mathcal{P}(-\mathcal{M}_{t_0,t_f}x_0) \tag{D.11}$$

where $\mathcal{P}$ is the projection on the closed convex cone $\mathcal{L}_{t_0,t_f}(P[t_0,t_f])$ in the Hilbert space $L_2[t_0,t_f]^q$. Also, since $\mathcal{P}$ is positive homogeneous, it holds that for $\alpha \geq 0$

$$u_{t_0,t_f,\alpha u} = \alpha u_{t_0,t_f,x_0} \quad \text{and} \quad V(t_0,\alpha x_0) = \alpha^2 V(t_0,x_0).$$

D.2.4. The maximum principle

Classic in optimal control theory is Pontryagin's minimum/maximum principle. This principle is stated in Theorem D.1. Consider for Equation (D.8) the Hamiltonian given by

$$H(x,p,\mu) = x^{\mathrm{T}}A^{\mathrm{T}}p + \mu^{\mathrm{T}}B^{\mathrm{T}}p + x^{\mathrm{T}}C^{\mathrm{T}}Cx + 2x^{\mathrm{T}}C^{\mathrm{T}}D\mu + \mu^{\mathrm{T}}D^{\mathrm{T}}D\mu$$

for $(x,p,\mu) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p$. The costate equation (D.3) is then given by

$$\dot{p}(t) = -A^{\mathrm{T}}p(t) - 2C^{\mathrm{T}}Cx(t) - 2C^{\mathrm{T}}Du(t) = A^{\mathrm{T}}p(t) - 2C^{\mathrm{T}}z(t), \quad p(t_f) = 0. \tag{D.12}$$

Theorem D.1 now states that the optimal control u_{t_0,t_f,x_0} satisfies for all $t \in [t_0,t_f]$

$$u_{t_0,t_f,x_0}(t) \in \arg\min_{\mu\in\mathbb{R}^m} H(x_{t_0,t_f,x_0}(t),\mu,p_{t_0,t_f,x_0}(t))$$

where $p_{t_0,t_f,x_0}(t)$ is the solution of Equation (D.12) with $z = z_{t_0,t_f,x_0}$. The following theorem is a straight application of the maximum principle. The optimal control x^* is found according to the maximum principle, i.e. for x^* it holds that $\frac{\partial H}{\partial \mu} = 0$. Then the optimal positive control u_{t_0,t_f,x_0} is found by projecting u^* onto $\mathbb{R}^m_+$.

Theorem D.5 (*Lemma 3 of [12]*) *Let the initial time t_0, final time $t_f > t_0$ and initial state $x_0 \in \mathbb{R}^n$ be fixed. In $\mathbb{R}^n$, let P_Ω denote the projection on $\Omega := \mathbb{R}^p_+$. The optimal control u_{t_0,t_f,x_0} satisfies*

$$u_{t_0,t_f,x_0} = P_\Omega\left(-\tfrac{1}{2}(D^{\mathrm{T}}D)^{-1}\left\{B^{\mathrm{T}}p_{t_0,t_f,x_0}(t) + 2D^{\mathrm{T}}Cx_{t_0,t_f,x_0}(t)\right\}\right) \tag{D.13}$$

for all $t \in [t_0,t_f]$, where the functions x_{t_0,t_f,x_0} and p_{t_0,t_f,x_0} are given by

$$\begin{aligned} \dot{x}_{t_0,t_f,x_0} &= Ax_{t_0,t_f,x_0} + Bu_{t_0,t_f,x_0}, & x(t_0) &= x_0 \\ \dot{p}_{t_0,t_f,x_0} &= -A^{\mathrm{T}}p_{t_0,t_f,x_0} - 2C^{\mathrm{T}}Cx_{t_0,t_f,x_0} - 2C^{\mathrm{T}}Du_{t_0,t_f,x_0}, & p(t_f) &= 0 \end{aligned} \tag{D.14}$$

Moreover, the optimal control is continuous in time.

Substituting Equation (D.13) into (D.14) gives a two-point boundary value problem. Its solutions provide us with a set of candidates containing the optimal control, because the maximum principle is a necessary condition for optimality.

D.2.5. Dynamic programming

Heemels et al. show that the LQ-problem with positive controls satisfiest the Hamiltonian-Jacobi-Bellman equation (see Theorem D.2) in classical sense. They introduce the function L (which they call the 'Lagrangian') for $(x(t),\mu(t)) \in \mathbb{R}^n \times \Omega$ defined by

$$L(x,\mu) = z^{\mathrm{T}}z = x^{\mathrm{T}}C^{\mathrm{T}}Cx + 2x^{\mathrm{T}}C^{\mathrm{T}}D\mu + \mu^{\mathrm{T}}D^{\mathrm{T}}D\mu.$$

Here the arguments of $x(t)$ and $\mu(t)$ were omitted for notational purposes. The HJB equation (D.4) is for $x \in \mathbb{R}^n$ and $t \in [t_0, t_f]$ given by

$$W_t(t,x) + \inf_{\mu \in \Omega} \{W_x^{\mathrm{T}}(t,x)(Ax + B\mu) + L(x,\mu)\} = 0 \tag{D.15}$$

where W is a function with domain $[t_0, t_f] \times \mathbb{R}^n$ taking values in $\mathbb{R}$, and where $W_t(t,x) := \frac{\partial W(t,x)}{\partial t}$ and $W_x^{\mathrm{T}}(t,x) := \frac{\partial W(t,x)}{\partial x^{\mathrm{T}}}$ denote its partial derivatives with respect to time t and state x, respectively.

Now, according to Theorem D.2 the following holds: If W is a continuously differentiable solution of the HJB equation (D.4), satisfying the boundary conditions $W(t_f, x) = 0$, $x \in \mathbb{R}^n$, then $W(\tau, x)$ is a lower bound for the cost over $[\tau, t_f]$, i.e. $W(\tau, x) \leq V(\tau, x)$ where V denotes the value function as in Equation (D.10). Moreover, if there exists a $u^* \in L_\infty[t_0, t_f]^p$, the normed space of all essentially bounded Lebesgue measurable functions on $[t_0, t_f]$, such that

$$u^*(t) = \arg\min_{\mu \in \Omega} \{W_x^{\mathrm{T}}(t, x^*(t))Ax^*(t) + W_x^{\mathrm{T}}(t, x^*(t))B\mu + L(x^*(t), \mu)\} \tag{D.16}$$

for almost all $t \in [t_0, t_f]$, then u^* is optimal for initial (t_0, x_0). In (D.16), x^* is the state trajectory corresponding to (t_0, x_0) and input u^*. If such W is found, Equation (D.16) is a sufficient condition for optimality.

In Equation (D.16), with $W = V$, the continuous differentiability of the value function is required. Heemels et al. show that indeed V is differentiable with respect to x as well as to t. In addition, it is shown that

$$V_x(t_0, x_0) = 2 \int_{t_0}^{t_f} \mathrm{e}^{A^{\mathrm{T}}(t-t_0)} C^{\mathrm{T}} z_{t_0,t_f,x_0}(t)\, \mathrm{d}t$$

where z_{t_0,t_f,x_0} is the output corresponding to the initial condition (t_0, x_0) with optimal control u_{t_0,t_f,x_0}. This expression connects to the Maximum Principle as follows:

$$p_{t_0,t_f,x_0}(t) = V_x(t, x_{t_0,t_f,x_0}(t))$$

where p_{t_0,t_f,x_0} is the solution to the adjoint equation corresponding to u_{t_0,t_f,x_0}. If this $p_{t_0,t_f,x_0}(t)$ is substituted into Equation (D.13), one finds the time-varying optimal feedback given by

$$u_{t_0,t_f,x_0}(t) = P_\Omega\left(-\tfrac{1}{2}(D^{\mathrm{T}}D)^{-1}\left\{B^{\mathrm{T}}V_x(t, x_{t_0,t_f,x_0}(t)) + 2D^{\mathrm{T}}Cx_{t_0,t_f,x_0}(t)\right\}\right). \tag{D.17}$$

D.2.6. Comments on approximate aspects

Both methods described by Heemels et al. lead to a two-point boundary value problem. The maximum principle leads to an open-loop optimal control function for one particular initial condition. Dynamic programming does yield the preferred feedback law (D.17). However, this feedback involves solving a partial differential equation, which in general is complicated and time consuming. Therefore Heemels et al. investigate an alternative approach resulting in a feedback law.

The paper reveals that in Equation (D.11) only the projection $\mathcal{P}$ (the projection on $\mathcal{L}_{t_0,t_f}(P[t_0,t_f])$) cannot explicitly be computed. The approach is to approximate this projection, based upon the following theorem.

Theorem D.6 (Approximation of projections – *theorem 5 of [12]*) *Let K be a closed convex set in a Hilbert space H. Let $\{K_n\}_{n\in\mathbb{N}}$ be a sequence of closed convex subsets of K with $K_n \uparrow K (n \to \infty)$. Then for all $x \in H$ we have $P_{K_n}x \to P_K x (n \to \infty)$.*

Now, consider the discretization of the time domain into N time increments $h := \frac{t_f - t_0}{N}$ for some $N \in \mathbb{N}$, and $t_j = t_0 + jh$, $j = 0, \ldots, N$. Then define

$$P_h[t_0, t_f] := \left\{ u \in P[t_0, t_f] \middle| u|_{[t_j, t_{j+1}]} \text{ is constant, } j = 0, \ldots, N-1 \right\}.$$

Now, Heemels et al. state that

$$\mathcal{L}_{t_0,t_f}\left(P_h[t_0, t_f]\right) \uparrow \mathcal{L}_{t_0,t_f}\left(P[t_0, t_f]\right) \quad \text{as} \quad h \downarrow 0.$$

It follows from Theorem D.6 that the projection $P_{\mathcal{L}_{t_0,t_f}(P_h[t_0,t_f])}$ converges pointwise to $\mathcal{P} := P_{\mathcal{L}_{t_0,t_f}(P[t_0,t_f])}$. Using $u^h_{t_0,t_f,x_0} = \mathcal{L}_{t_0,t_f} P_{\mathcal{L}_{t_0,t_f}(P_h[t_0,t_f])}(-\mathcal{M}_{t_0,t_f} x_0)$, the following theorem follows:

Theorem D.7 (Convergence of the discretized optimal control – *Theorem 6 of [12]*) *The optimal discrete controls converge in the norm of L_2 to the exact optimal control when the time-step converges to zero. That is*

$$u^h_{t_0,t_f,x_0} \xrightarrow{L_2} u_{t_0,t_f,x_0} \quad \textit{as} \quad h \downarrow 0.$$

Each control $u^k \in P_h[t_0, t_f]$ can be parametrized as $u^h = \sum_{j=1}^N u^h_j \mathbb{1}_{[t_{j-1}, t_j)}$ with $u_j \in \Omega$. The corresponding discrete state trajectory for initial condition (t_0, x_0) is given by

$$\begin{aligned} x^h(j+1) &= \mathrm{e}^{Ah} x^h(j) + \int_{t_j}^{t_{j+1}} \mathrm{e}^{A(t_{j+1}-\tau)} B u^h(\tau)\,\mathrm{d}\tau = \mathrm{e}^{Ah} x^h(j) + \int_0^h \mathrm{e}^{A(h-\theta)} B\,\mathrm{d}\theta u^h_j \\ &=: A_h x^h(j) + B_h u^h_j \end{aligned}$$

where $x^h(j) := x(t_j)$ for $j = 0, \ldots, N$, and $x^h(0) = t_f$.

Heemels et al. also introduce a discrete version of the value function.

Definition D.5 (Discretized value function – *Definition 3 of [12]*) Fix times t_0, t_f such that $t_0 < t_f < \infty$. Fix time-step $h = \frac{t_f - t_0}{N}$ for some positive integer N. The function V^h from $\{0, \ldots, N\} \times \mathbb{R}^n$ to $\mathbb{R}$ is for $(j, x_0) \in \{0, \ldots, N\} \times \mathbb{R}^n$ defind by

$$V^h(j, x_0) = \min_{u^h \in P_h[t_j, t_f]} J(t_j, x_0, u^h)$$

where the cost function J is defined in Equation (D.8). ◇

The optimality principle states that the optimal control stays optimal along its trajectory. So the optimization problem can be solved backwards in time, because the tail part of the optimal control is optimal as well. More specifically for $j = 1, \ldots, N$

$$V^h(j-1, x) = \min_{v \geq 0} \left\{ V^h(j, A_h x + B_h v) + \int_{t_{j-1}}^{t_j} z^{\mathrm{T}}_{j,v}(t) z^{\mathrm{T}}_{j,v}(t)\,\mathrm{d}t \right\}$$

where $z_{j,v}$ is the output of (D.5) and (D.6) with initial condition (t_{j-1}, x) and control identically equal to $v \in \mathbb{R}^m$ on the interval $[t_{j-1}, t_j)$. Since $V^h(N, x) = 0$ for all $x \in \mathbb{R}^n$, the value function V^h van be determined recursively, and the optimal control values for every point (j, x) can be stored.

An open problem is how to choose h.

D.2.7. *Extension to infinite horizon case*

Consider Problem 2. For the infinite-horizon case the following assumptions are made:

1. (A, B) is positively stabilizable;
2. D has full column rank;
3. (A, B, C, D) is minimum phase.

The assumption of minimum phase is needed to get the convergence between the finite and infinite horizon problem.

In Heemels et al. [12] the operators $\Pi_{t_f} : L_2[0,\infty)^m \to L_2[t_0, t_f]^p$ for $u \in L_2[0,\infty)^m$ and $\Pi^*_{t_f} : L_2[t_0, t_f]^p \to L_2[0,\infty)^m$ for $u \in L_2[t_0, t_f]^p$ are defined by

$$(\Pi_{t_f} u)(t) = u(t), \quad t \in [0, t_f]$$

$$(\Pi^*_{t_f} u)(t) = \begin{cases} u(t), & t \le t_f \\ 0, & t > t_f \end{cases}$$

Theorem D.8 (Convergence of V to infinite horizon – *Theorem 8 of [12]*) *For all $x_0 \in \mathbb{R}^n$, there holds that $V^T(x_0) \to V^\infty(x_0)(t_f \to \infty)$, where $V^T(x_0) = V^T(0, x_0)$. For $T \to \infty$ we have $\Pi^*_{t_f} u_{t_f} \to u_\infty$, $\Pi^*_{t_f} z_{t_f} \to z_\infty$ and $\Pi^*_{t_f} x_{t_f} \to x_\infty$ in the L_2-norm.*

Also the maximum principle was extended to the infinite horizon case. It is shown that the optimal controls also converge pointwise.

Theorem D.9 (*Theorem 9 of [12]*) *The optimal control u_∞ corresponding to initial conditions (t_0, x_0) satisfies*

$$u_{t_0,t_f,x_0} = P_\Omega \left(-\tfrac{1}{2} (D^\mathrm{T} D)^{-1} \left\{ B^\mathrm{T} p_\infty(t) + 2 D^\mathrm{T} C x_\infty(t) \right\} \right) \tag{D.18}$$

where the continuous function $p_\infty \in L_2[0,\infty)^n$ is given by

$$\dot{p}_\infty = -A^\mathrm{T} p_\infty - 2 C^\mathrm{T} z_\infty$$

for some initial condition $(0, p_\infty(0))$. Moreover, u_∞ is a continuous function.

Furthermore, another result is that *for all $\tau > 0 : \Pi_\tau u_{t_f} \to \Pi_\tau u_\infty (t_0 \to \infty)$ in $L_\infty[0,\tau]^m$.*

In the finite horizon problem, the optimal control could be given by Equation (D.17). In contrast with the finite horizon, the time-dependence vanishes in the infinite horizon case. It is obvious that for all $t \ge 0$ we have that $u_{\tau,\infty,x_0}(t+\tau) = u_{0,\infty,x_0}(t)$. From this, it follows that there exists a time-invriant optimal feedback $u_{f\,\mathrm{db}}$ defined by

$$u_{f\,\mathrm{db}}(x_0) := u_{\tau,\infty,x_0}(\tau) = u_{0,\infty,x_0}(0) = \lim_{t_f \to \infty} u_{0,t_f,x_0}(0).$$

Now, it t_f is large enough, u_{0,t_f,x_0} can be used as an approximation for $u_{f\,\mathrm{db}}(x_0)$. An unsolved problem at the moment of writing is how to choose t_f such that it is large enough.