GUARANTEED RESILIENCE OF AUTONOMOUS SYSTEMS TO PARTIAL LOSS OF CONTROL AUTHORITY OVER ACTUATORS

BY

JEAN-BAPTISTE BOUVIER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Aerospace Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2023

Urbana, Illinois

Doctoral Committee:

Assistant Professor Melkior Ornik, Chair
Assistant Professor Robyn Woollands
Professor Cedric Langbort
Professor Daniel Liberzon

Abstract

After docking to the International Space Station (ISS), the Nauka module suffered a software error causing its thrusters to misfire. In turn, these uncontrolled thrusters rotated the whole space station by 540° before being counteracted by other thrusters of the ISS. Motivated by such a scenario, this thesis investigates the guaranteed resilience of autonomous systems to a similar class of malfunctions called partial loss of control authority over actuators. These malfunctions are characterized by actuators producing uncontrolled and undesirable outputs instead of following the controller's commands. A loss of control authority can be caused, for instance, by a software bug as in the ISS example or by an adversarial takeover of some actuators of the system. In this setting, we investigate the malfunctioning system's remaining capabilities to complete its mission in terms of resilient reachability and resilient trajectory tracking. We also quantify the resilience of linear systems by comparing the reachability performance of the nominal dynamics with that of the worst-case malfunctioning dynamics. We extend our resilience investigation to systems further inflicted with actuation delays preventing an immediate cancellation of the undesirable outputs. We illustrate our theory on a wide range of applications including an octocopter, a fighter jet model, and an orbital inspection mission.

To my family.

Acknowledgments

This thesis would not have been possible without the support of many people. Many thanks to my advisor, Melkior Ornik, who guided me, reviewed all my papers and continually supported me throughout these four years of PhD. I also want to thank all the members of my thesis committee for their guidance and advice, Profs. Robyn Woollands, Cedric Langbort, and Daniel Liberzon. I want to thank my fellow group members Pranay, Farhad, Taha, Hamza, Alen, Karan, Gokul and Manav who were a great support group. Thanks to all my friends in the Fighting Illini Triathlon team and to my roommate Matt. Finally, I would like to thank my girlfriend Emily and my family for their encouragement throughout this long process.

Table of contents

List of Abbreviations

FTCS	Fault-Tolerant Control System
FDI	Fault Detection and Isolation
GPS	Global Positioning System
HJB	Hamilton-Jacobi-Bellman equation
ISS	International Space Station
KOS	Keep-Out Sphere
NASA	National Aeronautics and Space Administration
UAV	Unmanned Aerial Vehicle

Notation

$\in$	belong to
$\subseteq$	subset
$\cup$	union
$\cap$	intersection
$:=$	defined as

$\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$	fields of natural, real and complex numbers
$\mathbb{R}^n$	field of real vectors of dimension n
$\mathbb{R}^+$	nonnegative real numbers
$\mathbb{R}^+_*$	positive real numbers
$\llbracket a, b \rrbracket$	inclusive set of integers between $a \in \mathbb{N}$ and $b \in \mathbb{N}$
$\emptyset$	empty set
$Re(\mathcal{S}) \leq 0$	real part of all $s \in \mathcal{S}$ is nonpositive
$Re(\mathcal{S}) = 0$	real part of all $s \in \mathcal{S}$ is null
$\|x\| := \sqrt{\sum x_i^2}$	Euclidean norm of a vector $x = (x_1, ..., x_n) \in \mathbb{R}^n$
$\|x\|_\infty := \max \|x_i\|$	infinity-norm of a vector $x = (x_1, ..., x_n) \in \mathbb{R}^n$
$\|\cdot\|_X$	canonical norm on a space X other than $\mathbb{R}^n$

I_n	$n \times n$ identity matrix
$0_{n,m}$	$n \times m$ zero matrix
$\mathbf{1}_n$	$n \times 1$ vector of ones
$\mathrm{Im}(A)$ and $\mathrm{Ker}(A)$	image and kernel of a matrix $A \in \mathbb{R}^{n \times m}$
$\mathrm{rank}(A) = \dim \mathrm{Im}(A)$	rank of matrix A
$\det(A)$	determinant of matrix A
$A^\top$ and A^{-1}	transpose and inverse of matrix A
$\|A\| := \sup\limits_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max\limits_{\|x\|=1} \|Ax\|$	norm of matrix A
$A \succ 0$ and $A \succeq 0$	positive definite and positive semidefinite matrix A
$\lambda(A)$ and $\sigma(A)$	set of eigenvalues and singular values of A
$\left(\lambda^A_{min}, \lambda^A_{max}\right)$ and $\left(\sigma^A_{min}, \sigma^A_{max}\right)$	extremal eigenvalues and singular values of A
$\|x\|_A := \sqrt{x^\top A x}$	A-norm of vector $x \in \mathbb{R}^n$ if $A \succ 0$
$\mathcal{C}(A, B) := \left[BAB \dots A^{n-1}B\right]$	controllability matrix of pair (A, B)

$\mathbb{S} := \{x \in \mathbb{R}^n : \|x\| = 1\}$	unit sphere in $\mathbb{R}^n$
$\mathbb{B}_X(c, \varepsilon) := \{x \in X : \|x - c\| \leq \varepsilon\}$	closed ball of center c and radius ε in the space X
$\mathcal{E}(c, P) := \{x : (x - c)^\top P(x - c) \leq 1\}$	ellipsoid of center c and shape matrix $P \succ 0$

$\binom{m}{p}$	number of p-combinations among m elements for $p \leq m \in \mathbb{N}$
$p!$	factorial of $p \in \mathbb{N}$

$\mathrm{span}(\cdot)$	mapping of vectors to their linear span
$\langle \cdot, \cdot \rangle$	standard scalar product in $\mathbb{R}^n$

$\|u\|_{\mathcal{L}_2}^2 := \int_0^T \|u(t)\|^2 dt$	$\mathcal{L}_2$-norm for $T > 0$
$\mathcal{L}_2\big([0, T],\ \mathbb{R}^m\big)$	space of square integrable functions $u : [0, T] \to \mathbb{R}^m$
$\mathcal{L}\big(X, Y\big)$	space of continuous linear maps from a space X into space Y
$X^* := \mathcal{L}\big(X, \mathbb{R}\big)$	topological dual space of a Banach space X
$x^* \in X^*$	dual vector of $x \in X$, i.e., associated linear form from X to $\mathbb{R}$
$S^* \in \mathcal{L}\big(Y^*, X^*\big)$	adjoint linear map of $S \in \mathcal{L}\big(X, Y\big)$

$\partial \mathcal{X}$	boundary of a set $\mathcal{X}$
$\mathrm{int}(\mathcal{X}) := \mathcal{X} \backslash \partial \mathcal{X}$	interior of $\mathcal{X}$
$\mathrm{relint}(\mathcal{X})$	relative interior of $\mathcal{X}$ as defined in [1]
$-\mathcal{X} = \mathcal{X}$	symmetric set $\mathcal{X}$
$\mathrm{co}(\mathcal{X})$	convex hull of $\mathcal{X}$
$\mathcal{X} \oplus \mathcal{Y} := \{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$	Minkowski addition of sets $\mathcal{X}$ and $\mathcal{Y}$
$\mathcal{X} \ominus \mathcal{Y} := \{z : \{z\} \oplus \mathcal{Y} \subseteq \mathcal{X}\}$	Minkowski difference

$\mathcal{F}(\mathcal{X}) := \{f : f(t) \in \mathcal{X} \text{ for all } t \geq 0\}$	set of functions $f : \mathbb{R}^+ \to \mathcal{X}$
$f \circ g = f(g)$	composition of functions f and g
$\dot{x}(t) := \frac{d}{dt} x(t)$	time derivative of a function x
$f^{(k)}$	k^{th} derivative of function f for $k \in \mathbb{N}$
$\mathrm{proj}_r(x_1, \ldots, x_n) := (x_1, \ldots, x_r) \in \mathbb{R}^r$	projection map from $\mathbb{R}^n$ onto $\mathbb{R}^r$ with $r \leq n$

Chapter 1

Introduction

Autonomous systems are becoming ever more widespread as they accomplish increasingly varied and complex tasks such as controlling utility systems [2]–[4], industrial processes [5], [6], driving cars [7], flying aircraft [8]–[10] and controlling spacecraft [11]–[13]. Because of the critical nature of these applications, their autonomous controllers must be reliable and guarantee safe operations even after possible malfunctions. Let us describe some motivating examples of malfunctions impacting safety critical autonomous systems.

On July 29^{th} 2021, after the Nauka module docked to the International Space Station (ISS), a software failure caused a misfire of all the module's thrusters [13]. As a result, the whole station lost attitude control for 15 minutes and rotated by 540° possibly endangering the ISS crew. Eventually, other thrusters on the ISS were fired to counteract the uncontrolled and undesirable thrust until the Nauka module ran out of fuel.

Commercial aircraft can also be considered as safety critical autonomous systems since they are mostly operated by autopilots. However, these autopilots are not designed to fly in off-nominal conditions, where they disengage sometimes without advance warning [9]. Such off-nominal situations may arise in upset flight conditions, after a loss of control effectiveness over some actuators, or after changes of the airplane's dynamics. These are the primary causes of in-flight loss of control, which is the largest fatal accident category for large commercial jets [9]. More specifically, a loss of control effectiveness over a flight control surface can be caused by a leak in the hydraulic system responsible for moving said surface [14]. As a result the pilot has less control over the aircraft and might not be able to fly it safely anymore.

Autonomous systems' malfunctions are not always accidental as in the previous two examples, they can also be caused purposefully by adversarial attacks. Indeed, control systems are increasingly connected to the Internet [6], and hence they are more prone to cyber attacks. A well documented example is the attack on the Maroochy sewage control system in Australia [2], [5], [6], where a disgruntled ex-employee took remote control over sewer valves and managed to flood a hotel, a park and a river with a million liters of sewage. Researchers have also demonstrated the vulnerability of national power networks to cyber attacks [5], [15] as a result of the increased connectivity necessary to build a smart grid. Work [5] also discuses cyber attacks on the autonomous control systems of gas pipeline and other power utilities.

1.1 Framework

Motivated by these accidental and orchestrated failures of autonomous systems, we want to build a resilience framework to guarantee the safety of autonomous systems in the face of such malfunctions. We start by defining a precise malfunction model encompassing all aforementioned scenarios. We introduce the notion of *partial loss of control authority over actuators* where some of the actuators of an autonomous system start producing uncontrolled and thus possibly undesirable inputs within their full range of actuation. These actuators do not obey to the controller's commands anymore. Such a failure can be caused, for instance, by a software bug as in the ISS example [13], by a loss of pressure in hydraulic powered control surfaces of an aircraft [9], [14] or by a cyber attack [2], [5], [6], [15].

After an autonomous system lost control authority over some of its actuators, the malfunctioning system must first identify these faulty actuators. This is the role of the Fault Detection and Isolation (FDI) module popularized by the field of fault-tolerant control [16], [17]. Following [16], we assume that sensors monitor in real-time the outputs of each actuator of the system. Then, a model-based FDI module compares each actuator's output with its nominal prediction to assess whether a fault occurred as in [16].

Once the malfunctioning actuators are identified, the sensors used by the FDI module still play a crucial role by providing the controller with real-time measurements of the malfunctioning actuators' outputs. With this information, an adaptive controller will use the remaining controlled actuators to counteract the deleterious effect of the malfunctioning actuators, as in the aforementioned ISS example [13]. This approach is only feasible if the system possesses sufficient control redundancy. Since duplicating actuators is expensive, control redundancy is usually restricted to safety critical systems like aircraft [16] or spacecraft [18]. This trade-off between redundancy and cost has been investigated by NASA during the space race in order to provide guidance for spacecraft design [19]. More specifically, [19] provides a method to calculate how much redundancy is necessary in a subsystem by minimizing the total cost of the subsystem plus the expected loss incurred in case of failure.

Classically, the design of controllers operating in off-nominal conditions has relied on robust control theory [20] and on adaptive control theory [21]. Combining approaches from these two theories, the field of fault-tolerant control provides tools to study various actuator failures [17]. However, partial loss of control authority over actuators has so far largely escaped treatment by any of these three control theories. We will review related works of robust, adaptive and fault-tolerant control in Section 2.1 and discuss their limitations when applied to our malfunction of interest. Instead of arising from these classical control theories, the most useful contributions to this dissertation came from the fields of differential games theory [22], [23] and perturbed reachability analysis [24] which we will discuss in Section 2.3.

Let us introduce some notation to describe the framework employed throughout this dissertation. We study a *nominal* control system whose state at time t is denoted by $x(t)$ and follows dynamics

$$\dot{x}(t) = f\big(\, x(t),\ \bar{u}(t)\,\big), \qquad x(0) = x_0, \qquad \bar{u}(t) \in \bar{\mathcal{U}} \quad \text{for all } t \geq 0. \tag{1.1}$$

State x starts from x_0 at $t = 0$ and is then steered through function f by nominal input $\bar{u}$ from the controller. The actuators of system (1.1) all operate within some range that might be dictated, for instance, by energy or mechanical constraints. These constraints are enforced on the inputs of system (1.1) through the inclusion $\bar{u}(t) \in \bar{\mathcal{U}}$.

After a partial loss of control authority over actuators, nominal input signal $\bar{u}$ is split in two parts: controlled input signal u from the controlled actuators and undesirable input signal w from the malfunctioning actuators. The dynamics of the *malfunctioning* control system are described by the same function f as in (1.1), but the input $\bar{u}$ is now constituted of two parts u and w, so that (1.1) becomes

$$\dot{x}(t) = f\big(\, x(t),\ \big[u(t)\, w(t)\big]\,\big), \qquad x(0) = x_0, \qquad u(t) \in \mathcal{U}, \qquad w(t) \in \mathcal{W} \quad \text{for all } t \geq 0. \tag{1.2}$$

The same way as $\bar{u}$ is split into u and w, set $\bar{\mathcal{U}}$ is split into $\mathcal{U}$ and $\mathcal{W}$ to separate the controlled actuators from the malfunctioning ones. Then, the input constraints on the malfunctioning dynamics (1.2) are enforced through the inclusions $u(t) \in \mathcal{U}$ and $w(t) \in \mathcal{W}$. The loss of control authority is clearly illustrated by the fact that the controller chooses $\bar{u}(t)$ in (1.1), but it chooses only $u(t)$ in (1.2), while the undesirable input $w(t)$ is determined by other uncontrolled factors.

As discussed previously, the sensors of the FDI module provide real-time measurements of the malfunctioning actuators' outputs $w(t)$ to the controller, which is able to modify its command $u(t)$ to adapt in a reactive fashion to $w(t)$. This dependency can be written simply as $u(t) = u\big(t, w(t)\big)$. Control systems usually possess additional sensors like GPS, cameras, radars, etc, to measure at least partially the state $x(t)$. In this dissertation we assume that the controller has perfect knowledge of the state, which translates to $u(t) = u\big(t, x(t), w(t)\big)$. If the controller is also able to store in memory the history of the state $x([0,t])$ and of the undesirable input $w([0,t])$, this additional knowledge might be beneficial for performance. Then, we would write $u(t) = u\big(t, x([0,t]), w([0,t])\big)$. Since $x(t)$ and $w(t)$ are acquired through sensors, in practice they are received by the controller with some delay [25]–[27]. This setting will be investigated in Chapter 8 where we remove the simplifying assumption of instantaneous knowledge of $x(t)$ and $w(t)$ by the controller.

Now that the investigation framework is clearly stated, let us define our problems of interest.

1.2 Problems of interest

Because of a loss of control authority over some actuators, malfunctioning system (1.2) has less actuation than nominal system (1.1), and hence system (1.2) might not be able to complete the nominal mission assigned to system (1.1). To assess the remaining capabilities of the malfunctioning system we will first consider simple missions of target reachability. Let $\mathcal{T}$ be a target set that we assume to be reachable by the nominal system (1.1). Indeed, if $\mathcal{T}$ is not reachable by the nominal system, then it cannot be reached by malfunctioning system (1.2) because it has the same dynamics and reduced actuation. Reachability of $\mathcal{T}$ by the nominal system means that there exists a control input $\bar{u}$ driving the state of (1.1) from x_0 to $x(T_N) \in \mathcal{T}$ in some time $T_N \geq 0$.

For malfunctioning system (1.2), we need to introduce a slightly different notion of reachability. Indeed, $w(t)$ is uncontrolled and possibly unpredictable, hence reachability by system (1.2) should not depend on w. We say that target $\mathcal{T}$ is *resiliently reachable* by system (1.2) if for every signal w there exists a control u driving the state of (1.2) from x_0 to $x(T_M) \in \mathcal{T}$ in some time $T_M \geq 0$. The main obstacle preventing resilient reachability is the actuation constraint $u(t) \in \mathcal{U}$ that might prevent the controller from overcoming some undesirable inputs w and adequately steering the state to $\mathcal{T}$. This leads us to our first problem of interest.

Problem 1: Under what conditions is a target $\mathcal{T}$ resiliently reachable by malfunctioning system (1.2)?

We first tackle this question for systems with energy bounded inputs in Chapters 3 and 4, and for systems with amplitude bounded inputs in Chapters 5 and 7. For these two types of systems, we employ drastically different approaches. We rely on perturbed reachability theory [24] for systems with inputs of bounded energy, while we build on differential games theory [23] for systems with inputs of bounded amplitude.

Solving Problem 1 gives a method to study the remaining capabilities of a system that has suffered a partial loss of control authority over its actuators. When studying safety-critical systems, such a post-failure analysis should conclude that the malfunctioning system is still capable of completing its nominal mission. Therefore, safety-critical systems must be designed *resilient* to a partial loss of control authority over their actuators. This train of thoughts leads us to our second problem of interest.

Problem 2: How to design a system that can resiliently complete its mission despite a loss of control authority over any one of its actuators?

Building on the resilient reachability conditions of Chapter 3, we address Problem 2 in Chapter 4 for driftless linear systems with inputs of bounded energy. We are then able to design resilient systems capable of completing their nominal mission despite a partial loss of control authority over their actuators.

However, resilience only requires that target set $\mathcal{T}$ remains reachable by malfunctioning system (1.2) in some finite time T_M as discussed before Problem 1. Resilience does not place any constraint on T_M other than being finite. After an extremely damaging loss of control, T_M could be several orders of magnitude larger than T_N, the nominal reach time for system (1.1). Many applications have some time constraints that would prevent mission completion if T_M was too large. For instance, drones have a limited flight time due to their battery [28] and spacecraft have limited quantities of propellant on board [12], which prevent mission extension after some fixed timeframe.

Our third objective is then to estimate the maximal time penalty caused by a partial loss of control authority. More specifically, we introduce the *nominal reach time* T_N^* as the fastest time in which nominal system (1.1) can reach target set $\mathcal{T}$. Similarly, we define the *malfunctioning reach time* T_M^* as the fastest time in which malfunctioning system (1.2) can reach target set $\mathcal{T}$ when w is chosen to make that time the longest. Then, we can quantify the resilience of system (1.1) by studying the ratio T_N^*/T_M^*. The larger this ratio is, the less impact the loss of control authority has on system (1.1) and hence the more resilient it is. To quantify the resilience of system (1.1) independently of its mission, we will define *quantitative resilience* as the maximal ratio T_N^*/T_M^* over all targets $\mathcal{T}$.

The only thing left is to solve this optimization problem of the ratios of reach times. However, these ratios are nonlinear in their dependency on target sets $\mathcal{T}$ making this optimization non-trivial. Additionally, our definitions of T_N^* and T_M^* introduce nested optimization problems rendering quantitative resilience an extremely difficult quantity to calculate. Our third problem of interest is then to find a method of solving these nonlinearly nested optimizations.

Problem 3: How to calculate efficiently the quantitative resilience of control systems?

In Chapter 5 we establish a method to solve Problem 3 for driftless linear systems by relying on the novel Maximax Minimax Quotient Theorem that we prove in Chapter 6. However, this approach does not extend to the case of linear systems with drift, mainly due to the lack of analytical expression for the minimal reach times T_N^* and T_M^* of these systems [29], [30]. An exact calculation of quantitative resilience being impossible, Chapter 7 instead establishes bounds on the nominal and malfunctioning reach times T_N^* and T_M^* in order to bound the quantitative resilience of linear systems with drift.

Problems 1, 2 and 3 are at the core of the resilience theory established in this dissertation. The solutions of these foundational problems derived between Chapters 3 and 7 open the door for numerous extensions of resilience theory, which leads to our final problem of interest.

Problem 4: How to extend the scope of resilience theory?

We investigate Problem 4 in Chapters 8 to 10 where we explore a few of the possible extensions of resilience theory. More specifically, in Chapter 8 we detail how to adapt resilience when removing the assumption of instantaneous knowledge of $w(t)$ by controller $u(t)$. We also extend our theory to more complex mission scenarios than resilient reachability by studying resilient trajectory tracking. In Chapter 9, we further extend the scope of our theory to study the resilience of linear networks. Despite most of the work of this dissertation concerning linear dynamics, we show in Chapter 10 that resilience theory can be extended to nonlinear systems such as spacecraft dynamics. We have now stated the four problems of interest to be studied in this dissertation. Let us give an overview of the remainder of this work.

1.3 Overview

Following Problems 1 to 4, the contributions of this dissertation are fourfold and summarized as follows.

1. We establish analytical conditions to assess whether autonomous systems enduring a partial loss of control authority over their actuators remain capable of reaching a given target despite any undesirable input generated by the malfunctioning actuators.

2. We derive design criteria for autonomous systems to be resilient to the loss of control authority over any one of their actuators.

3. We quantify the impact of such a malfunction by comparing the optimal reach times of autonomous systems before and after a partial loss of control authority over their actuators.

4. We further extend the scope of resilience theory by establishing conditions for resilient trajectory tracking, resilience in the presence of actuation delay, resilience of networks and for the resilience of nonlinear systems.

These contributions are connected to our problems of interest discussed at length in Section 1.2, where we also mention which chapters of this dissertation address each problem. Similarly, each of these contributions is realized in the chapters tackling the associated problem of interest.

This dissertation is composed of the work accomplished between Fall 2019 and Spring 2023 and relies on four of our conference papers [31]–[34], our four journal papers [35]–[38] and two yet unpublished works [39], [40]. During these four years we also published another conference paper [4] but it does not fit within the framework of this dissertation. The remainder of this work is divided into ten chapters organized as follows.

- Chapter 2 is a literature review of the various fields related to resilience theory. We will study how resilience fits within the wider approaches of robust, adaptive and fault-tolerant control. We will also compare our theory with other notions of resilience found in the literature. Finally, we will introduce previous works studying reachability, controllability, differential games theory, and time optimal linear control upon which resilience theory is built.

- Chapter 3 is adapted from our earliest work on resilience [31] to address Problem 1. In this chapter we will establish the foundational resilience theory for linear systems with bounded energy. We will build on the perturbed reachability condition of [24] by transforming it into a usable form. We will then use this reworked condition to derive simple analytical conditions for the resilient reachability of driftless linear systems and to understand how their resilient reachability capabilities evolves with time.

- Chapter 4 relies on our work [35] to investigate the design of resilient driftless linear systems with bounded energy and hence solve Problem 2. In this chapter we will build on the resilient reachability condition of Chapter 3 to calculate the minimal degree of overactuation necessary for a system to be resilient to the loss of control over any single one of its actuators. Additionally, we will synthesize a control law achieving resilient reachability for linear systems.

- Chapter 5 switches gear to study the resilience of linear systems with component bounded inputs and to introduce the notion of quantitative resilience. This chapter is adapted from our published works [32], [38] and will use linear optimal control to design an efficient method to calculate the quantitative resilience of driftless linear systems. This chapter is then geared toward addressing Problem 3.

- Chapter 6 describes the proof of the Maximax Minimax Quotient Theorem. This optimization result is in fact needed to calculate the quantitative resilience of driftless systems and used in Chapter 5. The complete proof of this result was published in our work [36] and relies on a geometrical approach of input selection. Before solving said optimization problem we will prove the existence of a solution with the Berge Maximum theorem [1].

- Chapter 7 extends resilience theory to general linear systems with drift by addressing Problems 1 and 3. This chapter is drawn from our works [33], [37] which rely on differential games and linear control theories to establish necessary and sufficient conditions for the resilience of general linear systems. We will also calculate analytical bounds on the quantitative resilience of these systems.

- Chapter 8 represents the first step towards Problem 4 and is taken from our work [39]. We will start this chapter by investigating more complex mission scenarios than resilient reachability by deriving sufficient conditions for resilient trajectory tracking. Then, we will we extend resilience theory to linear systems with actuation delays to remove the assumption of instantaneous knowledge of the undesirable inputs by the controller. Finally, we will derive a sufficient resilience condition for systems with nonlinear dynamics.

- Chapter 9 extends resilience analysis to linear networks suffering partial loss of control authority. We will mostly study how an unresilient subsystem suffering from a partial loss of control authority can affect the stabilizability of the rest of the network. This chapter represents the current state of our work and has not been published yet.

- Chapter 10 investigates the resilience of an orbital inspection mission to the loss of control authority over a thruster of the inspecting spacecraft. This chapter is drawn from our works [34], [39] and contributes to Problem 4 by extending resilience theory to the nonlinear dynamics of a spacecraft. For these nonlinear dynamics we will also build a resilient trajectory tracking controller with guaranteed performance.

- Finally, Chapter 11 contains a summary of this dissertation as well as the main conclusions. It also gives recommendations to be considered for future research on the topic of resilience.

This concludes the introduction of this dissertation. We will now proceed with an extensive literature review of the various fields related to resilience theory.

Chapter 2

Literature Review

In this chapter, we proceed to an extensive overview of previous works in relation with our dissertation. Since resilience theory studies how well autonomous systems can endure a partial loss of control authority over their actuators, we need to compare resilience with the classical control theories studying changing or unknown dynamics, most notably robust and adaptive control. We will also compare our resilience theory within the resilience literature. Finally, since reachability is central in this dissertation, we review the approaches of robust control, optimal control and differential games to the topic of reachability analysis.

The remainder of this chapter is organized as follows. In Section 2.1, we will study how resilience compares with the well-established fields of robust, adaptive and fault-tolerant control. In Section 2.2, we compare our theory with other notions of resilience found in the literature. Finally, Section 2.3 introduces previous works studying various notions of reachability as this is a central topic of our dissertation.

2.1 Control theories for malfunctioning systems

In Chapter 1, we introduced the notion of resilience of autonomous systems to a partial loss of control authority over their actuators. Following such a malfunction, the faulty actuators produce uncontrolled and possibly undesirable inputs, which can severely hinder the capabilities of the system. Classically, changing or unknown dynamics are studied through robust and adaptive control theories grouped together under the wider umbrella of fault-tolerant theory. We will then review these theories and see how resilience fits within their framework. We start by reviewing the basis of robust control theory.

2.1.1 Robust control theory

A control system with guaranteed performance despite the presence of unmodeled and unknown disturbance is characterized as *robust* to these uncertainties [20]. Robust control methods are designed to function properly provided that uncertain parameters or disturbances are found within some typically compact set. To illustrate how robust theory works, let us introduce control system (2.1), whose state x follows dynamics f and is steered by control input u and unknown disturbance v. Then,

$$\dot{x}(t) = f\big(x(t), u(t), v(t)\big), \qquad x(0) = x_0, \quad u(t) \in \mathcal{U}, \quad v(t) \in \mathcal{V} \quad \text{for all } t \geq 0, \tag{2.1}$$

where $\mathcal{V}$ is a bounded set. We can now define *robust reachability*, also called *strong reachability* in [24], [41], *reachability under uncertain input disturbance* in [42] or *minimax reachability* in [43]. A target $\mathcal{T}$ is robustly reachable by system (2.1) if there exists an admissible control signal $u \in \mathcal{F}(\mathcal{U})$ such that for all disturbance signal $v \in \mathcal{F}(\mathcal{V})$ the state of system (2.1) reaches target $\mathcal{T}$ in some time T, i.e., $x(T) \in \mathcal{T}$.

The main strength of this approach is the absence of assumptions restricting disturbance v other than its boundedness $v(t) \in \mathcal{V}$ [44]. The disturbance input v can be stochastic, bang-bang, constant, or anything in between, as long as it remains in $\mathcal{V}$, a robust controller does not need to know its structure to drive the state to the target. However, this strength is also the main weakness of robust control. Indeed, the utter lack of knowledge concerning v forces robust controller u to be overly conservative. In order to obtain meaningful guarantees for robust reachability, the constraint set $\mathcal{V}$ must be sufficiently small. Hence, we should think of v as modeling small perturbations not accounted for in the dynamics f [44].

Let us now see how resilience compares with robust control. If we label as disturbances the uncontrolled inputs w produced by the malfunctioning actuators of a system enduring a partial loss of control authority like (1.2), then resilience theory would fit in the framework of robust control [20], [41], [42]. Indeed, note the similarity between the dynamics of malfunctioning system (1.2) and that of system (2.1). We will now compare robust reachability with resilient reachability as defined above Problem 1. Both of these concepts of reachability ask for a controller u to drive the state x to target $\mathcal{T}$ despite some uncontrolled input v or w. However, the crucial distinction between robust and resilient reachability lies in the order of the quantifiers. Robust reachability requires a single controller u to work for any disturbance v, while resilient reachability allows the controller u to depend on the uncontrolled input w. This dependency is justified by the assumption of instantaneous knowledge of $w(t)$ by $u(t)$ made in Section 1.1. Since the robust control setting treats v as unknown, it has access to much less information than a resilient controller allowed to adapt to the undesirable inputs w. Because of this crucial difference in information setting a resilient controller performs much better than an overly conservative robust controller, as we will later demonstrate in Section 4.6.2.

Another caveat hindering the application of robust control to a system enduring a partial loss of control authority over its actuators, is the magnitude of the undesirable inputs w. Indeed, these inputs are produced by actuators of the system and hence they can have the same magnitude as the controlled inputs. However, robust control needs the undesirable inputs to be significantly smaller than the controls to provide meaningful results [44]. Therefore, even if robust control methods can be applied to study systems enduring a partial loss of control authority over their actuators, these methods are too conservative to produce the desired reachability guarantees as they do not take advantage of the information available on the undesirable inputs.

We will now investigate whether the rival theory of adaptive control is more appropriate to study a loss of control authority over actuators.

2.1.2 Adaptive control theory

A controller that is autonomously modifying its parameters and structure in order to improve its performance is characterized as *adaptive* [21], [44]. These modifications aim at either adapting to time-varying parameters of the system or at learning the value of some initially uncertain system parameters. Adaptive control is different from robust control in that it does not need a priori information about the bounds on these uncertain or time-varying parameters [44]. Instead of having a single and conservative robust controller, adaptive control is concerned with how to change the control law itself to be better suited to changed or uncertain system dynamics.

This adaptation of the control law to time-varying parameters is present in the resilience framework as a resilient controller is allowed to modify its response when the undesirable inputs change. Hence, resilience theory is more closely affiliated to adaptive control than to robust theory. However, a straightforward application of adaptive control methods to a system enduring a partial loss of control authority over its actuators is likely to fail. Indeed, adaptive control tries to estimate unknown parameters before they have time to change significantly [21], which may not be possible for uncontrolled inputs that might operate on faster timescales. This timescale issue is one of the main limitations of adaptive control [21]. Such a situation would typically prevent convergence of the estimators and lead to mediocre adaptive control performance [44]. This apparent limitation of adaptive control did not discourage researchers who found ways of applying adaptive methods to compensate actuator failures [45], [46] under the wider umbrella of fault-tolerant theory.

Now that we have given some background on robust and adaptive control theories, let us see how they are implemented in the framework of actuator malfunctions, which is of particular interest to this dissertation.

2.1.3 Fault-tolerant theory

A control system capable of automatically preserving its stability and achieving acceptable performance despite component malfunctions is called a fault-tolerant control system (FTCS) as defined in [14]. These systems can be separated into two broad categories: *passive* and *active* FTCS [14], [17].

- In a *passive* FTCS all failure modes are known a priori and are all implemented together beforehand. Passive FTCS usually rely on robust control theory with a single controller operating both before and after any failure [46]. This method has the advantage of not requiring any information about the failure [17]. Then, its implementation does not require a fault detection and isolation (FDI) module, which makes it very appealing in practice [14]. Additionally, the controller being unique, there is no switching, adaptation period or transient during which performance could be unsafe as for active FTCS [14]. However, the main drawback of this approach is its conservativeness as the controller is built for the worst-case scenario and becomes more conservative as the number of considered fault scenarios increases [14].

- An *active* FTCS reacts to a detected fault and reconfigures its controller to adapt to the changed dynamics. This approach relies on adaptive control theory and switched systems to provide better performance than a passive FTCS [46]. Indeed, the adaptive controller will be designed specifically to perform well in a precise setting [14]. However, an active FTCS requires fast and correct diagnostic of the malfunction in order to adapt adequately. The necessary FDI module increases the design and implementation complexities of active FTCS. Another drawback to this approach is the need to establish stability guarantees for the transient period occurring after the malfunction, during the fault-diagnostic, and before the switching to the updated controller [46].

As discussed during the overview of robust control in Section 2.1.1, a loss of control authority generates undesirable inputs of magnitude too large to be meaningfully handled by a passive FTCS. We will then focus on active approaches. Indeed, this dissertation belongs to the framework of active fault-tolerant control since our approach relies on the detection of actuator malfunctions and switches to a controller specifically designed to counteract said actuator failure. Actuator failures as a whole are too broad to be studied together and are then usually subdivided into three main categories detailed in [17], [18].

- An actuator suffering from a *partial loss of effectiveness* is operating with reduced capability compared to its nominal range, but remains under the controller's authority [10], [14], [18], [46], [47]. For instance, a leak in the hydraulic system responsible for the motion of an actuator would cause a pressure loss effectively reducing the range of motion of said actuator [14]. To handle loss of actuator effectiveness, the work [47] establishes a new adaptive control allocation method that does not require unrealistic persistence of excitation like classical adaptive control, but requires actuator redundancies. When no actuator redundancy is available, work [18] devises a fault-tolerant scheme to compensate for the partial loss of actuator effectiveness in a spacecraft.

- A *locked-in-place* actuator is an extreme form of partial loss of effectiveness, as this actuator is stuck or jammed at a specific position and is then producing a constant output onward from a possibly unknown failure time [10], [46], [48]–[50]. A specific case of lock-in-place is *handover* or *runaway* when the actuator is stuck at the lower or upper end of its output range [17]. A typical example of this category of actuator failure would be a robot arm with a frozen joint [49]. To handle actuators locking in place, the work [48] relies on adaptive control to compensate uncertainties with adaptive tuning of controller parameters based on system response. If the output value of the locked-in-place actuator is known, one can evaluate the reachable space of the system to study its remaining capabilities as in [50].

- An actuator suffering from a *float* failure has no effect over the system performance [16], [46]. For instance, after a rupture of the control line, an aircraft elevon might be flapping in the wind instead of being actively controlled [16]. This elevon should then be removed from use by the control allocator. The works [51], [52] investigated float failures under the form of completely disabled spacecraft thrusters to guarantee the safety of orbital rendezvous.

Some works adopt a more generic malfunction model allowing to study several of the actuator failure categories together. For instance, the work [10] develops an adaptive control approach to stabilize hypersonic reentry vehicles facing a combination of both partial loss of effectiveness and locked-in-place actuators. The work [46] studies a combination of the three categories of actuator failures and establishes transient performance guarantees after actuator failures using direct adaptive control.

However, after a loss of control authority over an actuator, this malfunctioning actuator retains the capability to produce the same range of outputs as nominally but it does not follow the controller's commands anymore. This malfunction is then not covered by any of the three categories of actuator failures discussed above, hence preventing the use of these previous works. Another factor preventing the extension of existing fault-tolerant approaches to resilience theory is the specificity of the application of each work. Indeed, most of the works introduced above design fault-tolerant schemes specifically tailored for their application of interest like hypersonic reentry vehicles [10], supersonic fighter jet [16], hydraulic driven control surfaces [14], kinematically redundant manipulators [49], or steam generators [50]. The methods employed in these papers are usually too specific to be generalized to other applications with different dynamics.

We have now reviewed the three main control theories that most closely relate to the framework studied in this dissertation. Despite sharing some common traits with robust, adaptive and fault-tolerant control theories, resilience is better suited than these classical theories to investigate autonomous systems enduring a partial loss of control authority over their actuators. However, this dissertation is not the first work in the literature to study the resilience of control systems. We will then review previous approaches.

2.2 Resilience of control systems

In this section, we provide an overview of previous works who studied and quantified the resilience of control systems. We will investigate how these works compare to the framework of this dissertation.

Loss of control authority over actuators was first introduced in the work [53] which investigated how to guarantee safe operation of a control system in which some components are no longer under the controller's authority. However, it only studied extremely simple driftless discrete-time dynamics and its approach took full advantage of the discrete state and action spaces, hence preventing extensions to more complex continuous dynamics.

The work [6] studies control system under sensor attacks in a framework very similar to ours, except that all inputs are unbounded. Their goal is one of observability in reconstructing the state evolution based on corrupt sensor measurements. They prove that this reconstruction is only possible if less than half of the sensors are attacked. Since our problem of interest is more aligned with controllability than observability, we would want to transform their results through duality. However, because of the attacker's signal e bringing nonlinearity to the output y of the linear system

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad y(t) = Cx(t) + e(t),$$

the dual of this perturbed system is not clearly defined, what to do of $e(t)$? Additionally, the unboundedness of the inputs prevents to extend this theory to our framework. As noticed in the work [54], when bounding the initially unbounded inputs of an elegant control theory, the results are disappointing.

Instead of quantifying resilience with optimal reach times as in this dissertation, the work [55] chooses to quantify resilience with the minimal energy needed to reach a target. Work [55] actually extends the approach of the earlier work [15] from linear systems to nonlinear ones using the Koopman theory. An ideological difference between these works and our dissertation is their view of resilience as the capability of the system to bounce back after a malfunction. Then, the two factors contributing to resilience in [15], [55] are the capabilities of the controller to detect quickly all malfunction occurrences and to return the perturbed state to its nominal value. Hence, works [15], [55] quantify resilience as the observability and controllability of the system. Thanks to the observability and controllability Gramian matrices, these notions are much easier to quantify than optimal reach times as discussed in Sections 2.3.2 and 2.3.3. However, we believe our setting to be more meaningful as it is not restricted to inputs with bounded energy and reach times provide a more practical information than input energy.

Previous work have established notions of quantitative resilience for different fields of applications. In water infrastructure systems, the work [56] reviews and compares twenty-one previously established resilience metrics designed for water resource or distribution systems. A thorough comparison lead the authors of [56] to provide improvement guidelines for resilience metric specifically designed for water infrastructure systems. Because of their highly specific framework, these metrics do not apply to the generic control systems under study in this dissertation.

After the accident at the Fukushima nuclear power plant, nuclear safety guidelines all around the world were reviewed and new methods were established to assess and guarantee safety. The work [57] derives a quantitative resilience metric based on a statistical study of all the abnormal behaviors of nuclear power plants reported to the Korea Institute of Nuclear Safety. For each incident, resilience is based on an evaluation of factors describing the plant and its operating crew such as anticipation, adaptation, training, learning, decision making, and so on [57]. This framework is extremely specific to the problem of nuclear power plant safety and cannot be extended to general systems as this dissertation aims to do.

On the contrary to both previous works, [58] is too generic in its concept of resilience. Indeed, work [58] uses systems engineering tools to define resilience as the ratio ‘recovery over loss’ describing how much a system bounced back after a malfunction. This ratio is somewhat similar to the quantitative resilience metric introduced above Problem 3, but the desire of the authors of [58] to remain general prevents application to the control problems studied in this dissertation.

To the best of the author’s knowledge, we have reviewed all the previous works studying resilience of control systems and we concluded that the problems this dissertation set to solve cannot be addressed with previous approaches. We will now proceed to a review of different works pertaining to reachability analysis as this topic plays a central role in this dissertation.

2.3 Reachability analysis

In this section we review numerous works all related to reachability analysis, but belonging to different fields among which are robust, linear, and optimal control theories along with differential games theory and delayed systems. We start by discussing the problem of reachability of a controlled linear system.

2.3.1 Reachability

While computation of a reachable set is a classical problem in control theory [59], [60], it remains a computationally intensive problem usually addressed by under and over-approximations. The work [61] describes an algorithm to compute the reachable set of linear time-invariant systems requiring only Minkowski additions. The complexity of these additions is heavily dependent on the chosen set representation and [61] argues to select zonotopes for ease of calculation. Additionally, these zonotope additions are easily implemented on MATLAB with the dedicated reachability analysis toolbox [62]. However, zonotopes are not the only efficient set representation to perform reachability analysis. The work [63] instead focused on providing polyhedral approximations of the reachable set of a linear system, while [64] chose ellipsoidal representations. Research has also investigated how to obtain the optimal inputs responsible for the states on the boundary of the reachable set, using once again zonotopic representation [65].

In the resilience framework of this dissertation, we are also interested by the reachable set of the malfunctioning system (1.2), where the control u tries to steer the state to some target despite the worst undesirable input w trying to prevent this steering. In this setting, we cannot employ directly the reachability methods discussed above as they rely on full knowledge of system state and inputs. The most widely studied notion of reachability for systems with perturbations is *strong reachability* [24] which tackles the problem of how to reach a target set with a control input that works for all possible undesirable inputs. This approach relates to the field of robust control, and has been studied by, e.g., [41]–[43], [66]. The work [43] considers the worst case perturbation and try to obtain guaranteed performance for trajectory tracking by deriving conditions for a robust controller to maintain trajectory within a small distance of a target trajectory despite all admissible perturbation. Similarly, [42] establish ellipsoidal inner and outer approximations of robustly reachable sets obtained from the Hamilton-Jacobi-Bellman equation. On a similar note, the work [66] aims at computing the set of initial states from which some target is robustly reachable. Nonetheless, these robust methods are conservative and often produce meaningful results only when the amount of undesirable disturbances in the system is small [44]. Therefore, discussion of reachability in the face of loss of control authority, where the capabilities of the uncontrolled actuators may equal or exceed the capabilities of the remaining control actuators, calls for a different type of reachability to be discussed in Sections 2.3.3 and 2.3.4.

2.3.2 Time optimal linear control

To quantify the resilience of autonomous systems, we decided to compare the minimal time to reach a target for the nominal and malfunctioning dynamics. Then, the first step in addressing Problem 3 for linear dynamics is to calculate the minimal time for the nominal system to reach a target. This topic was of particular interest during the 1950's as discussed in the introduction of [67]. To be more specific, these works are concerned with calculating

$$T_N^*(x_0) := \inf_{u \in \mathcal{F}(\mathcal{U})} \big\{ T \geq 0 : x(T) = 0, \text{ with } x(0) = x_0, \text{ and } \dot{x}(t) = Ax(t) + Bu(t) \text{ for all } t \geq 0 \big\}, \tag{2.2}$$

and $\mathcal{U} = [-1, 1]^m$, with $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$ constant matrices. The work [67] establishes that T_N^* exists and that the optimal input u^* argument of the optimization (2.2) is unique and bang-bang for normal and Hurwitz linear systems. The notion of normality is related to the structure of the reachable set and directly implies uniqueness of the time-optimal inputs [68].

However, the problem of synthesizing u^* in a general setting was not solved until the work [69], which derives an iterative method converging to the optimal control. Shortly after, work [70] establishes a method of successive approximations of u^* designed for the engineers in need of a practical process. Instead of solving directly the minimal time problem (2.2), work [70] solves recursively the easier problem of minimizing the norm of the final state at a given time. The smallest of these times where the norm of the state is null is the minimal stabilizing time T_N^*. The same year [71] derives another algorithmic way of computing u^*, but based on geometrical intuition and hyperplanes.

Then, the work [72] extended the bang-bang principle of [67] to systems with nonlinear control but linear internal dynamics. More specifically, [72] established the compactness of the reachable set of dynamics $\dot{x}(t) = A(t)x(t) + \varphi(u, t)$ when $u(t) \in \mathcal{U}$ is compact, and A and φ are continuous. We will make good use of this result in Proposition 11.

In the following years, [73] derives another algorithm relying on gradient descent to compute u^* for semilinear systems of the form $\dot{x}(t) = A(t)x(t) + B(t)u(t) + f(t)$. Most of the results cited so far required the normality of the dynamics to ensure the uniqueness of the optimal solution. That is where the work [74] comes into play, since its algorithm to synthesize u^* does not need normality. The profusion of numerical methods to determine u^* and T_N^* is in fact due to the absence of a closed-form analytical description of these optimal quantities as stated in the review [29].

However, most of the algorithms derived at this period are not very efficient in terms of computation. Indeed, these gradient-based iterative methods are very sensitive to initial guess and usually exhibit poor convergence properties. To address this issue, the more recent work [75] studies the single input case where all the eigenvalues of A are real. In this case, work [75] determines the optimal sequence of switching times of the bang-bang input u^* and provides an algorithm with much better convergence properties than older approaches. This work is then extended to linear systems with complex poles in [76].

We should mention that contemporary research is still performed on the time optimal control of linear systems, as witnessed by the work [77] studying the case of time-optimal transfer between two non-zero states, which had not been solved previously according to the authors. To compute T_N^* and evaluate quantitative resilience in Chapter 7, we will rely on the numerical methods aforementioned. In order to prove the existence of certain optimal controls in several instances of our dissertation, we rely on the excellent optimization book [78]. However, calculating T_N^* is only half of the quantitative resilience ratio. We will now review previous work that studied the malfunctioning reach time T_M^* as introduced before Problem 3.

2.3.3 Time optimal differential games

Let us start by giving a proper definition of the malfunctioning reach time

$$T_M^*(x_0) := \sup_{w \,\in\, \mathcal{F}(\mathcal{W})} \Big\{ \inf_{u \,\in\, \mathcal{F}(\mathcal{U})} \big\{ T \geq 0 : x(T) = 0, \text{ with } x(0) = x_0, \text{ and } \dot{x}(t) = Ax(t) + Bu(t) + Cw(t) \big\} \Big\}, \quad (2.3)$$

where $\mathcal{U} = [-1, 1]^m$ and $\mathcal{W} = [-1, 1]^p$, with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{n \times p}$ constant matrices.

The first approach studying this minimax optimization is the work [79], which solves the minimal time intercept problem between a pursuer and evader of linear dynamics and single inputs. Additionally, work [79] concludes that if we know how to solve an optimal control problem, then we can solve the associated minimal time intercept problem. This statement was formally established as a duality theorem between optimal control and differential games by Hájek in his work [23]. This duality result is the bedrock for the resilience theory of linear systems established in Chapter 7 and used in subsequent chapters. Building on Hájek's work [23], his student continued the investigation of the equivalence between differential games and optimal controls by extending their approach to games with fixed winning times [80] and by proving the uniqueness of winning policies [68].

The duality theorem of [23] transforms a differential game with input sets $\mathcal{U}$ and $\mathcal{W}$ into an optimal control problem with a single input set defined as the Minkowski difference $B\mathcal{U} \ominus C\mathcal{W}$. This idea is independently discovered at the same time by [81], which develops an iterative process of set additions and differences in order to calculate the backward reachable set of a linear discrete time system with a controller u_k immediately aware of the disturbance w_k. This is almost exactly the framework of this dissertation. However, the reachability method developed in [81] is specific to discrete time dynamics and it does not have the theoretical generality of the duality theorems of [23], which is why do not employ the approach of [81].

To compute the Minkowski difference $B\mathcal{U} \ominus C\mathcal{W}$ mentioned in [23], [68], [80], [81], we will model input sets $B\mathcal{U}$ and $C\mathcal{W}$ as zonotopes and employ the methods of [62], [82]. Minkowski difference, also called Pontryagin difference [23], geometric set difference [81] or star-difference [68], [80] is studied in depth in the works [68], [83], [84].

The framework of [23] opened the door to prolific analytical results in our resilience theory, but it does not provide a way of computing T_M^*. Similarly, the work [79] can only be used for single input linear systems, which is too restrictive for applications. Following the intuition of [79], the solution of T_M^* is not found in time optimal control theory, but in the field of differential games.

The foundational theory of this field is attributed to Isaacs [22], [85] after remaining unpublished for ten years. Building on these early results of differential games, the work [30] provides an iterative process to solve (2.3). We will use this algorithm to calculate T_M^* in the numerical applications of Chapter 7. However, for the supremum and infimum of (2.3) to have solutions, the framework of [30] must assume that the controller knows ahead of time the optimal undesirable signal w^*, which contradicts the information setting of this dissertation. However, without this assumption, there would be no single obvious choice for $u^*\big(t, w(t)\big)$, rendering T_M^* ill-defined and certainly not time-optimal, whereas T_N^* is time-optimal. In this case, our concept of quantitative resilience T_N^*/T_M^* becomes meaningless as these two times are not comparable if only T_N^* is optimal.

The work [86] states that to calculate u^* without future knowledge of w^* the only technique is to solve the intractable Isaacs's equation. Thus, the paper [86] derives only suboptimal solutions and concludes that its practical contribution is minimal. Instead, we follow [30] where the inputs u^* and w^* are both chosen to make the transfer from x_0 to 0 time-optimal in the sense of (2.3). The controller knows that w^* will be chosen to make T_M^* the longest. Thus, u^* is chosen to react optimally to this worst undesirable input. Then, w^* is chosen, and to make T_M^* the longest, it is the same as the controller had predicted. Hence, from an outside perspective it looks as if the controller knew w^* in advance, as reflected by (2.3). This type of strategy in game theory is called Stakelberg optimal [87].

While not referring to it as resilient reachability, [88] studied this same framework but focused on developing a numerical method to compute reachable sets in a differential game setting by solving the Hamilton-Jacobi-Isaacs partial differential equation. Compared to the analytical treatment of resilient reachability proposed in this dissertation, the algorithm of [88] does not provide any insight in the reasons why a target set is resiliently reachable or not. Because such a knowledge is necessary to design resilient systems and solve Problem 2, we do not exploit the numerical approach of [88].

2.3.4 Max-min controllability

In parallel with the development of differential games, control theory also investigated the same problem, but under the name of perturbed reachability or max-min controllability. These works distinguish themselves from the robust control approaches in that the control input is allowed to depend on the disturbance, as in our resilience setting.

The first of these works is [24] which transformed the problem of resilient reachability into a minimax formula assessing whether a target set is reachable. This transformation required the use of topological dual spaces and the resulting reachability condition is highly abstract, lack intuition, and is difficult to compute, as we will notice in Chapter 3. Indeed, our approach to resilient reachability with inputs of bounded energy heavily draws from the work [24] and mostly consists in simplifying the reachability condition into a usable form.

Inspired by differential games, the work [89] studied a linear system with dual controls, which is essentially a two-player game. This system is called *max-min controllable* if after one player announces its control input for some time interval, the other player can devise an input to bring the state to the origin. The same authors further develop this approach in [90] and obtain elegant max-min controllability conditions very similar to the controllability condition of Kalman [91]. However, the unbounded inputs of [89], [90] prevent a wider use of these works. To generalize their approach, the work [54] bounded the admissible inputs and sadly showed that in this case the max-min controllability conditions become highly abstract and difficult to compute.

The works discussed here and in Section 2.3.3 assume that the controller has either immediate knowledge of his opponent's input following the 'snap decision rule' of [23], or the controller has advanced knowledge of the opponent's strategy as in [30], [89], [90]. While the differential games framework is particularly well adapted for resilience, the undesirable inputs generated by the malfunctioning actuators are not necessarily adversarial or rational and hence they cannot be anticipated in general. Additionally, even the immediate knowledge assumption is not realistic due to sensors and actuators delays, on top of the necessary computational time for the controller. We will then discuss the problem of reachability for systems with actuation delays.

2.3.5 Reachability of delayed systems

In real-world applications, sensors and actuators do not operate instantly. There is a delay between the time when the command signal is received by the actuator and the time when the commanded output is actually produced. This transient was mentioned in the Fault Detection and Isolation (FDI) module design of [16]. This transient is also studied for UAV propellers, as they do not respond immediately to commands as illustrated in the experimental work of [92]. Similarly, spacecraft thrusters operate with an actuation delay [26], [27].

On the other hand, the delay created by sensors prevents the controller from acquiring immediate knowledge of the malfunctioning input. If this delay is not taken into account in the construction of the controller, the latter might be destabilizing the system. For instance, consider a linear controller operating in phase opposition with the pendulum it is supposed to stabilize. Such a controller would maintain or amplify the oscillations of the pendulum. To study actuation delays, the extensive literature review [25] recommends the Artstein predictor [93]. This predictor estimates the state $x(t)$ based on the information available to the controller at time $t - \tau$, where $\tau > 0$ would be its actuation delay. With this state prediction, the controller can perform a more accurate state feedback than if it was using $x(t - \tau)$. For systems suffering from unknown disturbances, work [94] improved the Artstein predictor by taking into account past history of the disturbance. We will employ the predictor of [94] to design a resilient controller with actuation delay in Chapter 10.

We have now reviewed reachability literature from a variety of different fields to help us build our resilience theory.

2.4 Summary

In this chapter we proceeded to a literature review of the main fields connected with this dissertation. We first showed how our resilience theory differs from robust control, while being closer to the adaptive control ideology. Our theory fits within the wide umbrella of fault-tolerant theory, but studies a specific actuator malfunction requiring a distinct approach from existing fault-tolerant works. Then, we compared our theory with other works on resilience differing in their frameworks and in the type of problems they can solve. Finally, we delved into the reachability analysis literature and reviewed methods from a variety of fields among which are perturbed reachability analysis and differential games.

Chapter 3

Resilient Reachability for Linear Systems with Bounded Energy

3.1 Introduction

This chapter constitutes the foundational resilience theory for linear systems with bounded energy and relies on our work [31]. The objective of this chapter is to develop simple verification conditions determining whether after a loss of control authority over actuators a linear system is still able to reach its initial target. The contributions of this chapter are fourfold. First, we adopt the reachability condition of [24] and develop it into a usable equation describing resilient reachability for linear systems with bounded energy. Second, we tackle the specific case of driftless systems by deriving a computable condition for resilient reachability. Third, we analyze the evolution with time of resilient reachability for driftless systems, and show that the resilient reachability problem can be formulated as a minimax optimization of a concave-convex objective function. Fourth, we establish several sufficient conditions to avoid solving the aforementioned minimax optimization problem.

This chapter is organized as follows. Section 3.2 defines the problem of interest and states the related necessary definitions. Section 3.3 introduces preliminary results obtained by [24], upon which we build our theory. In Section 3.4 we develop a resilient reachability condition for linear systems. Section 3.5 applies this condition to driftless systems, while Section 3.6 explores how resilient reachability of a target set evolves with time and establishes a sufficient condition for resilient reachability. Section 3.7 illustrates our theory on two scenarios comprising a one-dimensional system and an underwater robot.

3.2 Problem statement

We consider a system governed by the dynamics

$$\dot{x}(t) = Ax(t) + \bar{B}\bar{u}(t), \qquad x(0) = x_0 \in \mathbb{R}^n,$$

where $A \in \mathbb{R}^{n\times n}$ and $\bar{B} \in \mathbb{R}^{n\times(m+p)}$ are constant matrices. Let $\mathcal{T} \subseteq \mathbb{R}^n$ be the target set to be reached by the system. Assume that, during its mission, the system loses authority over p of its $m+p$ actuators. We can then separate the controlled inputs $u \in \mathcal{F}(\mathcal{U})$ from the undesirable inputs $w \in \mathcal{F}(\mathcal{W})$, with $\mathcal{U} = -\mathcal{U} \subseteq \mathbb{R}^m$

the symmetric set of admissible control signals and $\mathcal{W} = -\mathcal{W} \subseteq \mathbb{R}^p$ the symmetric set of undesirable signals. Matrix $\bar{B}$ is split accordingly into $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{n\times p}$, representing respectively the controlled and uncontrolled actuators. The dynamics of this malfunctioning system can be written as follows

$$\dot{x}(t) = Ax(t) + Bu(t) + Cw(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \quad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W}. \tag{3.1}$$

The technical work of this chapter follows the assumptions of [24] and considers the admissible inputs to be square integrable signals over their time domain $[0, T]$, i.e.,

$$\mathcal{F}(\mathcal{U}) := \big\{u \in \mathcal{L}_2\big([0,T],\, \mathcal{U}\big) : \|u\|_{\mathcal{L}_2} \leq 1\big\}, \quad \text{and} \quad \mathcal{F}(\mathcal{W}) := \big\{w \in \mathcal{L}_2\big([0,T],\, \mathcal{W}\big) : \|w\|_{\mathcal{L}_2} \leq 1\big\}.$$

The target set is $\mathcal{T} := \big\{x \in \mathbb{R}^n : \|x - x_{goal}\| \leq \varepsilon\big\} = \mathbb{B}(x_{goal}, \varepsilon)$, where $x_{goal} \in \mathbb{R}^n$ and $\varepsilon \geq 0$. Our objective is then to find simple conditions characterizing whether target set $\mathcal{T}$ is reachable in a given time by system (3.1), regardless of the inputs w imposed by the malfunctioning actuators. We now define formally this notion of *resilient reachability.*

Definition 1: Target set $\mathcal{T}$ is *resiliently reachable at time* T by system (3.1) if for any undesirable input $w \in \mathcal{F}(\mathcal{W})$, there exists a control law $u_w \in \mathcal{F}(\mathcal{U})$ such that the state of system (3.1) verifies $x(T) \in \mathcal{T}$.

Definition 2: Target set $\mathcal{T}$ is *resiliently reachable by time* T by system (3.1) if $\mathcal{T}$ is resiliently reachable at some time $t \leq T$ by system (3.1).

We note the possible dependence of u_w on the undesirable input w. Unlike the concept of *strong reachability* [24] belonging to classical robust control [41], [66], our objective is not to a priori design a control signal that would bring the state to the target set for any undesirable inputs, but instead to guarantee that whatever the undesirable inputs are, one can determine a control signal *dependent on the undesirable inputs* to drive the system to its goal. The intuition behind posing such a problem is that the system inputs, even if not desirable, can often be measured. In turn, counteracting undesirable inputs is simpler when these inputs are known and a subsequent controller can thus handle perturbations of a larger magnitude than a standard robust controller. We can then formulate the associated problems of interest to be solved in this chapter.

Problem 5: Determine conditions under which $\mathcal{T}$ is resiliently reachable at time T by system (3.1).

Problem 6: Determine conditions under which $\mathcal{T}$ is resiliently reachable by time T by system (3.1).

This chapter focuses primarily on determining whether a target set is resiliently reachable for a particular initial state. The problem of determining a suitable control signal u_w as mentioned in Definition 1 is left to Chapter 4. We now proceed to describe prior results that enable our work.

3.3 Preliminaries

The main result of this section is a resilient reachability condition derived from [24]. This condition will serve as primary foundation to build our theory. The work [24] studied the abstract system

$$x = s + Su + Rw, \tag{3.2}$$

where $x \in X_3$ is the current state, $s \in X_3$ is the initial state, $u \in X_1$ is the control and $w \in X_2$ is the disturbance. Maps $S \in \mathcal{L}\big(X_1, X_3\big)$ and $R \in \mathcal{L}\big(X_2, X_3\big)$ represent respectively the impacts of the controlled and undesirable inputs on the state. For our case of interest, we consider $X_1 = \mathcal{F}(\mathcal{U})$, $X_2 = \mathcal{F}(\mathcal{W})$, $X_3 = \mathbb{R}^n$, and we define the following continuous linear operators:

$$S(u) := \int_0^T e^{A(T-\tau)} Bu(\tau) d\tau \ , \qquad R(w) := \int_0^T e^{A(T-\tau)} Cw(\tau) d\tau.$$

By taking $s := e^{AT} x_0 \in \mathbb{R}^n$, the solution of (3.1) can then be written as $x(T) = s + S(u) + R(w)$, which is exactly (3.2). We also need to define norms on dual spaces, and for that we rely on [95]. Considering a Banach space X and its adjoint X^*, the norm of $f \in X^*$ is defined by

$$\|f^*\|_{X^*} = \sup_{\|x\|_X = 1} \big\{ |f^*(x)| \big\} = \sup_{\|x\|_X \leq 1} \big\{ |f^*(x)| \big\}. \tag{3.3}$$

We can now state our first result, which will serve as the basis of the work in the next sections.

Proposition 1: Target set $\mathcal{T}$ is resiliently reachable at time T by system (3.1) if and only if

$$\sup_{\|x^*\|_{X_3^*} = 1} \Big\{ x^*(s - x_{goal}) - \|S^* x^*\|_{X_1^*} + \|R^* x^*\|_{X_2^*} - \varepsilon \Big\} \leq 0.$$

Proof. Let us start from Corollary 5.8 of [24], which, while not using the same terminology, states that $\mathcal{T}$ is resiliently reachable by system (3.1) at time T if and only if

$$\sup_{\|x^*\|_{X_3^*} = 1} \Big\{ x^*(s) + \inf_{u \in \mathcal{F}(\mathcal{U})} \big(S^* x^*(u) \big) + \sup_{w \in \mathcal{F}(\mathcal{W})} \big(R^* x^*(w) \big) - \sup_{y \in \mathcal{T}} \big(x^*(y) \big) \Big\} \leq 0. \tag{3.4}$$

Since $S^* x^*$ and $R^* x^*$ are linear, and sets $\mathcal{U}$ and $\mathcal{W}$ are symmetric, we obtain

$$\inf_{u \in \mathcal{U}} \big(S^* x^*(u) \big) = - \sup_{u \in U} \big(|S^* x^*(u)| \big) = -\|S^* x^*\|_{X_1^*}, \quad \text{and} \quad \sup_{w \in \mathcal{W}} \big(R^* x^*(w) \big) = \|R^* x^*\|_{X_2^*}. \tag{3.5}$$

For $y \in \mathcal{T} = \mathbb{B}(x_{goal}, \varepsilon)$, we can write $y = x_{goal} + \delta y$ with $\delta y \in \mathbb{B}(0, \varepsilon)$. Then, $\sup\limits_{\|\delta y\| \leq \varepsilon} \big(x^*(\delta y) \big) = \varepsilon \sup\limits_{\|\delta y\| \leq 1} \big(x^*(\delta y) \big) = \varepsilon \|x^*\|_{X_3^*}$ by linearity. Recalling that $\|x^*\|_{X_3^*} = 1$ in (3.4), we obtain $\sup\limits_{\|\delta y\| \leq \varepsilon} \big(x^*(\delta y) \big) = \varepsilon$. Since x^* is linear, $\sup\limits_{y \in \mathcal{T}} \big(x^*(y) \big) = x^*(x_{goal}) + \sup\limits_{\|\delta y\| \leq \varepsilon} \big(x^*(\delta y) \big) = x^*(x_{goal}) + \varepsilon$. We conclude the proof by plugging in (3.5) into (3.4). □

The reachability condition derived in Proposition 1 is highly abstract due to the dual terms and is impractical to use for devising solutions of our two problems of interest. The following two sections aim to develop more workable conditions.

3.4 Integral resilient reachability condition

We will now work on the simplification of Proposition 1. First, we can use the Riesz representation theorem [95] to simplify x^*. Indeed, $x^* \in \mathcal{L}\big(\mathbb{R}^n, \mathbb{R}\big)$ is bounded in Proposition 1, because $\|x^*\|_{X_3^*} = 1$, so there exists a unique $h \in \mathbb{R}^n$ such that $x^*(\cdot) = \langle h, \cdot \rangle$ and $\|h\| = \|x^*\|_{X_3^*} = 1$. Thus, the supremum in Proposition 1 is over the unit sphere in $\mathbb{R}^n$, i.e., for $h \in \mathbb{S}$. With $s = e^{AT} x_0$, the first term of the reachability condition from

Proposition 1 can be rewritten as

$$x^*(s - x_{goal}) = \langle h, e^{AT}x_0 - x_{goal}\rangle. \tag{3.6}$$

Now we can simplify the adjoint maps S^* and R^* that appear in Proposition 1. Since $S : \mathcal{F}(\mathcal{U}) \longrightarrow \mathbb{R}^n$, its adjoint map is by definition [95] $S^* : \big(\mathbb{R}^n\big)^* \longrightarrow \big(\mathcal{F}(\mathcal{U})\big)^*$ with $S^*(x^*) := x^* \circ S$ and the associated commutative diagram representing is illustrated in Fig. 3.1.

$$\begin{tikzcd}
 & \mathbb{R}^n \arrow[dr, "x^*"] & \\
\mathcal{F}(\mathcal{U}) \arrow[ur, "S"] \arrow[rr, "S^*x^* \in \mathcal{L}\big(\mathcal{F}(\mathcal{U}), \mathbb{R}\big)"'] & & \mathbb{R}
\end{tikzcd}$$

Figure 3.1: Commutative diagram of map S and its adjoint S^*.

Then, for any $u \in \mathcal{F}(\mathcal{U})$ we have

$$S^*x^*(u) = \big(x^* \circ S\big)(u) = x^*\big(S(u)\big) = \big\langle h, S(u)\big\rangle = \Big\langle h, \int_0^T e^{A(T-\tau)}Bu(\tau)d\tau\Big\rangle.$$

By (3.3), the norm on $\big(\mathcal{F}(\mathcal{U})\big)^*$ is defined by

$$\|f\|_{\left(\mathcal{F}(\mathcal{U})\right)^*} := \sup_{\|u\|_{\mathcal{F}(\mathcal{U})} = 1} \big\{|f(u)|\big\}, \qquad \text{with} \qquad \|u\|_{\mathcal{F}(\mathcal{U})} = \|u\|_{\mathcal{L}_2} = \sqrt{\int_0^T \|u(\tau)\|^2 d\tau}.$$

Then, we obtain

$$\|S^*x^*\|_{\mathcal{L}_2^*} = \sup_{\|u\|_{\mathcal{L}_2}=1}\left\{\left|\Big\langle h, \int_0^T e^{A(T-\tau)}Bu(\tau)d\tau\Big\rangle\right|\right\}, \quad \text{and} \quad \|R^*x^*\|_{\mathcal{L}_2^*} = \sup_{\|w\|_{\mathcal{L}_2}=1}\left\{\left|\Big\langle h, \int_0^T e^{A(T-\tau)}Cw(\tau)d\tau\Big\rangle\right|\right\}.$$

We can then simplify the resilient reachability condition of Proposition 1.

Theorem 1: Target set $\mathcal{T}$ is resiliently reachable at time T by system (3.1) if and only if

$$\max_{h\,\in\,\mathbb{S}}\left\{\langle h, e^{AT}x_0 - x_{goal}\rangle - \sup_{\|u\|_{\mathcal{L}_2}=1}\left\{\left|\Big\langle h, \int_0^T e^{A(T-\tau)}Bu(\tau)d\tau\Big\rangle\right|\right\} + \sup_{\|w\|_{\mathcal{L}_2}=1}\left\{\left|\Big\langle h, \int_0^T e^{A(T-\tau)}Cw(\tau)d\tau\Big\rangle\right|\right\}\right\} \leq \varepsilon. \tag{3.7}$$

Proof. After replacing $\|S^*x^*\|_{\mathcal{L}_2^*}$ and $\|R^*x^*\|_{\mathcal{L}_2^*}$ in Proposition 1, the only work left is to prove that the supremum from Proposition 1 turns into $\max_{h\in\mathbb{S}}$, which follows from the discussion preceding (3.6), $\mathbb{S}$ being closed, and the function to maximize being continuous in h. □

Because it directly uses matrices A, B and C instead of the adjoints of maps derived from those matrices, the condition from Theorem 1 is more direct than the condition (3.4) we started from. Yet, computing the two supremums on the unit sphere of $\mathcal{L}_2$ is a difficult task because of the infinite dimension of $\mathcal{L}_2$. We now focus on driftless systems where the integrals in (3.7) can be simplified.

3.5 Driftless systems

Driftless systems are widely studied in robotics, as their dynamics represent the kinematics constraints of the system: numerous examples are described in [96]. For these systems matrix A equals 0, so that (3.1) becomes

$$\dot{x}(t) = Bu(t) + Cw(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \quad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W}. \tag{3.8}$$

The removal of A enables us to distill Theorem 1 into a simpler resilient reachability condition.

Theorem 2: Target set $\mathcal{T}$ is resiliently reachable at time T by system (3.8) if and only if

$$\max_{h \in \mathbb{S}} \Big\{ \langle h, x_0 - x_{goal} \rangle - \sqrt{T}\, \big\| B^\top h \big\| + \sqrt{T}\, \big\| C^\top h \big\| \Big\} \leq \varepsilon.$$

Proof. When $A = 0$, the leftmost term in (3.7) becomes $\langle h, x_0 - x_{goal} \rangle$. We now simplify the next term by using the Cauchy-Schwarz inequality:

$$\left| \left\langle h, \int_0^T e^{A(T-\tau)} Bu(\tau) d\tau \right\rangle \right| = \left| \left\langle h, B \int_0^T u(\tau) d\tau \right\rangle \right| = \left| \left\langle B^\top h, \int_0^T u(\tau) d\tau \right\rangle \right| \leq \big\| B^\top h \big\| \left\| \int_0^T u(\tau) d\tau \right\|. \tag{3.9}$$

The equality case happens when $B^\top h$ and $\int_0^T u(\tau) d\tau$ are *positively collinear*, i.e., when $\int_0^T u(\tau) d\tau$ is a nonnegative scalar multiple of $B^\top h$ [95].

If $(e_1, ..., e_m)$ is the canonical basis of $\mathbb{R}^m$, there exist $u_1, ..., u_m \in \mathcal{L}_2([0, T], \mathbb{R})$ such that $u = \sum_{i=1}^m u_i e_i$. The norm of the integral of u can then be simplified with the Cauchy-Schwarz inequality:

$$\left\| \int_0^T u(\tau) d\tau \right\| = \sqrt{\sum_{i=1}^m \left(\int_0^T u_i(\tau) \times 1 \; d\tau \right)^2} \leq \sqrt{\sum_{i=1}^m \left(\int_0^T u_i^2(\tau) d\tau \right) \left(\int_0^T 1^2 d\tau \right)}$$
$$\leq \sqrt{T} \sqrt{\int_0^T \sum_{i=1}^m u_i^2(\tau) d\tau} = \sqrt{T} \|u\|_{\mathcal{L}_2}.$$

The equality case happens when each u_i is almost everywhere (in the measure-theoretical sense) collinear with the function $\tau \mapsto 1$, i.e., when u is almost everywhere constant. Then,

$$\sup_{\|u\|_{\mathcal{L}_2} = 1} \left\{ \left| \left\langle h, \int_0^T Bu(\tau) d\tau \right\rangle \right| \right\} \leq \|B^\top h\| \sqrt{T}. \tag{3.10}$$

If we can find a function u_h of unit norm in $\mathcal{L}_2$ for which the inequality in (3.10) is an equality, then the supremum in (3.10) would be a maximum. The function u_h must realize both equality cases of the Cauchy-Schwarz inequality used previously. Hence, for $h \in \mathbb{S}$ we define the following constant function: $u_h(t) := \frac{B^\top h}{\sqrt{T} \|B^\top h\|}$, with $\|u_h(t)\| = \frac{1}{\sqrt{T}}$ for all $t \in [0, T]$. Thus, u_h is of unit norm on $\mathcal{L}_2$:

$$\|u_h\|_{\mathcal{L}_2} = \sqrt{\int_0^T \|u_h(t)\|^2 dt} = \sqrt{\int_0^T \frac{1}{T} dt} = 1.$$

Moreover, u_h is positively collinear with $B^\top h$ and is constant over time, therefore it satisfies both of the

Cauchy-Schwarz equality cases aforementioned, which leads to

$$\left| \left\langle h, \int_0^T Bu_h(\tau)d\tau \right\rangle \right| = \|B^\top h\| \left\| \int_0^T u_h(\tau)d\tau \right\| = \|B^\top h\| \sqrt{T}. \tag{3.11}$$

From (3.10) and (3.11), we clearly obtain

$$\max_{\|u\|_{\mathcal{L}_2}=1} \left\{ \left| \left\langle h, \int_0^T Bu(\tau)d\tau \right\rangle \right| \right\} = \|B^\top h\| \sqrt{T}, \tag{3.12}$$

The same process can be applied to the final term in (3.7), yielding the theorem claim. □

We remark on the precise meaning of reachability as we defined it: for all undesirable inputs, one can find a control law such that the goal is reachable. Thus, when Theorem 2 predicts an unreachable goal, it does not mean that the system can never attain $\mathcal{T}$, but that reaching the goal is not guaranteed for *any* undesirable input.

The reader desiring intuition on Theorem 2 should recall that $\mathbb{S}$ is the unit sphere in $\mathbb{R}^n$, so the maximum over $\mathbb{S}$ explores every direction for h. The scalar product $\langle h, x_0 - x_{goal} \rangle$ gives the intuition that h represents a direction of the system's travel in the state space. The h maximizing this scalar product is positively collinear with $x_0 - x_{goal}$, so it is driving the system away from x_{goal}. On the other hand, the terms $B^\top h$ and $C^\top h$ represent how the controls and the undesirable inputs drive the system when they are both along the direction h. Hence, the h that maximizes $\|C^\top h\| - \|B^\top h\|$ is the direction giving the most strength to the undesirable inputs over the controls. Therefore, the h that realizes the overall maximum represents the worst direction for resilient reachability.

We can strengthen our faith in Theorem 2 by looking at a few special cases. If we assume that $x_0 = x_{goal}$, then $\mathcal{T}$ is reachable at time $T = 0$ as, for all $h \in \mathbb{S}$, $\langle h, x_0 - x_{goal} \rangle = 0$. Another simple case is when $B = 0$ and $C = 0$. In this case, $\dot{x}(t) = 0$, so for all $t \geq 0$, $x(t) = x_0$ and the reachability condition becomes as expected $\|x_0 - x_{goal}\| \leq \varepsilon$, i.e., equivalent to $x_0 \in \mathcal{T}$.

With Theorems 1 and 2 we have solved Problem 5. We now have all the tools to start working on Problem 6, and study how the resilient reachability of target set $\mathcal{T}$ evolves with time.

3.6 Evolution of reachability with time

To simplify the notation of Theorem 2, let us first write $d := x_0 - x_{goal}$ and define functions

$$g(h) := \|C^\top h\| - \|B^\top h\|, \quad J(h,t) := \langle h, d \rangle + g(h)\sqrt{t}, \quad \text{and} \quad f(t) := \max_{h \in \mathbb{S}} \big\{ J(h,t) \big\}.$$

Thus, the condition of Theorem 2 is equivalent to $f(T) \leq \varepsilon$.

For a given d, the inner product $\langle h, d \rangle$ in J is bounded for $h \in \mathbb{S}$, so, as time T becomes sufficiently large, resilient reachability of $\mathcal{T}$ largely depends on the sign of the coefficient of $\sqrt{T}$, namely $g(h)$. If there are controls ($B \neq 0$), but no undesirable inputs ($C = 0$), the coefficient is $g(h) = -\|B^\top h\| < 0$, so J decreases with time, which intuitively means that the reachable set grows. The opposite also happens as expected when $B = 0$ and $C \neq 0$. We now wish to analyze the resilient reachability of target set $\mathcal{T}$ over time for general driftless systems.

Note first that for $t > 0$, $J(\cdot, t)$ is not a concave function, and thus its maximization over $\mathbb{S}$ is not an easy task. Indeed, both functions $h \mapsto \|C^\top h\|$ and $h \mapsto \|B^\top h\|$ are convex, so $g(h)$ is the difference between two convex functions. This type of maximization is referred to as a *difference of convex* (DC) problem, and analytical solutions are only available for a few special cases. Numerous algorithms have been developed to solve DC problems, like for instance [97], [98] and [99]. While these numerical results, combined with Theorem 2, enable us to determine whether set $\mathcal{T}$ is resiliently reachable *at* every given time, they do not provide any insight regarding reachability *by* a certain time. Therefore, we will not attempt to solve directly our DC maximization problem.

Following Theorem 2, target set $\mathcal{T}$ is resiliently reachable by time T by system (3.8) if and only if

$$\min_{t \,\in\, [0,T]} \Big\{ \max_{h \,\in\, \mathbb{S}} \big\{ J(h,t) \big\} \Big\} \leq \varepsilon.$$

Hence the reachability by time T can be described as a minimax problem with a DC cost function. We will omit the discussion of possible numerical solutions to such a problem and instead focus on analytical results. As noticed above, as time grows, $\sqrt{t}$ becomes the leading term in J, with its sign determined by $g(h)$. We therefore study the sign of $\max\limits_{h \,\in\, \mathbb{S}} \big\{ g(h) \big\}$. We will show the following:

- if $\max\limits_{h \,\in\, \mathbb{S}} \big\{ g(h) \big\} > 0$, target set $\mathcal{T}$ is only resiliently reachable up to a certain time,
- if $\max\limits_{h \,\in\, \mathbb{S}} \big\{ g(h) \big\} = 0$, target set $\mathcal{T}$ can be either always resiliently reachable, never resiliently reachable, or its resilient reachability depends on time,
- if $\max\limits_{h \,\in\, \mathbb{S}} \big\{ g(h) \big\} < 0$, target set $\mathcal{T}$ is resiliently reachable from some time onwards.

We prove these claims in the following three subsections.

3.6.1 Maximum of g is positive

When $\max \big\{ g(h) \big\} > 0$, there exists some $h \in \mathbb{S}$ such that $\|C^\top h\| \; > \; \|B^\top h\|$, i.e., in line with our intuition, there is an input direction where the matrix C produces a stronger undesirable input than what the control matrix B is capable of counteracting. Since we want to guarantee reaching the goal for *any* undesirable input, a single direction where the undesirable inputs are stronger than the controlled ones is sufficient to prevent resilient reachability. We formalize this intuition as follows.

Theorem 3: If $\max\limits_{h \,\in\, \mathbb{S}} \big\{ g(h) \big\} > 0$, then there exists $t_{lim} > 0$ such that target set $\mathcal{T}$ is not resiliently reachable at any time $t \geq t_{lim}$ by system (3.8).

Proof. Because $\max\limits_{h \,\in\, \mathbb{S}} \big\{ g(h) \big\} > 0$, there exists $h_+ \in \mathbb{S}$ such that $g(h_+) > 0$. Then,

$$f(t) \geq \langle h_+, d \rangle + g(h_+)\sqrt{t} \xrightarrow[t \to \infty]{} +\infty.$$

Thus, $\lim\limits_{t \to \infty} f(t) = +\infty$. Then, there exists $t_{lim} > 0$ such that for all $t \geq t_{lim}$, $f(t) > \varepsilon$, i.e., $\mathcal{T}$ is not reachable at any time $t \geq t_{lim}$ by system (3.8). □

Theorem 3 states that, all resilient reachability can only happen in finite time. It also means that the state cannot be maintained forever in set $\mathcal{T}$ for some undesirable inputs w. This result is interesting from a safety point of view. Assume that $\mathcal{T}$ is the safe set, i.e. the set where states meet all the security conditions enacted for this system. Then, even if the state x_0 is initially safe, i.e., $x_0 \in \mathcal{T}$, after some time t_{lim}, the state $x(t_{lim})$ might have been pushed out of the safe set $\mathcal{T}$ no matter what admissible control input u has been applied. Having a positive maximum of g is then detrimental for long-term safety.

3.6.2 Maximum of g equals zero

When $\max\big\{g(h)\big\} = 0$, there is at least one $h \in \mathbb{S}$ such that $g(h) = 0$. Intuitively, in this direction h the undesirable inputs match the controls. But, in directions where g is negative, the controls have a greater magnitude than the undesirable inputs. Thus, overall the controls can at least compensate the effects of the undesirable inputs.

Following this intuition, if the state initially belong to the target set, $x_0 \in \mathcal{T}$, it can be maintained within it, $x(t) \in \mathcal{T}$ for all $t \geq 0$. Then, the state cannot be forced to exit the safe set by undesirable inputs. This is the desired comportment for a safety region.

Since undesirable inputs can match the controls in certain directions, the resiliently reachable region does not expand in every direction with time. Thus, the resilient reachability of $\mathcal{T}$ depends on its location. Let us define $H_0 := \{h \in \mathbb{S} : g(h) = 0\}$. Set H_0 is closed, bounded, and nonempty by assumption. Hence, with $d = x_0 - x_{goal}$, we can define $h_0 := \arg \max\limits_{h \in H_0} \big\{h^\top d\big\}$. We note that vector h_0 need not be uniquely defined. The theorem below holds for every h_0.

Theorem 4: Assume $\max\limits_{h \in \mathbb{S}}\big\{g(h)\big\} = 0$. If $\|d\| \leq \varepsilon$, then target set $\mathcal{T}$ is resiliently reachable at all times $t \geq 0$ by system (3.8). On the other hand, if $\varepsilon < h_0^\top d$, then $\mathcal{T}$ is never resiliently reachable by system (3.8).

Proof. We note that $\max\limits_{h \in \mathbb{S}}\big\{h^\top d\big\} = f(0) = \|d\|$. Thus,

$$f(t) \leq \max_{h \in \mathbb{S}}\big\{h^\top d\big\} + \max_{h \in \mathbb{S}}\big\{g(h)\sqrt{t}\big\} = \|d\| + 0 = \|d\|.$$

Hence, $\max\limits_{t \geq 0}\big\{f(t)\big\} = \|d\|$. Additionally, $h_0 \in \mathbb{S}$, so $f(t) \geq h_0^\top d + g(h_0)\sqrt{t} = h_0^\top d$. Thus, $h_0^\top d \leq f(t) \leq \|d\|$ for all $t \geq 0$.

If $\|d\| \leq \varepsilon$, then for $t \geq 0$, $f(t) \leq \varepsilon$, i.e., by Theorem 2, target set $\mathcal{T}$ is resiliently reachable at all times $t \geq 0$ by system (3.8).

On the other hand, if $\varepsilon < h_0^\top d$, then for $t \geq 0$, $f(t) > \varepsilon$, i.e., by Theorem 2, target set $\mathcal{T}$ is never resiliently reachable by system (3.8). □

There is obviously an intermediate case to Theorem 4, where $\varepsilon \in \big[h_0^\top d,\ \|d\|\big]$ and the resilient reachability of $\mathcal{T}$ depends on time.

3.6.3 Maximum of g is negative

We can now tackle the third case, where $\max\big\{g(h)\big\} < 0$. In this situation, our intuition stipulates that controls are stronger than the undesirable inputs in every direction, so the reachable set grows unbounded with time. The theorem below confirms this intuition. And obviously the state can be maintained within any set already reached. This is ideal from a safety point of view, because the state can be constrained to any safe set.

Theorem 5: If $\max\limits_{h \in \mathbb{S}}\big\{g(h)\big\} < 0$, then there exists $t_{lim} \geq 0$ such that target set $\mathcal{T}$ is resiliently reachable at all times $t \geq t_{lim}$ by system (3.8).

Proof. We have

$$f(t) = \max_{h \in \mathbb{S}}\big\{h^\top d + g(h)\sqrt{t}\big\} \leq \max_{h \in \mathbb{S}}\big\{h^\top d\big\} + \max_{h \in \mathbb{S}}\big\{g(h)\big\}\sqrt{t} = \|d\| + \max_{h \in \mathbb{S}}\big\{g(h)\big\}\sqrt{t}.$$

We compare this upper bound with ε to obtain a reachability condition: $\|d\| + \max\limits_{h \in \mathbb{S}}\big\{g(h)\big\}\sqrt{t} \leq \varepsilon$ is equivalent to $t \geq \left(\frac{\|d\|-\varepsilon}{\gamma}\right)^2 := t_{lim}$. Thus, for all $t \geq t_{lim}$, $f(t) \leq \varepsilon$, i.e., target set $\mathcal{T}$ is resiliently reachable by system (3.8). □

The t_{lim} defined in Theorem 5 might not be the minimal time for resilient reachability since a first inequality has been used in order to decouple $h^\top d$ and $g(h)\sqrt{t}$. Nonetheless, Theorem 5 proves that, after some time, any target set becomes resiliently reachable.

Theorems 3, 4 and 5 show that the sign of the maximum of g leads to interesting conclusions. It is thus natural to attempt to analytically determine an upper bound for g.

3.6.4 Bounding g

Let $\sigma_{max}^{C^\top}$ be the maximal singular value of $C^\top$, and $\sigma_{min}^{B^\top}$ be the minimal singular value of $B^\top$. We claim that the relationship between these two values impacts the maximal value of g.

Theorem 6: If $\sigma_{max}^{C^\top} < \sigma_{min}^{B^\top}$, then $\max\limits_{h \in \mathbb{S}}\big\{g(h)\big\} < 0$.

Proof. Let us define $M = CC^\top \succcurlyeq 0$. Matrix M is symmetric, and we can use the following classical inequality [100]: $\lambda_{min}^M \|x\|^2 \leq x^\top M x \leq \lambda_{max}^M \|x\|^2$, for all $x \in \mathbb{R}^n$, with λ_{min}^M and λ_{max}^M respectively, the minimum and maximum eigenvalues of M. Since M is trivially positive semi-definite, $\lambda_{min}^M \geq 0$. Note that $\|C^\top h\| = \sqrt{h^\top CC^\top h} = \sqrt{h^\top M h}$. Thus we obtain $\sqrt{\lambda_{min}^M} \leq \|C^\top h\| \leq \sqrt{\lambda_{max}^M} = \sigma_{max}^{C^\top}$, for all $h \in \mathbb{S}$. By doing the same for $B^\top$, g can be bounded as $\sqrt{\lambda_{min}^{C^\top}} - \sqrt{\lambda_{max}^{B^\top}} \leq g(h) \leq \sigma_{max}^{C^\top} - \sigma_{min}^{B^\top}$, for all $h \in \mathbb{S}$. Hence, $\sigma_{max}^{C^\top} < \sigma_{min}^{B^\top}$ yields $\max\limits_{h \in \mathbb{S}}\big\{g(h)\big\} < 0$. □

Theorems 5 and 6 trivially imply the following corollary.

Corollary 1: If all singular values of $C^\top$ are strictly smaller than those of $B^\top$, then target set $\mathcal{T}$ is resiliently reachable in finite time by system (3.8).

We now illustrate the above theoretical results on two numerical examples.

3.7 Numerical examples

3.7.1 1D system

To illustrate Theorem 1, we consider the following simplistic one-dimensional system:

$$\dot{x}(t) = x(t) + bu(t) + cw(t), \qquad x_0 = 0,$$

with $b \geq 0$, $c \geq 0$, $x_{goal} = 1$, and $\mathcal{F}(\mathcal{U}) = \mathcal{F}(\mathcal{W}) = \big\{v \in \mathcal{L}_2\big([0,T],\ \mathbb{R}\big) : \|v\|_{\mathcal{L}_2} \leq 1\big\}$. We will calculate the resilient reachability condition given by Theorem 1, and verify whether the goal can actually be reached in that case.

The condition in Theorem 1 can be simplified since the inner product on $\mathbb{R}$ is just a scalar multiplication:

$$\left|\left\langle h, \int_0^T e^{T-t} b u(t) dt \right\rangle\right| = \underbrace{|h|}_{=1} \cdot b e^T \left|\int_0^T e^{-t} u(t) dt\right| \leq b e^T \sqrt{\int_0^T e^{-2t} dt} \sqrt{\int_0^T u^2 dt} = b e^T \sqrt{\frac{1 - e^{-2T}}{2}} \times \|u\|_{\mathcal{L}_2}.$$

We used the Cauchy-Schwarz inequality above, and note that the equality case is realized for $u(t) = e^{-t}$. We proceed the same way as we did to transform (3.10) into (3.12) and obtain

$$\sup_{\|u\|_{\mathcal{L}_2}=1} \left\{ \left|\left\langle h, \int_0^T e^{T-t} b u(t) dt \right\rangle\right| \right\} = b \sqrt{\frac{e^{2T} - 1}{2}}.$$

The last term left to calculate in Theorem 1 is the scalar product from (3.6): $\langle h, e^\top x_0 - x_{goal} \rangle = -h$. Theorem 1 can then be simplified to

$$1 + \sqrt{\frac{e^{2T} - 1}{2}}(c - b) \leq \varepsilon. \tag{3.13}$$

Defining $m(\varepsilon, T) = (1 - \varepsilon)\sqrt{2}/(\sqrt{e^{2T} - 1})$, (3.13) is equivalent to

$$c + m(\varepsilon, T) \leq b. \tag{3.14}$$

Intuitively, if (3.14) holds, then the control magnitude b is larger than the magnitude c of undesirable inputs, plus the minimum margin $m(\varepsilon, T)$ required to reach the target set at time T. The more time is allowed, the less margin is necessary: $m(\varepsilon, T)$ decreases with T.

We can now test condition (3.14), i.e., Theorem 1, by taking $w = \frac{-1}{\sqrt{T}}$ and $u = \frac{1}{\sqrt{T}}$. These inputs were chosen to be of unit norm in $\mathcal{L}_2$, with w pulling the state away from its goal and u counteracting w. We can then calculate the final state:

$$x(T) = \int_0^T e^{T-t} \left(b \frac{1}{\sqrt{T}} + c \frac{-1}{\sqrt{T}} \right) dt = (b - c) \frac{(e^\top - 1)}{\sqrt{T}}.$$

The closest point of the target set from the initial state is $x_{goal} - \varepsilon = 1 - \varepsilon$. Thus, target set $\mathcal{T}$ is reached at T *for the particular control input* u *and the particular undesirable signal* w, if $x(T) \geq 1 - \varepsilon$, i.e., $b - c \geq (1 - \varepsilon)\sqrt{T}/(e^\top - 1)$. Since $\frac{(1-\varepsilon)\sqrt{T}}{e^\top - 1} \geq m(\varepsilon, T)$ for all $T \geq 0$, the above result validates Theorem 1; if $\mathcal{T}$ is reached above (in which case $\mathcal{T}$ is certainly resiliently reachable), (3.14) will hold.

We now proceed to illustrate the developed theory for driftless systems.

3.7.2 A driftless underwater vehicle

We consider an underwater robot propelled by a main engine and two side engines for operations in a 2D plane, as shown in Fig. 3.2. We consider the scenario where the main engine u_1 has a small bias in the y direction. The system dynamics are thus given by

$$\dot{X} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 10 & 1 & 1 \\ 0.2 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

Our example is motivated by the works [101], [102], which have also considered underwater driftless

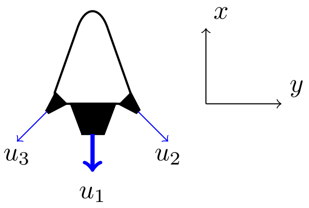


Figure 3.2: A model of an underwater robot with three actuators.

dynamics. The assumption of driftlessness can intuitively be justified by the viscosity of the water combined with a small speed of the robot.

We assume that during its mission the controller loses authority over the third engine. The terms in (3.8) can thus be written as follows:

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad w = u_3, \quad B = \begin{bmatrix} 10 & 1 \\ 0.2 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Intuitively, the robot should still be able to reach any goal set, since the second actuator u_2 can counteract the undesirable inputs of u_3, and the small bias of u_1 on y provides a net motion on y, while the desired displacement along x is also realized by the main engine. While the conditions of Theorem 6 are not satisfied, we can compute $\max\limits_{h \in \mathbb{S}} \big\{ g(h) \big\} = -0.02$, and use Theorem 5 to show that any goal is eventually resiliently reachable, as suggested by our intuition.

In the situation where the controller loses authority over both the second and third actuators, our intuition suggests that a controlled motion along x is still possible, but the displacements along y cannot be controlled. Therefore, we cannot guarantee to reach any target position. We numerically compute g and find that $\max\limits_{h \in \mathbb{S}} \big\{ g(h) \big\} = 1.4 > 0$. The conclusion of Theorem 3 validates our intuition.

If the controller only loses authority over the first actuator, then $\max\limits_{h \in \mathbb{S}} \big\{ g(h) \big\} = 8.6 > 0$. Of course none of the side engines can make up for the loss of the main one, as predicted by Theorem 3.

Another interesting case to note is when u_1 thrusts only along x without bias on y, i.e.,

$$\dot{X} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 10 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

Then, a loss of control authority over one of the side engines results in $\max\limits_{h \in \mathbb{S}} \big\{ g(h) \big\} = 0.02 > 0$. Indeed, we cannot guarantee to reach a goal that is not on the x axis, because no net motion on y is guaranteed, since both side engines can cancel each other out.

3.8 Summary

This chapter described and tackled the problem of resilient reachability for systems with bounded energy. We established analytical conditions under which a system can always be driven to a desired target, given that some of its actuators act in an undesirable manner and without prior knowledge of these undesirable inputs. To solve this problem, we derived a resilient reachability condition for linear systems and a more specific condition for driftless systems. We investigated the evolution of resilient reachability with time and rewrote the problem as a minimax optimization with a concave-convex objective function. We then derived results that do not require directly solving the optimization problem, at the price of providing sufficient or necessary conditions.

Chapter 4

Designing Resilient Linear Systems with Bounded Energy

4.1 Introduction

This chapter continues the resilience investigation of linear systems with bounded energy and relies on our work [35]. Building on Chapter 3, we define resilience as the capability to resiliently reach any state and we study how to design linear systems with bounded energy that are resilient to the loss of control authority over any one of their actuators. Chapter 3 showed that redundant actuators are necessary for resilient reachability and hence we aim at guaranteeing resilience with a minimal redundancy. The contributions of this chapter are twofold. First, we determine the minimal degree of overactuation necessary to design a resilient system. Second, we synthesize a control law driving a resilient system's state to its target despite loss of control authority over some actuators. To establish these results, we first focus on driftless systems, a common application in robotics [96], before extending our findings to systems with drift.

This chapter is organized as follows. Section 4.2 defines the problems of interest and introduces preliminary results. In Section 4.3, we develop the notion of resilient control matrices and we determine their minimal size in Section 4.4. Building on the driftless case, Section 4.5 focuses on the synthesis of a resilient control law for linear systems with and without drift. We illustrate our theory in Section 4.6 with three scenarios of a fighter jet undergoing a loss of control authority.

4.2 Problem statement and preliminaries

We use the same setting as in Chapter 3 and hence we consider a system initially governed by the differential equation

$$\dot{x}(t) = Ax(t) + \bar{B}\bar{u}(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \tag{4.1}$$

where $A \in \mathbb{R}^{n\times n}$ and $\bar{B} \in \mathbb{R}^{n\times m}$ are constant matrices. Let $\mathcal{T} := \mathbb{B}(x_{goal}, \varepsilon) \subseteq \mathbb{R}^n$ be the *target ball* to be reached by the state x. During its mission the system loses control authority over p of its m actuators. These p actuators are then producing uncontrolled and undesirable inputs w. Thanks to sensors on each actuators, we can separate signal $\bar{u}$ into its uncontrolled part w and the controls u. Similarly, matrix $\bar{B}$ is split into its uncontrolled part $C \in \mathbb{R}^{n\times p}$ and its controlled actuators $B \in \mathbb{R}^{n\times(m-p)}$, leading to

$$\dot{x}(t) = Ax(t) + Bu(t) + Cw(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \quad u \in \mathcal{F}(\mathcal{U}), \quad w \in \mathcal{F}(\mathcal{W}), \tag{4.2}$$

with admissible inputs of finite energy belonging to the sets $\mathcal{F}(\mathcal{U}) := \big\{u \in \mathcal{L}_2\big([0,T],\, \mathcal{U}\big) : \|u\|_{\mathcal{L}_2} \leq 1\big\}$ and $\mathcal{F}(\mathcal{W}) := \big\{w \in \mathcal{L}_2\big([0,T],\, \mathcal{W}\big) : \|w\|_{\mathcal{L}_2} \leq 1\big\}$. We want to characterize systems that are able to reach their target even after a loss of control over some of their actuators. We now recall the definition of resilient reachability from Chapter 3.

Definition 3: The target set $\mathcal{T}$ is *resiliently reachable at time T* by system (4.2) if for any undesirable inputs $w \in \mathcal{F}(\mathcal{W})$, there exists an admissible control law $u_w \in \mathcal{F}(\mathcal{U})$ such that $x(T) \in \mathcal{T}$.

As in Chapter 3, the control law u_w can depend on the undesirable input w. Indeed, we assume to have sensors on each actuators so that all inputs to the system are available to the controller. Therefore, resilient reachability guarantees that whatever the undesirable inputs are, there is a control law *dependent on the undesirable inputs* driving the system to its target. We have the intuition that a system resilient to the loss of control over some of its actuators must be initially *overactuated*, i.e., its control matrix $\bar{B}$ has strictly more columns than rows. Since adding actuators in practice comes with a cost, we consider the following problem.

Problem 7: Determine the minimal degree of overactuation required to build a resilient system.

Definition 3 calls for the existence of a control law, so we are naturally led to our second objective.

Problem 8: For a resilient system sustaining an undesirable input w, synthesize a control law u_w that drives the system's state $x(t)$ to the target $\mathcal{T}$.

The resilience of a linear system (4.1) is mostly determined by its control matrix $\bar{B}$. Therefore, in the next two sections we first focus on driftless systems, i.e., where (4.2) becomes

$$\dot{x}(t) = Bu(t) + Cw(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \quad u \in \mathcal{F}(\mathcal{U}), \quad w \in \mathcal{F}(\mathcal{W}). \tag{4.3}$$

According to Chapter 3, the resilience of these systems is linked to the sign of the maximum of $g(h) := \|C^\top h\| - \|B^\top h\|$ over $h \in \mathbb{S}$. Since this maximum is difficult to compute, we introduce an equivalent but more convenient criteria.

Theorem 7: For $F := BB^\top - CC^\top$, the following hold:

(a) If $F \succ 0$, there exists a time t_{lim} such that $\mathcal{T}$ is resiliently reachable for all $t \geq t_{lim}$.

(b) If $F \not\succeq 0$, there exists a time t_{lim} such that $\mathcal{T}$ is not resiliently reachable for all $t > t_{lim}$.

Proof. If $F \succ 0$, then for all $h \in \mathbb{S}$, we have $0 < h^\top F h$, i.e., $h^\top CC^\top h < h^\top BB^\top h$. This is equivalent to $\max\limits_{h\,\in\,\mathbb{S}} g(h) < 0$, and thus according to Theorem 5, there exists a time t_{lim} after which $\mathcal{T}$ is resilient reachable.

Similarly, if $F \not\succeq 0$, there is $h \in \mathbb{S}$ such that $h^\top F h < 0$, i.e., $\max\limits_{h\,\in\,\mathbb{S}} g(h) > 0$. Following Theorem 3 there exists a time t_{lim} after which $\mathcal{T}$ is not resiliently reachable. □

With this simple resilient reachability condition, we investigate the resilience of a system.

4.3 Resilient control matrices

When losing control authority over p actuators, we remove the corresponding columns $j_1, \dots, j_p$ from $\bar{B}$ to form the matrix C, and we name B the remaining control matrix.

Definition 4: The control matrix $\bar{B} \in \mathbb{R}^{n \times m}$ is *p-resilient* if for all pairwise distinct $j_1, \dots, j_p \in [\![1, m]\!]$ the system following the driftless dynamics (4.3) can resiliently reach any target ball.

Definition 5: The *degree of resilience* of matrix $\bar{B}$ is the largest $p \in \mathbb{N}$ for which $\bar{B}$ is p-resilient.

Proposition 2: The control matrix $\bar{B}$ is p-resilient if and only if $\max\limits_{h \in \mathbb{S}} g(h) < 0$ for all pairwise distinct $j_1, \dots, j_p \in [\![1, m]\!]$, if and only if $F \succ 0$ for all pairwise distinct $j_1, \dots, j_p \in [\![1, m]\!]$.

Proof. If $\max\limits_{h \in \mathbb{S}} g(h) < 0$ for all pairwise distinct $j_1, \dots, j_p \in [\![1, m]\!]$, then from Theorem 5, any target ball is resiliently reachable, so $\bar{B}$ is p-resilient. On the other hand, assume $\bar{B}$ is p-resilient. For all pairwise distinct $j_1, \dots, j_p \in [\![1, m]\!]$, the continuous function g reaches a maximum g_{max} over the compact set $\mathbb{S}$. If $g_{max} \geq 0$, then from Theorems 3 and 4 some target balls are not resiliently reachable. Thus, $g_{max} < 0$. From Theorem 7, the equivalence with $F \succ 0$ holds. □

Proposition 2 enables us to assess the p-resilience of a system with m actuators by verifying the positive definiteness of $\binom{m}{p}$ matrices. We now establish several results concerning resilient matrices in order to address Problem 7.

Proposition 3: If $\bar{B}$ is p-resilient with $p \geq 1$, then $\bar{B}\bar{B}^\top \succ 0$.

Proof. Assume that $\bar{B}\bar{B}^\top$ is not positive definite. Then, there exists $x \neq 0$ such that $x^\top \bar{B}\bar{B}^\top x \leq 0$. Let C be the last column of $\bar{B}$, so that $\bar{B}\bar{B}^\top = \begin{bmatrix} B & C \end{bmatrix} \begin{bmatrix} B^\top \\ C^\top \end{bmatrix} = BB^\top + CC^\top$. Then, $F = BB^\top - CC^\top = \bar{B}\bar{B}^\top - 2CC^\top$. Thus, $x^\top F x = x^\top \bar{B}\bar{B}^\top x - 2x^\top CC^\top x \leq 0 - 2\|C^\top x\|^2 \leq 0$, so $F \not\succ 0$. By Proposition 2, $\bar{B}$ is not 1-resilient and thus not p-resilient either since $p \geq 1$. □

Proposition 4: If $\bar{B}$ is p-resilient with $p \geq 1$, then the system is overactuated.

Proof. Assume $\bar{B} \in \mathbb{R}^{n \times m}$ is not overactuated, i.e., $m \leq n$. After losing control of one actuator, the remaining control matrix B has n rows and at most $n-1$ columns. From [100], $\text{rank}(BB^\top) \leq \text{rank}(B) \leq n - 1$. Since $BB^\top \in \mathbb{R}^{n \times n}$, it is not invertible. Then, $BB^\top \not\succ 0$, so $F = BB^\top - CC^\top \not\succ 0$ either. According to Proposition 2, $\bar{B}$ is not p-resilient. □

Intuitively a system without actuator redundancy cannot be resilient, because a malfunctioning actuator cannot be counteracted. On the other hand, numerous copies of each actuator guarantees resilience. In between these extremes lies a minimum degree of overactuation required for resilience.

Proposition 5: The degree of resilience of $\bar{B}$ is not affected by left multiplication by an invertible matrix.

Proof. For P invertible, $PFP^\top \succ 0$ if and only if $F \succ 0$. Proposition 2 concludes the proof. □

We can now simplify the resilience investigation with the Singular Value Decomposition (SVD). The compact SVD [103] of $\bar{B}$ is UDV, with U orthogonal of size $n \times n$, D a diagonal matrix gathering the n singular values of $\bar{B}$, and V of size $n \times m$ with orthonormal rows, i.e., $VV^\top = I_n$.

Proposition 6: The following statements hold for $p \geq 1$:

(a) If $\bar{B}$ is p-resilient, then V is also p-resilient.

(b) If V is p-resilient and $\bar{B}\bar{B}^\top \succ 0$, then $\bar{B}$ is p-resilient.

Proof. (a) According to Proposition 3, $\bar{B}\bar{B}^\top \succ 0$, so the singular values of $\bar{B}$ are non-zero [103]. Then, D is invertible and since U is also invertible, Proposition 5 states that $\bar{B} = UDV$ and V have the same degree of resilience.

(b) Since $\bar{B}\bar{B}^\top \succ 0$, the matrix D is invertible. Then, by Proposition 5, $\bar{B}$ has the same degree of resilience as V.

□

Following Proposition 6 we study V to determine the resilience of $\bar{B}$ since, contrary to $\bar{B}$, V has the practical property of having orthonormal rows. We split V similarly to $\bar{B}$ into a controlled part B_V and an uncontrolled part C_V.

Proposition 7: The matrix $V \in \mathbb{R}^{n\times m}$ is p-resilient if and only if $\sigma_{max}^{C_V^\top} < \frac{1}{\sqrt{2}}$ for all $\binom{m}{p}$ possible C_V matrices.

Proof. We investigate whether $F_V := B_V B_V^\top - C_V C_V^\top \succ 0$. Note that $VV^\top = B_V B_V^\top + C_V C_V^\top$. Since $VV^\top = I_n$, $F_V = VV^\top - 2C_V C_V^\top = I_n - 2C_V C_V^\top$. Let λ be an eigenvalue of F_V: $0 = \det\big(\lambda I_n - F_V\big) = \det\big((\lambda - 1)I_n + 2C_V C_V^\top\big) = \big(-2\big)^n \det\Big[\Big(\frac{1-\lambda}{2}\Big)I_n - C_V C_V^\top\Big]$. Define $s := \frac{1-\lambda}{2}$, so that s is an eigenvalue of $C_V C_V^\top$. Let $x \neq 0$ be an eigenvector such that $C_V C_V^\top x = sx$. A left multiplication by $x^\top$ lead to $\|C_V^\top x\|^2 = s\|x\|^2$, so $s \geq 0$.

Then, $\sqrt{s}$ is a singular value of $C_V^\top$. We note that $\lambda > 0$ if and only if $\sqrt{s} < \frac{1}{\sqrt{2}}$. Since $\sigma_{max}^{C_V^\top}$ is the maximal singular value of $C_V^\top$, $F_V \succ 0$ if and only if $\sigma_{max}^{C_V^\top} < \frac{1}{\sqrt{2}}$. □

With Propositions 6 and 7, Problem 7 is now within our reach for 1-resilient matrices.

4.4 Minimal size of resilient matrices

We now establish a necessary condition determining the minimal size of a 1-resilient control matrix.

Theorem 8: If $\bar{B} \in \mathbb{R}^{n\times m}$ is 1-resilient, then $m \geq 2n+1$.

Proof. Since $\bar{B}$ is 1-resilient, Proposition 6 states that V is also 1-resilient. Let C_j be the columns of V and r_i its orthonormal rows, i.e., $\|r_i\| = 1$. Then,

$$\sum_{j=1}^{m} \|C_j\|^2 = \sum_{j=1}^{m}\sum_{i=1}^{n} V_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} V_{ij}^2 = \sum_{i=1}^{n} \|r_i\|^2 = n. \tag{4.4}$$

Thus, $\max_j \|C_j\|^2 \geq \frac{n}{m}$. From [100], $\max_j \|C_j\|^2 = \big(\sigma_{max}^{C^\top}\big)^2$. Then, Proposition 7 yields $\frac{n}{m} < \frac{1}{2}$, i.e., $m \geq 2n+1$. □

Theorem 8 shows that at least $2n+1$ actuators are required to have a 1-resilient control system in n dimensions. We will now prove that $n \times (2n+1)$ is in fact the minimal size of 1-resilient matrices by producing such a matrix for all $n \in \mathbb{N}$. Let $\bar{B}_k := [I_n \dots I_n\ D]$ the matrix composed of k identity matrices I_n and a column vector $D := \frac{1}{\sqrt{n}}[1 \dots 1]^\top$.

Proposition 8: The matrix $\bar{B}_{2p}$ is p-resilient.

Proof. We calculate the maximum of $g(h) = \|C^\top h\| - \|B^\top h\|$ over $h \in \mathbb{S}$ for all possible losses of p columns among the $2p+1$ columns of matrix $\bar{B}_{2p}$.

First, assume the system loses control of p columns belonging all to the identity matrices. Without loss of generality we assume losing one column per matrix. The index of the column lost in the i^{th} identity matrix is $j_i \in [\![1, n]\!]$. These columns form the matrix $C = \big[e_{j_1} \dots e_{j_p}\big]$, while B is the remaining control matrix. Then, $h = \big(h_1, \dots, h_n\big) \in \mathbb{S}$, i.e., $\|h\|^2 = 1$, yields $C^\top h = \big(h_{j_1}, \dots, h_{j_p}\big)$ so $\|C^\top h\|^2 = \sum_{i=1}^p h_{j_i}^2$,

$$\|B^\top h\|^2 = 2p \sum_{i=1}^n h_i^2 - \sum_{i=1}^p h_{j_i}^2 + \left(\sum_{i=1}^n \frac{h_i}{\sqrt{n}}\right)^2 = 2p - \sum_{i=1}^p h_{j_i}^2 + \frac{1}{n}\left(\sum_{i=1}^n h_i\right)^2,$$

$$g(h) < 0 \iff \sum_{i=1}^p h_{j_i}^2 < 2p - \sum_{i=1}^p h_{j_i}^2 + \frac{1}{n}\left(\sum_{i=1}^n h_i\right)^2 \iff \sum_{i=1}^p h_{j_i}^2 < p + \frac{1}{2n}\left(\sum_{i=1}^n h_i\right)^2. \tag{4.5}$$

If $j_1 = \ldots = j_p$, and $h = e_{j_1}$, then (4.5) simplifies into $p < p + \frac{1}{2n}$, which is true.

If at least one of the j_i is different from the others, then at least two different components of h are present in the sum $\sum_{i=1}^p h_{j_i}^2$. Because $\|h\| = 1$, vector h cannot have two components both equal to 1, at least one of them is strictly inferior to 1. Assume without loss of generality that $h_{j_1} < 1$. Because $\|h\| = 1$, we also have $h_{j_i} \leq 1$. Thus, $h_{j_1}^2 + \sum_{i=2}^p h_{j_i}^2 < 1 + \sum_{i=2}^p h_{j_i}^2 \leq 1 + (p-1) = p$.

Another possible case, is that $j_1 = \ldots = j_p$ but $h \neq e_{j_1}$. Because $\|h\| = 1$, $h_{j_1} < 1$ otherwise we would have $h = e_{j_1}$. Then, $\sum_{i=1}^p h_{j_i}^2 = p h_{j_1}^2 < p$. These were the only two other possible cases and in each of them some $h_{j_i} < 1$, so $\sum_{i=1}^p h_{j_i}^2 < p$, so the inequality also holds true. Overall $g(h) < 0$ for all $h \in \mathbb{S}$ and all choice of columns $j_1, \dots, j_p \in [\![1, n]\!]$.

The other possibility is that $\bar{B}_{2p}$ loses $p-1$ columns among the identity matrices and the last column D. Then,

$$g(h) = \sqrt{\sum_{i=1}^{p-1} h_{j_i}^2 + \frac{1}{n}\left(\sum_{i=1}^n h_i\right)^2} - \sqrt{2p - \sum_{i=1}^{p-1} h_{j_i}^2}.$$

Since $\|h\| = 1$, $h_{j_i}^2 \leq 1$ for all $i \in [\![1, p-1]\!]$. Then,

$$\left(\sum_{i=1}^n h_i\right)^2 \leq \left(\sum_{i=1}^n h_i^2\right)\left(\sum_{i=1}^n 1^2\right) = \|h\|^2 n = n,$$

by the Cauchy-Schwarz inequality [95]. Thus,

$$g(h) \leq \sqrt{p - 1 + \frac{1}{n} n} - \sqrt{2p - (p-1)} \leq \sqrt{p} - \sqrt{p+1} < 0,$$

i.e., $\max\limits_{h \in \mathbb{S}} g(h) < 0$, so $\bar{B}_{2p}$ is p-resilient according to Proposition 2. □

Therefore, $\bar{B}_2 = \big[I_n \, I_n \, D\big]$ is 1-resilient and has only $2n+1$ columns. Following Theorem 8, the minimal size of a 1-resilient matrix is then exactly $n \times (2n+1)$. We will now investigate sufficient conditions allowing to generate 1-resilient control matrices by making use of Proposition 7.

Proposition 9: Any matrix $V \in \mathbb{R}^{n \times m}$ where $m \geq 2n+1$ which has orthonormal rows and whose columns have all the same norm, is 1-resilient.

Proof. Since the columns C of matrix V have the same norm, (4.4) implies $\|C\|^2 = \frac{n}{m}$. The maximal singular value of a column vector is its norm [100], so $\sigma_{max}^{C^\top} = \|C\| = \sqrt{\frac{n}{m}}$. Since $m \geq 2n+1$, we obtain

$$\frac{n}{m} \leq \frac{1}{2} - \frac{1}{2m} < \frac{1}{2}, \quad \text{i.e.,} \quad \sigma_{max}^{C^\top} < \frac{1}{\sqrt{2}}.$$

Then, Proposition 7 states that V is 1-resilient. □

Intuitively, the columns of V having the same norm means that the actuators are equally powerful, whereas the rows having the same norm means that all the states are equally actuated. Furthermore, the orthogonality of rows enforces the necessary condition for 1-resilience of Proposition 3 by making $VV^\top$ positive definite. With Proposition 9 we can now easily generate 1-resilient matrices for any size n. For instance,

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{bmatrix} \quad \text{are both 1-resilient.}$$

We now wish to expand our minimal size investigation to higher degrees of resilience. We start by extending Proposition 9 to 2-resilient matrices with a large increase in the calculations required.

Proposition 10: Any matrix $V \in \mathbb{R}^{n \times m}$ where $m \geq 4n+1$ which has orthonormal rows and whose columns have all the same norm, with at least two columns being collinear, is 2-resilient.

Proof. Similarly as in the proof of Proposition 9 the columns have a squared norm of $\|C\|^2 = \frac{n}{m}$. We extract any two columns C_1 and C_2 from V to form C, the remaining part of V is named B. Since $C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$, we have $CC^\top = C_1C_1^\top + C_2C_2^\top$.

The singular values $\sigma^{C^\top}$ of $C^\top$ are defined as the square roots of the eigenvalues s of $CC^\top$. Therefore we calculate $s = \left(\sigma^{C^\top}\right)^2$ to use Proposition 7. From the matrix determinant lemma [100],

$$0 = \det\left(sI_n - CC^\top\right) = \det\left(sI_n - C_1C_1^\top - C_2C_2^\top\right) = \left(1 - C_2^\top\left(sI_n - C_1C_1^\top\right)^{-1}C_2\right)\det\left(sI_n - C_1C_1^\top\right).$$

If $\det\left(sI_n - C_1C_1^\top\right) = 0$, then the resulting eigenvalue is either 0 or $\|C_1\|^2 = \frac{n}{m}$ by [100]. To investigate when the other term goes to zero, we develop the inverse into a Neumann series [100] for s such that $\left\|\frac{C_1C_1^\top}{s}\right\| < 1$:

$$\begin{aligned} s\left(sI_n - C_1C_1^\top\right)^{-1} &= \left(I_n - \frac{C_1C_1^\top}{s}\right)^{-1} = \sum_{p=0}^{\infty}\left(\frac{C_1C_1^\top}{s}\right)^p \qquad (4.6) \\ &= I + \sum_{p=1}^{\infty}\frac{1}{s^p}C_1\left(C_1^\top C_1\right)^{p-1}C_1^\top = I + \frac{C_1C_1^\top}{s}\sum_{p=1}^{\infty}\left(\frac{\|C_1\|^2}{s}\right)^{p-1} \\ &= I + \frac{C_1C_1^\top}{s}\frac{1}{1 - \frac{\|C_1\|^2}{s}} = I + \frac{C_1C_1^\top}{s - \|C_1\|^2}. \end{aligned}$$

$$\begin{aligned} \text{Then,} \quad & \left(1 - C_2^\top\left(sI_n - C_1C_1^\top\right)^{-1}C_2\right) = 0 \iff s = C_2^\top s\left(sI_n - C_1C_1^\top\right)^{-1}C_2 \\ \iff & C_2^\top\left(I + \frac{C_1C_1^\top}{s - \|C_1\|^2}\right)C_2 = s = \|C_2\|^2 + \frac{\left(C_1^\top C_2\right)^2}{s - \|C_1\|^2} \\ \iff & s^2 - \left(\|C_1\|^2 + \|C_2\|^2\right)s + \|C_1\|^2\|C_2\|^2 - \left(C_1^\top C_2\right)^2 = 0. \end{aligned}$$

Recall that $\|C_1\|^2 = \|C_2\|^2 = \frac{n}{m}$. Then the previous equation becomes $s^2 - \frac{2n}{m}s + \frac{n^2}{m^2} - \left(C_1^\top C_2\right)^2 = 0$. The maximal root of this quadratic equation is

$$s_{max} = \frac{n}{m} + \left|C_1^\top C_2\right|. \tag{4.7}$$

This expansion is only valid for the case where s satisfies $\left\|\frac{C_1 C_1^\top}{s}\right\| < 1$. We note that $\|C_1 C_1^\top\| = \lambda_{max}(C_1 C_1^\top) = \|C_1\|^2 = \frac{n}{m}$, from [100]. Therefore, in the other case $s \leq \frac{n}{m}$. From (4.7) we deduce that s_{max} is the maximal eigenvalue of $CC^\top$.

The matrix C maximizing s_{max} is the one composed of two collinear columns of V. Indeed, by the Cauchy-Schwarz inequality $\left|C_1^\top C_2\right| \leq \|C_1\|\,\|C_2\|$, and the equality only happens when C_1 and C_2 are collinear. In that case, $s_{max} = \frac{2n}{m}$. Then, the resilience condition of Proposition 7 is equivalent to $2s_{max} < 1$, i.e., $m \geq 4n + 1$. Thus, V is 2-resilient. □

Note that two collinear columns of same norm are either the same or opposites. Proposition 10 thus deals with the case where at least one actuator of the system is doubled. With the guidelines provided by Proposition 10 we produce an example of a 2-resilient matrix V of size 2×10:

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}.$$

With Proposition 8 we can generate p-resilient matrices of size $n \times (2pn + 1)$. For $p = 1$ it corresponds to $n \times (2n + 1)$, which is the minimal size for 1-resilient matrices. For $p = 2$, we obtain a matrix with $4n + 1$ columns, which is consistent with the minimal size detailed in Proposition 10.

In order to determine the minimal size of a p-resilient matrix, with $p \geq 2$, the only missing result is an equivalent of Theorem 8 for higher degrees of resilience. However, the proof of Theorem 8 becomes increasingly complicated as the desired degree of resilience increases. Indeed, the case $p = 2$ treated in Proposition 10 requires significantly more complex calculations than for the case $p = 1$ of Proposition 9. Without the assumption of same column norm for the case $p = 2$ the calculations do not even reach a conclusion. For $p \geq 3$, the calculations become even more cumbersome. The Neumann series (4.6) becomes

$$s\left(sI_n - \sum_{j=1}^{p-1} C_j C_j^\top\right)^{-1} = \sum_{k=0}^{\infty}\left(\sum_{j=1}^{p-1}\frac{C_j C_j^\top}{s}\right)^k.$$

We would then need the multinomial formula to calculate each term of the series:

$$\left(\sum_{j=1}^{p-1} C_j C_j^\top\right)^k = \sum_{i_1+\ldots+i_{p-1}=k}\binom{k}{i_1,\ldots,i_{p-1}}\prod_{j=1}^{p-1}\left(C_j C_j^\top\right)^{i_j}.$$

Proceeding to the separation of $\left(C_j C_j^\top\right)^{i_j}$ into a scalar part with the power $i_j - 1$ and a matrix part like we did for $p = 2$ is still possible but brings numerous cross-terms that did not appear for $p = 2$. Because of the complexity of the calculations for $p \geq 2$, we were unable to obtain a simple necessary condition on the minimal size of such p-resilient matrices.

Remark: If we based our intuition about the minimal size of p-resilient matrices on Theorem 8 and on Proposition 8, then we might conjecture a minimal size of $n \times (2pn + 1)$ for p-resilient matrices $\bar{B}$.

Such a conjecture holds for 2-resilient matrices with a state dimension $n = 1$. Indeed, let us consider $\bar{B} = \begin{bmatrix} b_1 \; b_2 \; b_3 \; b_4 \end{bmatrix}$. Without loss of generality, assume that b_3 and b_4 have a greater absolute value than b_1 and b_2. When losing control of the last two columns we form $B = \begin{bmatrix} b_1 \; b_2 \end{bmatrix}$ and $C = \begin{bmatrix} b_3 \; b_4 \end{bmatrix}$. Then, $F = BB^\top - CC^\top = b_1^2 + b_2^2 - b_3^2 - b_4^2 \leq 0$. Therefore, there are no 2-resilient matrices of size 1×4. The minimal size of a 2-resilient matrix for $n = 1$ is then 1×5, since $\begin{bmatrix} 1 \; 1 \; 1 \; 1 \; 1 \end{bmatrix}$ is 2-resilient.

However, we are able to generate 2-resilient matrices of size $n \times 4n$ for $n = 6$ and $n = 8$, and even of size $n \times (4n - 2)$ for $n = 12$. We will now provide the intuition that led us to these counterexamples.

We consider a matrix $V \in \mathbb{R}^{n \times m}$ with orthogonal rows whose only elements are ± 1. Obviously, all columns have the same norm: $\|C\|^2 = n$, and the maximal singular value of $CC^\top$ defined in (4.7) becomes $s_{max} = \left| C_1^\top C_2 \right| + n$, with the notations from the proof of Proposition 10. To build a 2-resilient matrix of minimal size, we need to minimize s_{max}. Indeed, for these matrices the resilience condition of Proposition 7 becomes $2s_{max} < m$. For a small s_{max}, we should then be able to have a small number m of columns. To minimize s_{max}, V should not have any collinear columns, because they would maximize the scalar product $\left| C_1^\top C_2 \right|$, as seen in the proof of Proposition 10.

There are 2^n different vectors composed of n elements ± 1. These vectors are only collinear with the vector of opposite sign. Thus, there are 2^{n-1} of such non-collinear vectors. To build a matrix with $4n$ columns, we then require $2^{n-1} \geq 4n$. The minimal dimension realizing that condition is $n = 6$. We believe that it is impossible to build a 2-resilient matrix of $4n$ columns for $n \leq 5$.

We propose two ways of generating a 2-resilient matrix with $4n$ columns for $n \geq 6$. The first approach consists in producing all the non-collinear vectors and then selecting $4n$ of them to create a matrix with orthogonal rows. With this approach, we were able to generate a 2-resilient matrix of size 6×24 shown in (4.8).

$$\bar{B} = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1
\end{bmatrix}. \tag{4.8}$$

The other approach uses Hadamard matrices [104]. These matrices are square, orthogonal, and composed of only ± 1. By carefully selecting n rows of a $4n \times 4n$ Hadamard matrix, it is possible to have $4n$ non-collinear columns. We extracted 8 chosen rows of a 32×32 Hadamard matrix and we built a 2-resilient matrix of minimal size 8×32 as shown in (4.9).

$$\bar{B}^\top = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}. \tag{4.9}$$

In order to generate a 2-resilient matrix with an even lower degree of overactuation, the maximal scalar product in (4.7) must be made even smaller. We succeeded by taking $n = 12$ and selecting n partial rows from a $4n \times 4n$ Hadamard matrix in order to obtain a 2-resilient matrix of size $n \times (4n - 2)$. This matrix is not shown here because its size exceeds a single page and showing its 552 entries of ±1 has little interest.

Therefore the above conjecture is wrong. Its demise also explains why the proof of Theorem 8 cannot be extended to higher degrees of resilience. It is now time to tackle Problem 8, the generation of a control law for resilient systems.

4.5 Control synthesis

The definition of resilient reachability, Definition 3 requires the existence of a control law u_w depending on the undesirable input w, but so far we never synthesized one. A natural follow-up question is thus one of designing such a control law. When B is not invertible, this question is not trivial as the control u needs to counteract $w \in \mathcal{F}(\mathcal{W})$ while remaining in $\mathcal{U}$.

Theorem 9: If $F \succ 0$, then there exists $\alpha > 0$ such that

$$u(t) := B^\top \big(BB^\top\big)^{-1}\Big(- Cw(t) + \alpha\big(x_{goal} - x(t)\big)\Big) \tag{4.10}$$

drives the state of (4.3) to x_{goal} and $u \in \mathcal{F}(\mathcal{U})$ for any $w \in \mathcal{F}(\mathcal{W})$.

Proof. Since $F = BB^\top - CC^\top \succ 0$, obviously $BB^\top \succ 0$, so $BB^\top$ is invertible, and (4.10) is well-defined. If we plug (4.10) into the state equation (4.3) we obtain $\dot{x} = \alpha\big(x_{goal} - x\big)$. The solution is $x(t) = x_{goal} + e^{-\alpha t}d$, with $d = x(0) - x_{goal}$. Since $\alpha > 0$, $x(t)$ converges globally exponentially to x_{goal}, the control law is successful.

We now need to prove that $\|u\|_{\mathcal{L}_2} \leq 1$ for all $w \in \mathcal{F}(\mathcal{W})$. Note that $x_{goal} - x(t) = -e^{-\alpha t}d$, and let $\upsilon(t) := Cw(t) + \alpha e^{-\alpha t}d$, so that $u(t) = -B^\top \big(BB^\top\big)^{-1}\upsilon(t)$. Then,

$$\|u\|_{\mathcal{L}_2}^2 = \int_0^T u(t)^\top u(t)\, dt = \int_0^T \upsilon(t)^\top \big(BB^\top\big)^{-1} \upsilon(t)\, dt.$$

To simplify, let $P := \big(BB^\top\big)^{-1} \succ 0$, and expand $\upsilon(t)$ as

$$\upsilon(t)^\top P \upsilon(t) = \underbrace{w(t)^\top C^\top P C w(t)}_{=\,T_1} + \underbrace{w(t)^\top C^\top P \alpha e^{-\alpha t} d}_{=\,T_2} + \underbrace{\alpha e^{-\alpha t} d^\top P C w(t)}_{=\,T_3} + \underbrace{\alpha^2 d^\top e^{-\alpha t} P e^{-\alpha t} d}_{=\,T_4} \tag{4.11}$$

We start by upper bounding the term T_1. Let λ_M be the maximal eigenvalue of the positive semidefinite matrix $C^\top \big(BB^\top\big)^{-1} C$. Then,

$$\int_0^T T_1 dt = \int_0^T w(t)^\top C^\top \big(BB^\top\big)^{-1} C w(t)\, dt \leq \int_0^T w(t)^\top \lambda_M w(t)\, dt \leq \lambda_M \|w\|_{\mathcal{L}_2}^2 \leq \lambda_M, \tag{4.12}$$

since $\|w\|_{\mathcal{L}_2} \leq 1$ for all $w \in \mathcal{F}(\mathcal{W})$. We will now prove that $\lambda_M < 1$ through some convoluted linear algebra. From the Woodbury formula [100], $\big(I + C^\top F^{-1} C\big)$ is invertible, because F is invertible. Then, we can rewrite matrix P as

$$P = \big(BB^\top\big)^{-1} = \big(F + CC^\top\big)^{-1} = F^{-1} - F^{-1}C\big(I + C^\top F^{-1} C\big)^{-1} C^\top F^{-1}.$$

Let $D := C^\top F^{-1} C$. Then, $C^\top P C = D - D\big(I + D\big)^{-1} D$. Expanding $\big(I + D\big)^{-1}\big(I + D\big) = I$ yields $\big(I + D\big)^{-1} D = I - \big(I + D\big)^{-1}$, so that $C^\top P C = D - D + D\big(I + D\big)^{-1}$. Similarly, from $\big(I + D\big)\big(I + D\big)^{-1} = I$, we finally obtain

$$C^\top \big(BB^\top\big)^{-1} C = I - \big(I + D\big)^{-1}.$$

Let λ be an eigenvalue of $C^\top \big(BB^\top\big)^{-1} C$. Then,

$$\begin{aligned} 0 &= \det\big(\lambda I - C^\top \big(BB^\top\big)^{-1} C\big) = \det\big(\lambda I - I + \big(I + D\big)^{-1}\big) = \det\big((\lambda - 1)(I + D)(I + D)^{-1} + I(I + D)^{-1}\big) \\ &= \det\big((\lambda - 1)(I + D) + I\big)\det(I + D)^{-1}. \end{aligned}$$

From the Woodbury formula we know that $(I + D)$ is invertible, so $\det(I + D))^{-1} \neq 0$. If $\lambda = 1$, then $\det(I) = 0$, which is absurd. Thus $\lambda \neq 1$, so we can divide by $(\lambda - 1)$:

$$0 = \det\Big(I + D + \frac{1}{\lambda - 1} I\Big) = \det\Big(\frac{\lambda}{\lambda - 1} I + D\Big).$$

Since $F^{-1} \succ 0$, $C^\top F^{-1} C \succeq 0$, i.e. the eigenvalues of D are nonnegative, so $\frac{-\lambda}{\lambda-1} \geq 0$. Since $BB^\top \succ 0$, $\big(BB^\top\big)^{-1} \succeq 0$ and then $C^\top \big(BB^\top\big)^{-1} C \succeq 0$, thus $\lambda \geq 0$. Then $\lambda - 1 < 0$, i.e. $\lambda < 1$, and hence $\lambda_M < 1$.

We can now tackle the integral of the second term of (4.11):

$$\int_0^T T_2 dt = \int_0^T \alpha w(t)^\top C^\top P e^{-\alpha t} d\, dt = \alpha \int_0^T w(t)^\top e^{-\alpha t} dt\ C^\top P d.$$

We use the Cauchy-Schwarz inequality

$$\begin{aligned} \left\| \int_0^T w(t)^\top e^{-\alpha t} dt \right\| &= \sqrt{\sum_{i=1}^m \left(\int_0^T w_i(t) e^{-\alpha t} dt \right)^2} \leq \sqrt{\sum_{i=1}^m \left(\int_0^T w_i^2(t) dt \right) \left(\int_0^T e^{-2\alpha t} dt \right)} \\ &= \sqrt{\left[\frac{e^{-2\alpha t}}{-2\alpha} \right]_0^\top \int_0^T \sum_{i=1}^m w_i^2(t) dt} = \sqrt{\frac{1 - e^{-2\alpha T}}{2\alpha}}\ \|w\|_{\mathcal{L}_2}. \end{aligned} \tag{4.13}$$

$$\text{Thus,} \quad \int_0^T T_2\ dt \leq \sqrt{\frac{\alpha}{2}} \|C^\top P d\|\ \|w\|_{\mathcal{L}_2}, \quad \text{and similarly,} \quad \int_0^T T_3\ dt \leq \sqrt{\frac{\alpha}{2}} \|d^\top P C\|\ \|w\|_{\mathcal{L}_2}. \tag{4.14}$$

We also simplify the integral of the fourth term of (4.11):

$$\begin{aligned} \int_0^T T_4 &= \int_0^T \alpha^2 d^\top e^{-\alpha t} P e^{-\alpha t} d\ dt = \alpha^2 d^\top P d \int_0^T e^{-2\alpha t}\ dt = \alpha^2 d^\top P d \left[\frac{e^{-2\alpha t}}{-2\alpha} \right]_0^\top \\ &= \frac{\alpha}{2} d^\top P d \big(1 - e^{-2\alpha T}\big) \leq \frac{\alpha}{2}\ d^\top P d. \end{aligned} \tag{4.15}$$

Then, we combine (4.12), (4.13), (4.14) and (4.15):

$$\|u\|^2_{\mathcal{L}_2} \ \leq\ \frac{\alpha}{2}\ d^\top P d + 2\sqrt{\frac{\alpha}{2}} \|C^\top P d\| + \lambda_M. \tag{4.16}$$

Since $\lambda_M < 1$, and d, P and C are constant, we can choose α small enough so that the right hand side of (4.16) is smaller than 1, which finally leads to $\|u\|^2_{\mathcal{L}_2} \leq 1$, i.e. $u \in \mathcal{F}(\mathcal{U})$. □

The proof of Theorem 9 provides a constructive method of finding α satisfying the claim of the theorem. The maximum α satisfying Theorem 9 and thus ensuring the fastest convergence to x_{goal} is given by

$$\alpha^* = 2\ \frac{\big(\sqrt{b^2 + (1 - \lambda_M) a} - b\big)^2}{a^2}, \qquad \text{with} \quad a = d^\top P d \quad \text{and} \quad b = \|C^\top P d\|. \tag{4.17}$$

Theorem 9 gives an intuitive validation of the work developed in the previous sections. Indeed, we established that resilient reachability implies $F \succ 0$. From Theorem 9, we see that such a condition is indeed sufficient to build a control law of the form (4.10).

The positive definiteness of F brings two results. The part $BB^\top \succ 0$ guarantees the existence of u. But $BB^\top$ is more than just positive definite, in fact $BB^\top \succ CC^\top$. This relation ensures that u of the form (4.10) remains within the bounds of $\mathcal{U}$ even when w is maximal.

We finally return to the general linear system (4.2). We will show that a control law similar to (4.10) can be used if matrix A is not overly unstable. The intuition is that the magnitude of u in excess of w can be used to counteract instability of A to a certain extent. We formalize our intuition below.

For all $\eta > \max(Re(\lambda(A)))$, we can find $\beta > 0$ such that $\|e^{At}\| \leq \beta e^{\eta t}$ for all $t \geq 0$ [105]. Using

$P = \big(BB^\top\big)^{-1}$ and $\lambda_M = \max\big(\lambda(C^\top PC)\big)$ we define for all $\eta \in \mathbb{R}$ the set

$$\mathcal{A}_\eta := \Big\{ \alpha > \eta : \ \lambda_M + \frac{\alpha}{\sqrt{\alpha - \eta}}\sqrt{2}\beta \|C^\top P\| \|x_0\| + \frac{\alpha^2}{\alpha - \eta}\frac{\beta^2}{2}\|P\|\|x_0\|^2 \leq 1 \Big\}. \tag{4.18}$$

We showed in the proof of Theorem 9 that $F \succ 0$ implies $\lambda_M < 1$. Then, taking α sufficiently small would satisfy the condition of (4.18), as long as η is even smaller. There is of course a trade-off here because taking η very close to $\max(Re(\lambda(A)))$ leads to a larger β and thus requires an even smaller α to satisfy the inequality in (4.18).

Theorem 10: If $F \succ 0$ and if there exists $\eta > \max(Re(\lambda(A)))$ such that $\mathcal{A}_\eta$ is not empty, then for all $\alpha \in \mathcal{A}_\eta$ the control law

$$u(t) := B^\top \big(BB^\top\big)^{-1}\big(-Cw(t) - \alpha x(t)\big) \tag{4.19}$$

drives the resilient system (4.2) to the origin, and $u \in \mathcal{F}(\mathcal{U})$ for any $w \in \mathcal{F}(\mathcal{W})$.

Remark: In contrast with the driftless case of Theorem 9, having $F \succ 0$ is not sufficient anymore for resilience. Indeed, the existence of $\alpha \in \mathcal{A}_\eta$ in Theorem 10 depends on the eigenvalues of matrix A having sufficiently small real part.

Proof. When plugging control law (4.19) into (4.2), the dynamics become $\dot{x}(t) = Ax(t) - \alpha x(t)$. Then, $x(t) = e^{\tilde{A}t}x_0$ with $\tilde{A} := A - \alpha I$. Since $\alpha > \eta > \max(Re(\lambda(A)))$, matrix $\tilde{A}$ is Hurwitz, which guarantees the convergence of the state to the origin. Then, we need to verify whether $u \in \mathcal{F}(\mathcal{U})$ for all $w \in \mathcal{F}(\mathcal{W})$.

We first bound the state transition matrix: $\|e^{\tilde{A}t}\| = \|e^{(A-\alpha I)t}\| = \|e^{At}e^{-\alpha t}\| \leq \beta e^{\eta t}e^{-\alpha t} = \beta e^{-\gamma t}$, with $\gamma := \alpha - \eta > 0$. Now, we can proceed as in the proof of Theorem 9. Since $F \succ 0$, we have $P \succ 0$ and $\|u\|^2_{\mathcal{L}_2} = \int_0^T \nu(t)^\top P \nu(t)\, dt$, with $\nu(t) = Cw(t) + \alpha e^{\tilde{A}t}x_0$. We expand the terms of this expression as in the proof of Theorem 9:

$$\nu(t)^\top P\nu(t) = \underbrace{w(t)^\top C^\top PCw(t)}_{=\,T_1} + 2\underbrace{w(t)^\top C^\top P\alpha e^{\tilde{A}t}x_0}_{=\,T_2} + \underbrace{\alpha^2 x_0^\top e^{\tilde{A}^\top t}Pe^{\tilde{A}t}x_0}_{=\,T_4} = T_1 + 2T_2 + T_4.$$

Note that T_1 is exactly the same as in (4.11), so that $\int_0^T T_1\, dt \leq \lambda_M$ according to (4.12). However, in the terms T_2 and T_4 the scalar exponential $e^{-\alpha t}$ of (4.11) has now been replaced by a matrix exponential $e^{\tilde{A}t}$. We use Hölder's inequality [95] to split the following integral:

$$\begin{aligned}\int_0^T T_2\, dt &\leq \int_0^T \big|w(t)^\top \big(C^\top P\alpha e^{\tilde{A}t}x_0\big)\big|\, dt \leq \sqrt{\int_0^T \|w(t)\|^2\, dt}\sqrt{\int_0^T \big\|C^\top P\alpha e^{\tilde{A}t}x_0\big\|^2\, dt} \\ &\leq \|w\|_{\mathcal{L}_2}\, \|C^\top P\|\, \alpha\, \|x_0\| \sqrt{\int_0^T \beta^2 e^{-2\gamma t}\, dt} \leq \alpha\beta\, \|C^\top P\|\, \|x_0\| \sqrt{\frac{1 - e^{-2\gamma T}}{2\gamma}} \\ &\leq \frac{\alpha}{\sqrt{\alpha - \eta}}\frac{\beta}{\sqrt{2}}\|C^\top P\|\, \|x_0\|,\end{aligned}$$

where we used $\|w\|_{\mathcal{L}_2} \leq 1$ since $w \in \mathcal{F}(\mathcal{W})$. For the term T_4, we have

$$\int_0^T T_4\, dt \leq \alpha^2\|x_0\|^2\|P\| \int_0^T \beta^2 e^{-2\gamma t}\, dt = \alpha^2\|x_0\|^2\|P\|\beta^2\left(\frac{1 - e^{-2\gamma T}}{2\gamma}\right) \leq \frac{\alpha^2}{\alpha - \eta}\frac{\beta^2}{2}\|P\|\|x_0\|^2$$

Then,

$$\|u\|_{\mathcal{L}_2}^2 \leq \lambda_M + \frac{\alpha}{\sqrt{\alpha - \eta}}\sqrt{2}\beta\|C^\top P\|\|x_0\| + \frac{\alpha^2}{\alpha - \eta}\frac{\beta^2}{2}\|P\|\|x_0\|^2.$$

Since we assumed that $\alpha \in \mathcal{A}_\eta$, we have $\|u\|_{\mathcal{L}_2} \leq 1$ according to (4.18). Hence, $u \in \mathcal{F}(\mathcal{U})$. □

Note that the set $\mathcal{A}_\eta$ depends on $\|x_0\|$. Therefore, the further away the initial state is, the less instability can be counteracted by the control law. From Theorem 10 we can easily derive a sufficient condition for resilience and confirm our intuition about stable systems.

Corollary 2: If A is Hurwitz and $\bar{B}$ is p-resilient, then the system $\dot{x} = Ax + \bar{B}u$ is also p-resilient.

Proof. Since $\bar{B}$ is p-resilient, we can lose p actuators and create $F \succ 0$. For a target $\mathcal{T}$, set $\mathcal{A}$ has a maximum α^*. Since A is Hurwitz, $\alpha^* > 0 > \max(Re(\lambda(A)))$. Thus, the control law (4.10) drives the state to $\mathcal{T}$, the system is p-resilient. □

4.6 Numerical example

To validate our theory, we consider the ADMIRE fighter jet model developed by the Swedish Defense Research Agency [106] and used as an application case in several control frameworks [107], [108]. We explore three different scenarios featuring the fighter jet. First, we investigate the resilience of the simplified model used in [107]. We also use this model as a benchmark to compare our approach with a robust control method. We finally study the resilience of a more advanced driftless dynamics model of the aircraft.

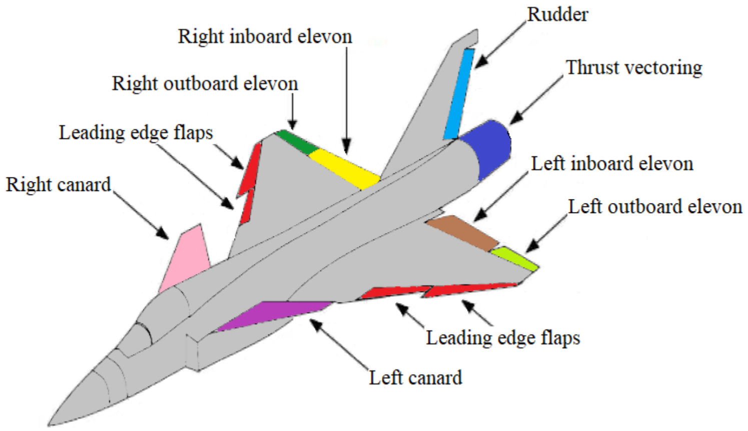


Figure 4.1: The ADMIRE fighter jet model. Image modified from [108].

4.6.1 Resilience of a fighter jet

We consider only four of the actuators of the jet: the canard, the left and right elevons and the rudder, as depicted on Figure 4.1. With these control surfaces, the pilot can directly affect the angular acceleration in roll, pitch and yaw.

The nominal linearized dynamics of the jet established in [107] are $\dot{x}(t) = Ax(t) + \bar{B}\bar{u}(t)$, with the state vector x gathering the angular velocities in roll, pitch and yaw (rad/s):

$$x = \begin{bmatrix} p \\ q \\ r \end{bmatrix} \quad A = \begin{bmatrix} -0.997 & 0 & 0.618 \\ 0 & -0.506 & 0 \\ -0.094 & 0 & -0.213 \end{bmatrix} \quad \bar{B} = \begin{bmatrix} 0 & -4.242 & 4.242 & 1.487 \\ 1.653 & -1.274 & -1.274 & 0.002 \\ 0 & -0.281 & 0.281 & -0.882 \end{bmatrix}.$$

Note that the system is stable since the eigenvalues of A have negative real parts. The inputs of the system are the deflections of the control surfaces: u_c for the canard wings, u_{re} and u_{le} for the right and left elevons, and u_r for the rudder. They are mechanically constrained:

$$u = \big(u_c,\ u_{re},\ u_{le},\ u_r\big) \quad \text{with} \quad u_c \in [-25, 55]\frac{\pi}{180}, \quad \text{and} \quad u_{re}, u_{le}, u_r \in [-30, 30]\frac{\pi}{180}. \tag{4.20}$$

Consider the scenario in which, after sustaining damage (e.g., during air combat), one of the control surfaces of the fighter jet stops responding to the commands. This surface is now producing undesirable inputs. The pilot wants to minimize the aircraft roll, pitch and yaw rates, so the target is a ball of radius 0.1 centered around the origin, $x = 0$.

By looking at the matrix $\bar{B}$ we can build our intuition on the resilience of the system. The first column represents the effect of the canard and only modifies the pitch rate of the aircraft. This actuator can be counteracted by the combined actions of both elevons, because $1.2735 + 1.2735 > 1.6532$. The elevons can counteract each other in terms of roll but doing so would induce a high pitching moment that cannot be counteracted. The yawing moment produced by the rudder cannot be counteracted by the other actuators: $0.8823 > 0.2805 + 0.2805$. Therefore, our intuition states that the fighter jet is only resilient to the loss of control authority over the canard.

We check whether the matrix $F = BB^\top - CC^\top \succ 0$ for each of the four possible actuator losses. Table 4.1 gathers the minimal eigenvalues of F for the four cases. As predicted by our intuition, the jet is only resilient to the loss of control authority over the canard.

Table 4.1: Minimal eigenvalue of F for each actuator losses

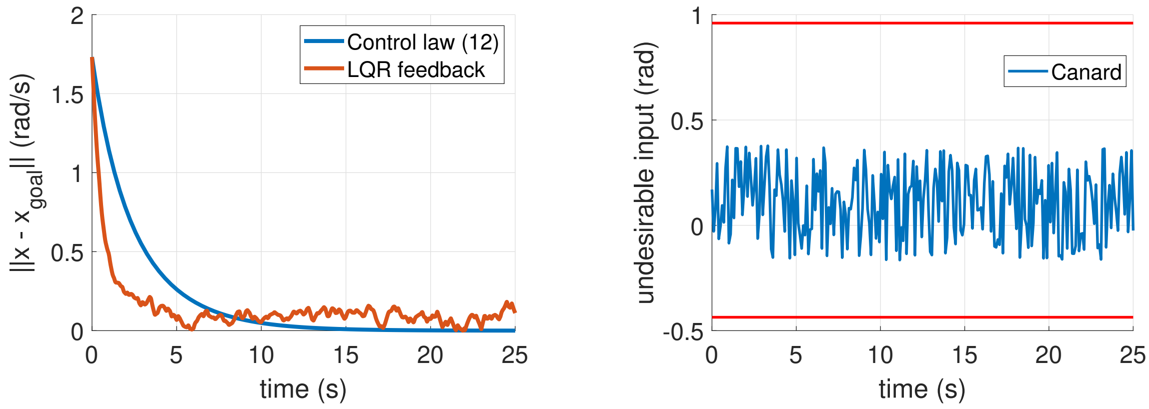
Loss of control of:	Canards	Right elevon	Left elevon	Rudder
$\min \lambda(F)$	0.51	-8.5	-8.5	-1.0

We study more in-depth the loss of control over the canard with Theorem 10. We reuse the notations employed in the proof and after some calculations we obtain: $\lambda_M = 0.8426 < 1$, $\max(Re(\lambda(A))) = -0.259 < \alpha^*$, so the control law (4.19) should work.

We simulate our system on MATLAB with $ode45$ on the time interval $[0, 25]$. We generate w as a stochastic signal between the bounds of u_c defined in (4.20), i.e., $w(t) \in [-25, 55]\frac{\pi}{180}$ for $t \in [0, 25]$. If $\|w\|_{\mathcal{L}_2} > 1$, we divide w by its $\mathcal{L}_2$ norm so that once normalized, $\|w\|_{\mathcal{L}_2} = 1$. If instead we initially had $\|w\|_{\mathcal{L}_2} \leq 1$, then we keep w as is. In order to respect the constraints (4.20) we add a saturation to the control law (4.19) and to the LQR feedback control law $u_{LQR} = -Kx$ that we use as a reference. On MATLAB we obtain

$$K = \begin{bmatrix} -0.5825 & -0.5358 & -0.1659 \\ 0.5826 & -0.5360 & 0.1653 \\ 0.2198 & 0.0007 & -0.7564 \end{bmatrix}, \quad \text{with} \quad Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad R = 1.$$

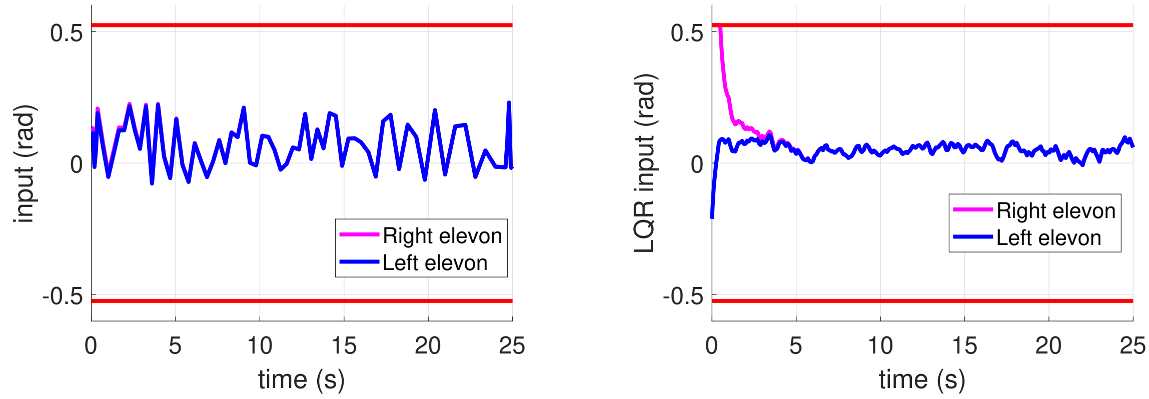
As predicted, the state converges exponentially from $x_0 = (1,\ 1,\ 1)$ rad/s to the origin, as shown by the blue curve in Figure 4.2(a). With the LQR feedback unaware of the undesirable input, the state does not converge to the origin, as shown in red in Figure 4.2(a). As can be seen on Figure 4.2(b), the undesirable input has a high variation and an amplitude non-negligible compared to the controlled inputs. It is not reaching its upper and lower bound because of the $\mathcal{L}_2$ normalization we operated.



(a) Distance of the state from the origin.

(b) Undesirable canard inputs.

Figure 4.2: State evolution with the two controllers for undesirable canard inputs.



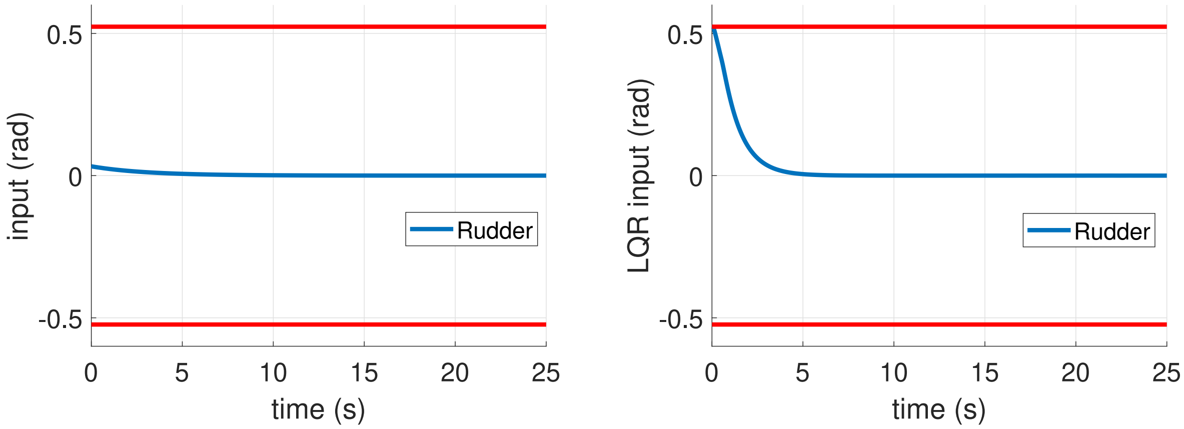
(a) Control law (4.10).

(b) LQR feedback control.

Figure 4.3: Right and left elevons inputs for the two controllers.

The control strategies employed by our two controllers are very different, as illustrated by the differences between Figure 4.3(a) and 4.3(b), and between Figure 4.4(a) and 4.4(b). The LQR input is initially saturated as can be seen on Figures 4.3(b) and 4.4(b).

If the pilot loses control authority over any one of the elevons, then F is not positive definite, but $BB^\top$ is invertible. The control law (4.19) is still well-defined, so it can be implemented, but for some $w \in \mathcal{F}(\mathcal{W})$ the control is not admissible: $u_w \notin \mathcal{F}(\mathcal{U})$. If the pilot loses control of the rudder, $BB^\top$ is not invertible, so the control law (4.19) is not well-defined. The jet cannot be guaranteed to be able to reach the desired target.



(a) Control law (4.10).

(b) LQR feedback control.

Figure 4.4: Rudder inputs for the two controllers.

4.6.2 Comparison with robust control

To illustrate the strength of our approach in the considered scenario, we compare our results with those of classical robust control.

Let us first recall the differences in assumptions between robust control and resilient reachability. A control law is said to be robust if it drives the state to the target *whatever the disturbance is*, i.e., there exists a control law u such that for all undesirable input w, we have $x(T) \in \mathcal{T}$. On the other hand, resilient reachability considers a controller *aware of the undesirable input*, i.e., for all w, there exists a control law u_w such that $x(T) \in \mathcal{T}$.

In our setting, the undesirable input is produced by an actuator belonging to the system. With sensors measuring the output of each actuators and a fault-detection mechanism, it is reasonable to assume that w can be measured. Then, the resilient controller has access to more information than a robust controller, and should perform better.

We choose the robust control approach developed in [42]. Its objective is to approximate the closed-loop reach set $\mathcal{X}[T]$ with internal and external ellipsoids. The reach set gathers the states $x_{goal} \in \mathbb{R}^n$ for each of which there exists a control law such that, whatever the undesirable input is, $x(T) \in \mathbb{B}(x_{goal}, \mu)$ for a certain radius $\mu \geq 0$. The comparison criteria between our approach and that of [42] will be the size of the smallest target ball guaranteed to be reached, i.e., μ. The application case is the ADMIRE model with drift studied in the previous subsection 4.6.1. We assume that the pilot loses control authority over the canards.

The resilient inputs have $\mathcal{L}_2$ bounds. However, the robust control inputs u must be bounded by an ellipsoid. To make the comparison as fair as possible, we choose the maximal ellipsoid within the actuators range (4.20).

The disturbance ellipsoid is $\mathcal{W} = \mathcal{E}(w_c, Q)$, with its center $w_c := \frac{1}{2}(w_{max} + w_{min})$. The disturbance bounds w_{min} and w_{max} are the mechanical bounds of the uncontrolled actuator defined in (4.20). We consider loss of control over only one actuator. Thus, Q is a scalar, so $Q(w - w_c)^2 \leq 1$ and $w_{min} \leq w \leq w_{max}$. Hence, $Q = \frac{4}{(w_{\max} - w_{\min})^2}$.

Defining the control ellipsoid is more complicated. To have a fair comparison with the results of our paper, we would need to enforce $\mathcal{L}_2$ bounds on the inputs. However, this is not possible in the framework of

[42]: it allows only for time-invariant ellipsoidal sets of admissible control inputs. Let us find a compromise. We start from the bounds $\|u\|_{\mathcal{L}_2} \leq 1$ and $\|w\|_{\mathcal{L}_2} \leq 1$. So, we want to enforce

$$\int_0^T \|u(t)\|^2 dt \leq 1 \qquad \text{and} \qquad \int_0^T \|w(t)\|^2 dt \leq 1,$$

which can be done by choosing $\|u(t)\|^2$, $\|w(t)\|^2 \leq \frac{1}{T}$ for all $t \in [0, T]$. What matters here is the fact that $\|u(t)\|$ and $\|w(t)\|$ have the same bound. Therefore, we choose to limit each input to the smallest of the two intervals $[w_{min}, w_{max}]$ and the interval from (4.20). The control ellipsoid is then $\mathcal{E}\big(u_c, P\big)$, with its center $u_c := \frac{1}{2}(u_{max} + u_{min})$ and a diagonal shape matrix P with $P_{ii} = \min\left\{\frac{2^2}{\left(u^i_{\max} - u^i_{\min}\right)^2}, Q\right\}$.

Now that we have defined our ellipsoidal input sets, we need to calculate the ellipsoidal bounds on the on the reachable set $\mathcal{E}\big(x_-(T), X_-(T)\big) \subseteq \mathcal{X}[T]$. The center $x_-(t)$ of each of the internal ellipsoid follows the dynamics

$$\dot{x}_- = Ax_- + Bu_c + Cw_c, \qquad x_-(0) = x_0 \in \mathbb{R}^n, \tag{4.21}$$

where u_c and w_c are the respective centers of the control ellipsoid and of the disturbance ellipsoid. The differential equation for the shape matrix $X_-(t)$ of the internal ellipsoid [42] is

$$\dot{X}_- = AX_- + X_-A^\top + \sqrt{X_-}S_1(t)B\sqrt{P} + \sqrt{P}B^\top S_1(t)\sqrt{X_-^\top} + \mu\Big(\sqrt{X_-}S_2(t) + S_2(t)\sqrt{X_-^\top}\Big) - \pi(t)X_- - \frac{CQC^\top}{\sqrt{\pi(t)}}, \tag{4.22}$$

with $X_-(0) = X^0$. The functions π, S_1 and S_2 are defined as follows for a given vector $l \in \mathbb{R}^n$:

$$\begin{aligned} l(t) &:= e^{A^\top t} l & \qquad \pi(t) &:= \sqrt{l(t)^\top CQC^\top l(t)} \\ S_1(t)B\sqrt{P} &:= \sqrt{\frac{l(t)^\top BPB^\top l(t)}{l(t)^\top X_- l(t)}}\sqrt{X_-} & \qquad S_2(t) &:= \frac{\|l(t)\|}{\sqrt{l(t)^\top X_- l(t)}}\sqrt{X_-}. \end{aligned}$$

We now need to calculate the radius μ of the smallest robustly reachable target. We compute only the tight ellipsoidal internal approximation of the closed-loop reachable set: $\mathcal{E}\big(x_-(T), X_-(T)\big) \subseteq \mathcal{X}[T]$. Indeed, if a state belongs to the internal ellipsoid it is guaranteed to be robustly reachable, because included in the reachable set. We compute the trajectory of the center of the ellipsoid $x_-(t)$ with (4.21), and the evolution of the shape matrix $X_-(t)$ of the ellipsoid with (4.22) and (4.23). When the radius μ of the target ball is too small for the target to be reached, then the shape matrix X_- is not positive definite. We investigated for the smallest μ such that $X_-(T) \succ 0$, and found $\mu = 5.9$. Therefore, the smallest target ball the robust method guarantees to reach has a radius of 5.9. The initial state $x_0 = (1,\ 1,\ 1)$ was already inside that ball. Thus, the robust control cannot even guarantee that the state will get closer to the target than its initial state.

On the other hand, we know that the jet is resilient to the loss of control over the canards. Therefore, a target ball of any size is resiliently reachable. By having access to the undesirable input, a controller ensuring resilient reachability is then more effective than a robust controller.

4.6.3 A driftless model

The aircraft model used as previous example is very convenient for our study because of the linearization and the overactuation. However, to render the dynamics driftless, we needed a more in-depth analysis of the model. We obtained the original simulation code of the ADMIRE model from [109].

For our purposes, we removed the states representing the sensor dynamics and those not directly affected by the controls from the initial 28-states model [106]. We also removed four of the sixteen inputs as they are negligible compared to the other inputs.

The simulation generates a pair of matrices A and $\bar{B}$ following the nominal dynamics (4.1). The effect of the matrix A is negligible compared to $\bar{B}$, when considering the states $x = (V_t,\ q,\ r)$, i.e., the jet speed, pitch and yaw rates. Thus, we approximate their dynamics by a driftless system, setting $A = 0$.

Since the jet has a single engine, it is not resilient to its loss. For our study, we assume a guaranteed authority over the thrust command, except for the afterburners. In the model the thrust command actuator also encompasses the afterburners. Since they account for only 20% of the thrust, the corresponding column in $\bar{B}$ is scaled by 20%. At Mach 0.75 and altitude 3000 m, the control matrix is

$$\bar{B}^\top = \begin{array}{cl} \left[\begin{array}{ccc} -2.7 & 7.1 & -1.9 \\ -2.7 & 7.1 & 1.9 \\ -1.0 & -7.7 & -1.1 \\ -1.8 & -13 & -3.0 \\ -1.8 & -13 & 3.0 \\ -1.0 & -7.7 & 1.1 \\ -1.9 & 0.0 & -11 \\ -0.8 & -0.5 & 0 \\ -4.3 & -0.7 & 0 \\ 1.2 & 0 & 0 \\ -71 & 1.2 & -710 \\ -113 & -882 & 0 \end{array}\right] & \begin{array}{l} \text{right canard,} \\ \text{left canard,} \\ \text{right outboard elevon,} \\ \text{right inboard elevon,} \\ \text{left inboard elevon,} \\ \text{left outboard elevon,} \\ \text{rudder,} \\ \text{leading edge flaps,} \\ \text{landing gear,} \\ \text{afterburner,} \\ \text{yaw thrust vectoring,} \\ \text{pitch thrust vectoring.} \end{array} \end{array}$$

Each row of $\bar{B}^\top$ represents the effect of the actuator written on the right. All the values of the inputs are in radians except for the landing gear and the afterburner which are between 0 and 1. This control matrix is not 1-resilient, because the thrust vectoring inputs are several orders of magnitude greater than any of the other inputs. For the same reason, the system is resilient to the loss of any one of the other ten actuators.

Simply removing thrust vectoring capabilities does not render the system 1-resilient; the control of the yaw rate would then primarily depend on the rudder, hence rendering the aircraft not resilient to the loss of the rudder.

Instead of removing the thrust vectoring actuators, if their range of motion is restricted to 1.4% of their current range, then $\bar{B}$ becomes resilient. Indeed, the thrust vectoring actuators can now be counteracted by the rudder and the elevons. Since we reduced the magnitude of two columns of $\bar{B}$, we also had to verify that the driftless hypothesis was still valid by comparing the effects of A and $\bar{B}$.

We showed how to make the fighter jet resilient in terms of speed, pitch and yaw rates, by scaling down thrust vectoring and having a guaranteed thrust. The resilience improvement by reducing the thrust vectoring might seem counterintuitive. Yet, it is explained by the fact that these actuators were too powerful to be balanced if they became uncontrolled. While the new system is resilient, its capabilities have been reduced. For instance, reaching a target (while undamaged) would take significantly more time for the new resilient system than for the old one.

The resilience analysis developed for this fighter jet is affected by several limitations of the current state of our theory. The first and obvious limitation comes from the driftless hypothesis but is justified here by the difference of magnitude between the drift and controlled dynamics. The most limiting hypothesis is that the controls are bounded by a $\mathcal{L}_2$ norm. Indeed, each actuator is independent of the others so a joint bound may not be appropriate. The structure of $\mathcal{U}$ with admissible inputs satisfying $\|u\|_{\mathcal{L}_2} \leq 1$ also assumes that each actuator has a symmetric range of functioning, which makes sense for the rudder, for instance, but not for the landing gear which can only be stored or deployed.

4.7 Summary

This chapter introduced the notion of resilient systems that can withstand the loss of control over any single or multiple actuators and still guarantee to drive the state to its target. We determined the minimal number of actuators required to design a 1-resilient system. We highlighted the increased complexity arising from higher degrees of resilience, which prevent calculations of the minimal overactuation required for p-resilience, $p \geq 2$. We then focused on the synthesis of a resilient control law for linear systems. Eventually, we illustrated our results on a fighter jet model.

Chapter 5

Quantitative Resilience of Generalized Integrators

5.1 Introduction

This chapter constitutes the first step of resilience theory for linear systems with component bounded inputs and relies on our works [32], [38]. After a partial loss of control authority over actuators, a target is *resiliently reachable* if for any undesirable inputs of the malfunctioning actuators there exists a control driving the state to the target [31]. However, the malfunctioning system might need considerably more time to reach its target compared to the initial system. To measure the delays caused by the loss of control authority, we will define the notion of quantitative resilience. Homonymous concepts have been previously developed for nuclear power plants [57], water infrastructure systems [56] and systems engineering [58] but were limited to their specific applications.

We formulate quantitative resilience as the maximal ratio over all targets of the minimal reach times for the initial and malfunctioning systems. This formulation leads to a nonlinear minimax optimization problem with an infinite number of constraints. Our main contribution is to reduce the quantitative resilience of systems with multiple integrators to a linear optimization problem. To do so we combine two optimization results designed specifically for this application [36] and established in Chapter 6 with the theorems of [72], [78] stating the existence of time-optimal controls.

The contributions of this chapter are twofold. First, we propose an efficient method to compute the quantitative resilience of linear systems with multiple integrators and nonsymmetric inputs by simplifying a nonlinear problem of four nested optimizations into a single linear optimization problem. Second, we establish necessary and sufficient conditions to verify if a system is resilient to the loss of control over one of its actuators.

This chapter is organized as follows. Section 5.2 introduces preliminary notions on resilience. We calculate the optimal reach times for the initial and malfunctioning systems in Section 5.3. The pinnacle of this work is the efficient method to compute quantitative resilience in Section 5.4. This metric also allows to assess whether a system is resilient, as detailed in Section 5.5. We study the quantitative resilience of systems with multiple integrators in Section 5.6 before applying our theory to an octocopter in Section 5.7.

5.2 Preliminaries and problem statement

The control of a physical system usually involves steering its position with inputs only affecting its acceleration [110]. With these systems in mind, we focus on generalized k^{th} order integrators in $\mathbb{R}^n$, i.e.,

$$x^{(k)}(t) = \bar{B}\bar{u}(t), \qquad \bar{u}(t) \in \bar{\mathcal{U}}, \quad x(0) = x_0, \quad x^{(l)}(0) = 0, \tag{5.1}$$

for all $l \in [\![1, k-1]\!]$ and $k \in \mathbb{N}$. Matrix $\bar{B} \in \mathbb{R}^{n \times (m+p)}$ is constant. The control set is the hyperrectangle $\bar{\mathcal{U}} := \prod_{i=1}^{m+p} \left[\bar{u}_i^{min}, \bar{u}_i^{max}\right] \subseteq \mathbb{R}^{m+p}$, with $\bar{u} \in \mathcal{F}(\bar{\mathcal{U}})$.

After a malfunction, the system loses control authority over p of its $m+p$ actuators. We then split $\bar{B}$ into B and C, $\bar{\mathcal{U}}$ into $\mathcal{U}$ and $\mathcal{W}$, and $\bar{u}$ into the remaining control inputs $u \in \mathcal{F}(\mathcal{U})$ and the undesirable inputs $w \in \mathcal{F}(\mathcal{W})$. Then, the initial conditions are the same as in (5.1) but the dynamics become

$$x^{(k)}(t) = Bu(t) + Cw(t), \quad u(t) \in \mathcal{U} := \prod_{i=1}^{m} \left[u_i^{min}, u_i^{max}\right], \quad w(t) \in \mathcal{W} := \prod_{i=1}^{p} \left[w_i^{min}, w_i^{max}\right]. \tag{5.2}$$

We now recall the definition of *resilience* from Chapter 4.

Definition 6: System (5.1) is *resilient* to the loss of p of its actuators corresponding to the matrix C as above if for all undesirable inputs $w \in \mathcal{F}(\mathcal{W})$ and all target $x_{goal} \in \mathbb{R}^n$ there exists a control $u_w \in \mathcal{F}(\mathcal{U})$ and a time T such that the state of the system (5.2) reaches the target at time T, i.e., $x(T) = x_{goal}$.

While a resilient system is by definition capable of reaching any target after a partial loss of control authority, the malfunctioning system might be considerably slower than the initial system to reach the same target. We introduce the following two reach times for the target $x_{goal} \in \mathbb{R}^n$ and the target distance $d := x_{goal} - x_0 \in \mathbb{R}^n$.

Definition 7: The *nominal reach time of order k* $T_{k,N}^*$, is the shortest time required for the state x of (5.1) to reach the target x_{goal} under admissible control $\bar{u} \in \mathcal{F}(\bar{\mathcal{U}})$:

$$T_{k,N}^*(d) := \inf_{\bar{u} \,\in\, \mathcal{F}(\bar{\mathcal{U}})} \big\{T \geq 0 : x(T) - x_0 = d\big\}. \tag{5.3}$$

Definition 8: The *malfunctioning reach time of order k* $T_{k,M}^*$, is the shortest time required for the state x of (5.2) to reach the target x_{goal} under admissible control $u \in \mathcal{F}(\mathcal{U})$ when the undesirable input $w \in \mathcal{F}(\mathcal{W})$ is chosen to make that time the longest:

$$T_{k,M}^*(d) := \sup_{w \,\in\, \mathcal{F}(\mathcal{W})} \Big\{ \inf_{u \,\in\, \mathcal{F}(\mathcal{U})} \big\{T \geq 0 : x(T) - x_0 = d\big\}\Big\}. \tag{5.4}$$

The causality issue arising from (5.4) is discussed at the end of this section. By definition, if the system is controllable, then $T_{k,N}^*(d)$ is finite for all $d \in \mathbb{R}^n$, and if it is resilient, then $T_{k,M}^*(d)$ is also finite. The malfunctioning system (5.2) can take up to $\frac{T_{k,M}^*(d)}{T_{k,N}^*(d)}$ times longer than the initial system (5.1) to reach the target $d + x_0$.

Definition 9: The *quantitative resilience of order k* of system (5.2) is

$$r_{k,q} := \inf_{d \,\in\, \mathbb{R}_*^n} \frac{T_{k,N}^*(d)}{T_{k,M}^*(d)}. \tag{5.5}$$

For a resilient system, $r_{k,q} \in (0, 1]$. The closer $r_{k,q}$ is to 1, the smaller is the loss of performance caused by the malfunction. We are now led to our problem of interest.

Problem 9: How to calculate efficiently $r_{k,q}$?

Indeed, a naive computation of $r_{k,q}$ requires solving four nested optimization problems whose constraint sets are $\mathbb{R}^n_*$ and three infinite-dimensional function spaces. A brute force approach to this problem is doomed to fail.

We will explore thoroughly the simple case $k = 1$ in the following sections and generalize their results to the case $k \in \mathbb{N}$ in Section 5.6. For $k = 1$, systems (5.1) and (5.2) simplify into

$$\dot{x}(t) = \bar{B}\bar{u}(t), \qquad \bar{u}(t) \in \bar{\mathcal{U}}, \quad x(0) = x_0 \in \mathbb{R}^n, \tag{5.6}$$

$$\dot{x}(t) = Bu(t) + Cw(t), \quad u(t) \in \mathcal{U}, \; w(t) \in \mathcal{W}. \tag{5.7}$$

For brevity, in the case $k = 1$ we lose the subscript 1 and write the *nominal reach time* $T_N^* = T_{1,N}^*$ as

$$T_N^*(d) := \inf_{\bar{u} \,\in\, \mathcal{F}(\bar{\mathcal{U}})} \Big\{ T \geq 0 : \int_0^T \bar{B}\bar{u}(t)\, dt = d \Big\}, \tag{5.8}$$

and the *malfunctioning reach time* $T_M^* = T_{1,M}^*$ as

$$T_M^*(d) := \sup_{w \,\in\, \mathcal{F}(\mathcal{W})} \left\{ \inf_{u \,\in\, \mathcal{F}(\mathcal{U})} \Big\{ T \geq 0 : \int_0^T \big[Bu(t) + Cw(t)\big] dt = d \Big\} \right\}. \tag{5.9}$$

The *ratio of reach times* is $t(d) := T_M^*(d)/T_N^*(d)$. The *quantitative resilience* r_q of a system following (5.7) is then

$$r_q := \frac{1}{\sup\limits_{d \,\in\, \mathbb{R}^n_*} t(d)} = \inf_{d \,\in\, \mathbb{R}^n_*} \frac{T_N^*(d)}{T_M^*(d)} = r_{1,q}. \tag{5.10}$$

We now discuss the information setting in the malfunctioning system. The resilience framework of this dissertation assumes that u has only access to the past and current values of w, but not to their future. Then, the optimal control u^* in (5.9) cannot anticipate a truly random undesirable input w. Hence, this strategy is not likely to result in the global time-optimal trajectory of Definition 8.

In fact, there would be no single obvious choice for $u^*\big(t, w(t)\big)$, rendering T_M^* ill-defined and certainly not time-optimal, whereas T_N^* is time-optimal. In this case, our concept of quantitative resilience becomes meaningless. The work [86] states that to calculate u^* without future knowledge of w^* the only technique is to solve the intractable Isaac's equation. Thus, the paper [86] derives only suboptimal solutions and concludes that its practical contribution is minimal.

Instead, we follow [30] where the inputs u^* and w^* are both chosen to make the transfer from x_0 to x_{goal} time-optimal in the sense of Definition 8. The controller knows that w^* will be chosen to make T_M^* the longest. Thus, u^* is chosen to react optimally to this worst undesirable input. Then, w^* is chosen, and to make T_M^* the longest, it is the same as the controller had predicted. Hence, from an outside perspective it looks as if the controller knew w^* in advance, as reflected by (5.4).

We will prove in the following sections that with this information setting w^* is constant. Then, the controller can more easily and more reasonably predict what is the worst w^* and build the adequate u^*. With these two input signals, T_M^* is time-optimal in the sense of Definition 8 and can be meaningfully compared with T_N^* to define the quantitative resilience of control systems.

5.3 Optimal reach times

We start with the dynamical system (5.6) to calculate the nominal reach time T_N^* of (5.8). We easily show in Lemma 1 of Appendix 5.8, that if system (5.6) is controllable, the optimal control $\bar{u}^*$ of (5.8) exists and is constant:

$$T_N^*(d) = \min_{\bar{u}\,\in\,\bar{\mathcal{U}}} \big\{T \geq 0 : \bar{B}\bar{u}\,T = d\big\}. \tag{5.11}$$

Since the input set $\bar{\mathcal{U}}$ is bounded, the controllability of system (5.6) is equivalent to $\mathrm{rank}(\bar{B}) = n$ and $0 \in \mathrm{int}(\bar{\mathcal{U}})$ [91]. The multiplication of variables $\bar{u}$ and T makes (5.11) a bilinear optimization problem. For easier computation, we solve instead the linear optimization $T_N^*(d) = 1/\max\limits_{\bar{u}\,\in\,\bar{\mathcal{U}}}\big\{\lambda : \bar{B}\bar{u} = \lambda d\big\}$.

We now study the malfunctioning system (5.7) to compute the malfunctioning reach time T_M^* of (5.9). As above, we easily prove in Lemma 2 of Appendix 5.8 that if system (5.7) is resilient, the optimal control $u^*(w)$ of (5.9) exists and is constant for any undesirable input $w \in \mathcal{F}(\mathcal{W})$:

$$T_M^*(d) = \sup_{w\,\in\,\mathcal{F}(\mathcal{W})} \left\{ \min_{u^*(w)\,\in\,\mathcal{U}} \left\{ T : Bu^*(w)T + \int_0^T Cw(t)\,dt = d \right\}\right\}. \tag{5.12}$$

Tackling the supremum in (5.12) requires a different approach.

Proposition 11: If system (5.7) is resilient, then for all $d \in \mathbb{R}^n_*$ the supremum $T_M^*(d)$ of (5.9) is a maximum achieved by a constant undesirable input $w^* \in \mathcal{W}$.

Proof. For $w \in \mathcal{F}(\mathcal{W})$, let

$$w^c := \int_0^{T_M(w,d)} \frac{w(t)}{T_M(w,d)}\,dt,$$

with T_M defined in (5.26). Then, for $i \in [\![1,p]\!]$ we have $w_i^{min} \leq w_i(t) \leq w_i^{max}$. Integrating yields $w_i^{min} \leq w_i^c \leq w_i^{max}$, so $w^c \in \mathcal{W}$. Then,

$$\int_0^{T_M(w,d)} Cw(t)dt = Cw^c T_M(w,d) = d - Bu^*(w)T_M(w,d).$$

Conversely, note that for all $w^c \in \mathcal{W}$ and $T > 0$, we can define $w(t) := \frac{1}{T}w^c$ for $t \in [0,T]$ such that $\int_0^T Cw(t)\,dt = Cw^c$ and $w \in \mathcal{F}(\mathcal{W})$. Thus, the constraint space of the supremum of (5.9) can be restricted to constant inputs in $\mathcal{W}$.

We define the function $\varphi(w) := Bu^*(w) + Cw$ for $w \in \mathcal{W}$. When applying the constant inputs w and $u^*(w)$, dynamics (5.7) become $\dot{x} = \varphi(w)$. Because $\big(Bu^*(w) + Cw\big)T_M(w,d) = d$, we have $\varphi(w) = \frac{1}{T_M(w,d)}d$ and φ is continuous in w according to Lemma 3 in Appendix 5.8. Set $\mathcal{W}$ is compact and $x_0 \in \mathbb{R}^n$ is fixed. Then, Theorem 1 of [72] states that $\mathcal{A}_{\mathcal{W}} := \big\{(x_1, T) : \int_0^T \varphi(w)dt = x_1 - x_0,\ \text{for } w \in \mathcal{W}\big\}$ is compact. Note that $T_M^*(d) = \sup\big\{T : (x_{goal}, T) \in \mathcal{A}_{\mathcal{W}}\big\}$ is the supremum of a continuous function over the compact set $\mathcal{A}_{\mathcal{W}}$, so $T_M^*(d)$ is a maximum achieved on $\mathcal{W}$. □

Then, the malfunctioning reach time becomes

$$T_M^*(d) = \max_{w\,\in\,\mathcal{W}} \left\{ \min_{u\,\in\,\mathcal{U}} \big\{T \geq 0 : \big(Bu + Cw\big)T = d\big\}\right\}. \tag{5.13}$$

We will show that the maximum of (5.13) is achieved by an extreme undesirable input, i.e., at the set of vertices of $\mathcal{W}$, denoted by $\mathcal{V}$. However, we cannot directly apply the bang-bang principle, as it has been mostly derived for systems with a linear dependency on the input [67], [78], [111], while φ introduced in Proposition 11 is not linear in w. The works [72], [112], [113] consider a nonlinear φ, but they require conditions that are not satisfied in our case. Thus, we need a new optimization result, namely Theorem 2.1 from [36], which applies to polytopes.

Definition 10: A *polytope* in $\mathbb{R}^n$ is a compact intersection of finitely many half-spaces.

We define $\mathcal{X} := \big\{Cw : w \in \mathcal{W}\big\}$ and $\mathcal{Y} := \big\{Bu : u \in \mathcal{U}\big\}$. Since $\mathcal{U}$ and $\mathcal{W}$ are polytopes, so are $\mathcal{X}$ and $\mathcal{Y}$ [114].

Proposition 12: If system (5.7) is resilient, then $\dim \mathcal{Y} = n$ and $-\mathcal{X} \subseteq \text{int}(\mathcal{Y})$.

Proof. Following Proposition 11 we know that for all $x \in \mathcal{X}$ and all $d_0 \in \mathbb{R}^n$ there exist $y \in \mathcal{Y}$ and $T \geq 0$ such that $(x+y)T = d_0$. Since d_0 can be freely chosen in $\mathbb{R}^n$, we must have $\dim \mathcal{Y} = n$.

Take $d_0 = x \in \mathcal{X}$, $x \neq 0$. Then, there exists $y_1 \in \mathcal{Y}$ and $T_1 > 0$ such that $(x+y_1)T_1 = x$. Hence, $\lambda_1 x \in \mathcal{Y}$ with $\lambda_1 := -1 + 1/T_1$. Now take $d_0 = -x$. Then, there exists $y_2 \in \mathcal{Y}$ and $T_2 > 0$ such that $(x+y_2)T_2 = -x$. Hence, $\lambda_2 x \in \mathcal{Y}$ with $\lambda_2 := -1 - 1/T_2$. Since $\lambda_2 \leq -1 \leq \lambda_1$ and $\mathcal{Y}$ is convex, we have $-x \in \mathcal{Y}$.

If $x = 0$, this process fails because we would get $T = 0$ when taking $d = 0$. Instead, take $d_0 \in \mathbb{S}$. Then there exist $T > 0$ and $y \in \mathcal{Y}$ such that $yT = d_0$. Repeating the same for $-d_0$ and using the convexity of $\mathcal{Y}$ as in the previous paragraph, we obtain $0 \in \mathcal{Y}$. Thus $-\mathcal{X} \subseteq \mathcal{Y}$.

Assume that there exists $-x_1 \in -\mathcal{X} \cap \partial\mathcal{Y}$. For $d = -x_1$, $T_M(x_1, -x_1) = \min\limits_{y \,\in\, \mathcal{Y}} \big\{T \geq 0 : (x_1 + y)T = -x_1\big\}$, with T_M introduced in (5.26). Since $T \geq 0$, the optimal y (called y^*) must make $x_1 + y$ positively collinear with $-x_1$. Thus, y^* is positively collinear with $-x_1$ and the largest it can be is $y^* = -x_1$ because $-x_1 \in \partial\mathcal{Y}$. Then, the constraint in $T_M(x_1, -x_1)$ is $0T = -x_1$. The lack of solution contradicts the resilience of the system. Thus, $-\mathcal{X} \cap \partial\mathcal{Y} = \varnothing$, i.e., $-\mathcal{X} \subseteq \text{int}(\mathcal{Y})$. □

We now prove that the maximum of (5.13) is achieved on $\mathcal{V}$.

Proposition 13: If system (5.7) is resilient, then for all $d \in \mathbb{R}^n_*$, the maximum of (5.13) is achieved with a constant input $w^* \in \mathcal{V}$.

Proof. Replacing $\frac{1}{T}$ by λ in (5.13) leads to $T_M^*(d) = 1/\min\limits_{x \,\in\, \mathcal{X}} \big\{\max\limits_{y \,\in\, \mathcal{Y}} \{\lambda > 0 : x + y = \lambda d\}\big\}$. Since $\lambda \geq 0$, we write $\lambda = |\lambda| = \|\lambda d\|/\|d\| = \|x+y\|/\|d\|$. Then,

$$T_M^*(d) = \frac{\|d\|}{\min\limits_{x \,\in\, \mathcal{X}} \big\{\max\limits_{y \,\in\, \mathcal{Y}} \big\{\|x+y\| : x + y \in \mathbb{R}^+ d\big\}\big\}}. \tag{5.14}$$

Following Proposition 12, we can apply Theorem 2.1 of [36] and conclude that the argument of the minimum in (5.14) is at a vertex x^* of $\mathcal{X}$. Since the transformation between $\mathcal{W}$ and $\mathcal{X}$ is linear, $x^* = Cv$ with $v \in \mathcal{V}$ a vertex of $\mathcal{W}$ [114]. Therefore, the maximum of (5.13) is achieved on $\mathcal{V}$. □

We have then reduced the outer constraint set of (5.9) from an infinite-dimensional function set $\mathcal{F}(\mathcal{W})$ to a finite set $\mathcal{V}$ of cardinality 2^p with p the number of malfunctioning actuators. Then,

$$T_M^*(d) = \max_{w \,\in\, \mathcal{V}} \left\{ \min_{u \,\in\, \mathcal{U}} \big\{T \geq 0 : \big(Bu + Cw\big)T = d\big\} \right\}. \tag{5.15}$$

Because u is chosen to counteract w and make $Bu + Cw$ collinear with $d \in \mathbb{R}^n$, while w is chosen freely in $\mathcal{W}$, the minimum of (5.15) cannot be restricted to the vertices of $\mathcal{U}$. We will now prove that both reach times are linear in the target distance.

Proposition 14: For any $d \in \mathbb{R}^n$ and $\lambda \geq 0$ we have $T_N^*(\lambda d) = \lambda\, T_N^*(d)$ and $T_M^*(\lambda d) = \lambda\, T_M^*(d)$.

Proof. The case $\lambda = 0$ is trivial since $T_N^*(0) = T_M^*(0) = 0$, so consider $\lambda > 0$. The nominal reach time for d is $T_N^*(d)$, so there exists $\bar{u}_d \in \bar{\mathcal{U}}$ such that $\bar{B}\bar{u}_d T_N^*(d) = d$. Then, $\bar{B}\,\bar{u}_d\,\lambda T_N^*(d) = \lambda d$. The optimality of $T_N^*(\lambda d)$ to reach λd leads to $T_N^*(\lambda d) \leq \lambda T_N^*(d)$.

Similarly, there exists $\bar{u}_{\lambda d} \in \bar{\mathcal{U}}$ such that $\bar{B}\bar{u}_{\lambda d} T_N^*(\lambda d) = \lambda d$, so $\bar{B}\,\bar{u}_{\lambda d}\,\frac{T_N^*(\lambda d)}{\lambda} = d$. The optimality of $T_N^*(d)$ to reach d yields $T_N^*(d) \leq \frac{T_N^*(\lambda d)}{\lambda}$. Thus, $\lambda T_N^*(d) \leq T_N^*(\lambda d)$.

A similar proof does not work for T_M^* because of the minimax structure of (5.15).

For $d \in \mathbb{R}^n_*$ and $w \in \mathcal{W}$, we define $x = Cw$ and $y^*(x,d) := \arg\min\limits_{y \,\in\, \mathcal{Y}} \big\{ T \geq 0 : (y+x)T = d \big\}$. Note that $Bu^*(w) + Cw = y^*(x,d) + x$, with u^* defined in Lemma 2. Then, with T_M introduced in (5.26), we have $\big(Bu^*(w) + Cw\big) T_M(w,d) = d$, i.e., $y^*(x,d) = \frac{1}{T_M(w,d)} d - x$. For $\lambda > 0$, we define $\alpha(\lambda) := \frac{\lambda}{T_M(w,\lambda d)} - \frac{1}{T_M(w,d)}$, such that $y^*(x,\lambda d) - y^*(x,d) = \alpha(\lambda) d$.

The polytope $\mathcal{Y}$ in $\mathbb{R}^n$ has a finite number of faces, so we can choose $d \in \mathbb{R}^n_*$ not collinear with any face of $\mathcal{Y}$. Since $\mathcal{Y}$ is convex, the ray $\big\{ y^*(x,d) + \alpha d : \alpha \in \mathbb{R} \big\}$ intersects with $\partial\mathcal{Y}$ at most twice. Since $y^*(x,d) \in \partial\mathcal{Y}$, one intersection happens at $\alpha = 0$. If there exists another intersection, it occurs for some $\alpha_0 \neq 0$. Since $y^*(x,\lambda d) \in \partial\mathcal{Y}$, we have $y^*(x,d) + \alpha(\lambda) d \in \partial\mathcal{Y}$. Then, $\alpha(\lambda) \in \{0, \alpha_0\}$ for all $\lambda > 0$.

According to Lemma 3, T_M is continuous in d, so α is continuous in λ but its codomain is finite. Therefore, α is constant and we know that $\alpha(1) = 0$. So α is null for all $\lambda > 0$, leading to $T_M(w,\lambda d) = \lambda T_M(w,d)$ for all $\lambda > 0$ and d not collinear with any face of $\partial\mathcal{Y}$. Since the dimension of each face of $\partial\mathcal{Y}$ is at most $n-1$ in $\mathbb{R}^n$ and T_M is continuous in d, the homogeneity of T_M holds on the whole of $\mathbb{R}^n$. Note that $T_M^*(d) = \max\limits_{w \,\in\, \mathcal{W}} T_M(w,d)$. Thus, $\lambda T_M^*(d) = T_M^*(\lambda d)$. □

Combining the results obtained for the nominal and the malfunctioning dynamics, we can now evaluate the quantitative resilience of the system.

5.4 Quantitative resilience

Focusing on the loss of control over a single actuator we will simplify tremendously the computation of r_q by noting that the effects of the undesirable inputs are the strongest along the direction described by the malfunctioning actuator.

Theorem 11: If system (5.7) is resilient and C is a single column matrix, the ratio of reach times $t(d)$ is maximizing along C, i.e., $\max\limits_{d \,\in\, \mathbb{R}^n_*} t(d) = \max\big\{ t(C), t(-C) \big\}$.

Proof. Using Proposition 14 we reduce the constraint set of (5.10) from $\mathbb{R}^n_*$ to $\mathbb{S}$. We use the same process that yielded (5.14) but we start from (5.11) where we split $\bar{B}$ into B and C:

$$\frac{1}{T_N^*(d)} = \max_{\bar{u} \,\in\, \bar{\mathcal{U}}} \big\{ \lambda : \bar{B}\bar{u} = \lambda d \big\} = \max_{u \,\in\, \mathcal{U},\, w \,\in\, \mathcal{W}} \big\{ \lambda : Bu + Cw = \lambda d \big\} = \max_{x \,\in\, \mathcal{X},\, y \,\in\, \mathcal{Y}} \big\{ \|y + x\| : y + x \in \mathbb{R}^+ d \big\}. \quad (5.16)$$

We can now gather (5.14) with $d \in \mathbb{S}$ and (5.16) into

$$t(d) = \frac{T_M^*(d)}{T_N^*(d)} = \frac{\max\limits_{x \,\in\, \mathcal{X},\, y \,\in\, \mathcal{Y}} \big\{ \|x + y\| : x + y \in \mathbb{R}^+ d \big\}}{\min\limits_{x \,\in\, \mathcal{X}} \Big\{ \max\limits_{y \,\in\, \mathcal{Y}} \big\{ \|x + y\| : x + y \in \mathbb{R}^+ d \big\} \Big\}}.$$

Because C is a single column, $\dim \mathcal{X} = 1$. Then, following Proposition 12 we can apply the Maximax Minimax Quotient Theorem of [36] that states $\max\limits_{d \,\in\, \mathbb{S}} t(d) = \max\big\{t(C), t(-C)\big\}$. □

Theorem 11 is the strongest result of this work as it solves the nonlinear fractional optimization of r_q over $d \in \mathbb{S}$. Its proof is brief because all the heavy lifting is done in [36].

Since the sets $\mathcal{U}$ and $\mathcal{W}$ are not symmetric, in general $t(C) \neq t(-C)$. Thus, to calculate the quantitative resilience r_q we need to evaluate $T_N^*(\pm C)$ and $T_M^*(\pm C)$, i.e., solve four optimization problems. The computation load can be halved with the following result.

Theorem 12: If system (5.7) is resilient and C is a single nonzero column, then we have $r_q = r_{min}$, with $r_{min} := \min\big\{r(C), r(-C)\big\}$, $r(C) := \frac{w^{min}+\lambda^+}{w^{max}+\lambda^+}$, $r(-C) := \frac{w^{max}-\lambda^-}{w^{min}-\lambda^-}$ and $\lambda^{\pm} := \max\limits_{v \,\in\, \mathcal{U}}\big\{\lambda : Bv = \pm\lambda C\big\}$.

Proof. Let $\bar{u} \in \bar{\mathcal{U}}$, $u \in \mathcal{U}$ and $w \in \mathcal{W}$ be the arguments of the optimization problems (5.11) and (5.15) for $d = C \neq 0$. We write $\bar{u} = (u_B,\ u_C) \in \mathcal{U} \times \mathcal{W}$. Then,

$$\bar{B}\bar{u}\, T_N^*(C) = Bu_B\, T_N^*(C) + Cu_C\, T_N^*(C) = C, \qquad \text{and} \qquad Bu\, T_M^*(C) + Cw\, T_M^*(C) = C. \tag{5.17}$$

We consider the loss of a single actuator, thus $\mathcal{W} = [w^{min}, w^{max}] \subseteq \mathbb{R}$ which makes $CwT_M^*(C)$ and $Cu_CT_N^*(C)$ collinear with C. From Proposition 13, we know that $w \in \partial\mathcal{W}$. Since w maximizes $T_M^*(C)$ in (5.17), we obviously have $w = w^{min}$. On the contrary, u_C is chosen to minimize $T_N^*(C)$ in (5.17), so $u_C = w^{max}$.

According to (5.17), Bu_B and Bu are collinear with C, and they are chosen to minimize respectively $T_N^*(C)$ and $T_M^*(C)$. Thus, u and u_B are the vectors in $\mathcal{U}$ that maximize the norm of Bu and Bu_B and make them positively collinear with C, i.e., $u = u_B = \arg\min\limits_{v \,\in\, \mathcal{U}}\big\{\tau : Bv\tau = C\big\}$. Using $\lambda = \frac{1}{\tau}$ we render this problem linear:

$$\lambda^+ = \max_{v \,\in\, \mathcal{U}}\big\{\lambda : Bv = \lambda C\big\}, \qquad \text{and} \qquad u = u_B = \arg\max_{v \,\in\, \mathcal{U}}\big\{\lambda : Bv = \lambda C\big\}.$$

By combining all the results, (5.17) simplifies into:

$$C(\lambda^+ + w^{max})T_N^*(C) = C, \qquad \text{and} \qquad C(\lambda^+ + w^{min})T_M^*(C) = C.$$

Since C is a nonzero column,

$$\frac{T_N^*(C)}{T_M^*(C)} = \frac{\lambda^+ + w^{min}}{\lambda^+ + w^{max}} = r(C).$$

Following the same reasoning for $d = -C$, we obtain

$$C(-\lambda^- + w^{min})T_N^*(C) = -C \qquad \text{and} \qquad C(-\lambda^- + w^{max})T_M^*(C) = -C,$$

with $\lambda^- = \max\limits_{v \,\in\, \mathcal{U}}\big\{\lambda : Bv = -\lambda C\big\}$. Then,

$$\frac{1}{t(-C)} = \frac{T_N^*(-C)}{T_M^*(-C)} = \frac{w^{max} - \lambda^-}{w^{min} - \lambda^-} = r(-C).$$

Following Theorem 11,

$$r_q = \frac{1}{\max\{t(d)\,:\, d \,\in\, \mathbb{S}\}} = \min\left\{\frac{1}{t(C)}, \frac{1}{t(-C)}\right\} = \min\big\{r(C), r(-C)\big\} = r_{min}.$$

□

We introduced quantitative resilience as the solution of four nonlinear nested optimization problems and with Theorem 12 we reduced r_q to the solution of two linear optimization problem. We can thus quickly calculate the maximal delay caused by the loss of control of a given actuator.

5.5 Resilience conditions

So far, all our results need the system to be resilient. However, we know that verifying the resilience of a system with inputs of finite energy is not an easy task [35], and thus we can assume it is not trivial either with our component bounded inputs.

Proposition 15: A system following (5.6) is resilient to the loss of control over a column C if and only if it is controllable and both $T_M^*(C)$ and $T_M^*(-C)$ are finite.

Proof. If system (5.6) is resilient, then it is controllable a fortiori and Proposition 11 yields $T_M^*(C)$ and $T_M^*(-C)$ are finite.

On the other hand, assume that system (5.6) is controllable and $\max\big\{T_M^*(C), T_M^*(-C)\big\}$ is finite. Let $w \in \mathcal{W}$ and $d \in \mathbb{R}^n_*$. By controllability of system (5.6), there exists $\bar{u} \in \bar{\mathcal{U}}$ and $\lambda > 0$ such that $\bar{B}\bar{u} = \lambda d$. We split $\bar{B}$ into B and C, and $\bar{u}$ into u_d and w_d. Then, $u_d \in \mathcal{U}$ and $\bar{B}\bar{u} = Bu_d + Cw_d = \lambda d$. In the case $C = 0$, this equation yields $Bu_d = \lambda d = Bu_d + Cw$, so the system is resilient.

For $C \neq 0$, we will first show that for any $w \in \mathcal{W}$ we can find $u \in \mathcal{U}$ such that $Bu + Cw = 0$. Because $T_M^*(C)$ and $T_M^*(-C)$ are finite, $T_M(w, \pm C)$ is positive and finite for all $w \in \mathcal{W} = [w^{min}, w^{max}]$, with $T_M(\cdot,\cdot)$ defined in (5.26). Take $w \in \mathcal{W}$. Then, there exist $u_+^w \in \mathcal{U}$ and $u_-^w \in \mathcal{U}$ such that

$$\big(Bu_+^w + Cw\big)T_M(w, C) = C \qquad \text{and} \qquad \big(Bu_-^w + Cw\big)T_M(w, -C) = -C.$$

Define

$$\alpha := \frac{T_M(w, C)}{T_M(w, C) + T_M(w, -C)} \in (0, 1) \qquad \text{and} \qquad u := \alpha u_+^w + (1-\alpha)u_-^w.$$

Then, $u \in \mathcal{U}$ because $\mathcal{U}$ is convex. Notice that

$$\begin{aligned} Bu + Cw &= \alpha\big(Bu_+^w + Cw\big) + (1-\alpha)\big(Bu_-^w + Cw\big) \\ &= \frac{T_M(w, C)}{T_M(w, C) + T_M(w, -C)} \frac{C}{T_M(w, C)} + \frac{T_M(w, -C)}{T_M(w, C) + T_M(w, -C)} \frac{-C}{T_M(w, -C)} = 0. \end{aligned}$$

We want to make a convex combination of u and u_d to build the desired control. If $w \in \partial\mathcal{W}$ the resulting control will not be stronger than the adversary. So, we need to show that even if w is a little bit outside of $\mathcal{W}$ we can still counteract it. Let

$$\varepsilon := \min\left(\frac{1}{2T_M(w^{min}, C)}, \frac{1}{2T_M(w^{max}, -C)}\right) > 0.$$

Now take $w' \in (w^{max}, w^{max} + \varepsilon]$. There exists $u_- \in \mathcal{U}$ and $u_+ \in \mathcal{U}$ such that

$$\big(Bu_+ + Cw^{max}\big)T_M(w^{max}, C) = C \qquad \text{and} \qquad \big(Bu_- + Cw^{max}\big)T_M(w^{max}, -C) = -C.$$

Then, we can define $T^+ > 0$ such that

$$Bu_+ + Cw' = Bu_+ + Cw^{max} + C(w' - w^{max}) = C\left(\frac{1}{T_M(w^{max}, C)} + w' - w^{max}\right) = \frac{C}{T^+}.$$

Since $w' - w^{max} \leq \frac{1}{2T_M(w^{max}, -C)}$, we can similarly define $T^- > 0$ such that

$$Bu_- + Cw' = -C\bigg(\frac{1}{T_M(w^{max}, -C)} - (w' - w^{max})\bigg) = \frac{-C}{T^-}.$$

We take $\alpha = \frac{T^+}{T^+ + T^-} \in (0,1)$ which yields $u' = \alpha u_+ + (1-\alpha)u_- \in \mathcal{U}$ by convexity. Then, $Bu' + Cw' = 0$. With a similar approach we can build another u' to counteract any $w' \in [w^{min} - \varepsilon, w^{min})$.

Since $\mathcal{W}$ is convex, $w \in \mathcal{W}$ and $w_d \in \mathcal{W}$, we can take $w' \in [w^{min} - \varepsilon, w^{max} + \varepsilon]$ such that there exists $\gamma \in (0,1)$ for which $w = \gamma w_d + (1-\gamma)w'$. We build $u' \in \mathcal{U}$ as above to make $Bu' + Cw' = 0$. By convexity of $\mathcal{U}$, $u := \gamma u_d + (1-\gamma)u' \in \mathcal{U}$. Then,

$$Bu + Cw = \gamma\big(Bu_d + Cw_d\big) + (1-\gamma)\big(Bu' + Cw'\big) = \gamma\lambda d.$$

Since $\gamma > 0$, we have $\gamma\lambda > 0$ making the system resilient to the loss of column C. □

The intuition behind Proposition 15 is that a resilient system has two properties: the ability to reach any state prior to a malfunction, i.e., controllability, and the ability to do so after the malfunction despite the worst undesirable inputs, i.e., $T_M^*(\pm C)$ is finite. We can now derive resilience from a computation, making it easier to verify.

Corollary 3: System (5.6) is resilient to the loss of control over a nonzero column C if and only if it is controllable, and $r(C)$ and $r(-C)$ from Theorem 12 are in $(0,1]$.

Proof. If $C = 0$, the controllability is equivalent to resilience and $r(0) = 1$. If $C \neq 0$ and system (5.6) is resilient, then by Proposition 15, both $T_M^*(\pm C)$ are finite and system (5.6) is controllable, so both $T_N^*(\pm C)$ are finite too. Trivially $T_N^* \leq T_M^*$, so we have both $r(C) = \frac{T_N^*(C)}{T_M^*(C)} \in (0,1]$ and $r(-C) = \frac{T_N^*(-C)}{T_M^*(-C)} \in (0,1]$ according to Theorem 12.

On the other hand, assume that the system is controllable and that $\frac{w^{min}+\lambda^+}{w^{max}+\lambda^+}$ and $\frac{w^{max}-\lambda^-}{w^{min}-\lambda^-} \in (0,1]$. If $w^{min} + \lambda^+ < 0$, then $w^{max} + \lambda^+ \leq w^{min} + \lambda^+$ because $r(C) \in (0,1]$. This leads to the impossible conclusion that $w^{max} \leq w^{min}$. If $w^{min} + \lambda^+ = 0$, then $r(C) = 0$. Therefore, $w^{min} + \lambda^+ > 0$. Let $u \in \mathcal{U}$ such that $Bu = \lambda^+ C$. For $w \in \mathcal{W}$, we define $T_w := \frac{1}{w + \lambda^+}$, so that $(Bu + Cw)T_w = C$. Note that T_w is positive and finite because $w + \lambda^+ \geq w^{min} + \lambda^+ > 0$. Since

$$T_M^*(C) \leq \max_{w \,\in\, \mathcal{W}} T_w = \frac{1}{w^{min} + \lambda^+},$$

$T_M^*(C)$ is finite.

The same reasoning holds for $r(-C)$. We can show that $w^{max} - \lambda^- < 0$ and that $T_w := \frac{1}{\lambda^- - w} > 0$ for all $w \in \mathcal{W}$. With $u \in \mathcal{U}$ such that $Bu = -\lambda^- C$ we have $(Bu + Cw)T_w = -C$. Then,

$$T_M^*(-C) \leq \max_{w \,\in\, \mathcal{W}} T_w = \frac{1}{\lambda^- - w^{max}},$$

so $T_M^*(-C)$ is finite. Then, Proposition 15 states that the system is resilient. □

We now have all the tools to assess the quantitative resilience of a driftless system. If $\operatorname{rank}(\bar{B}) = n$ and $0 \in \operatorname{int}(\bar{\mathcal{U}})$, then system (5.6) is controllable [91]. After computing the ratios $r(\pm C)$, Corollary 3 states whether the system is resilient. If it is, then $r_q = r_{min}$ by Theorem 12, otherwise $r_q = 0$. We summarize this process in Algorithm 1.

```
Algorithm 1: Resilience algorithm for system (5.6)
Data: A column $C$ of $\bar{B}$, $r(C)$ and $r(-C)$ from Theorem 12
Result: $r_q$
if $\operatorname{rank}(\bar{B}) = n$ and $0 \in \operatorname{int}(\bar{\mathcal{U}})$ then
                    # system (5.6) is controllable
    if $r_C^+ \in (0,1]$ and $r_C^- \in (0,1]$ then
        $r_q = \min\{r_C^+, r_C^-\}$ # resilient to loss of $C$
    else
        $r_q = 0$        # not resilient to loss of $C$
    end
else
    $r_q = 0$            # not resilient to any loss
end
```

5.6 Systems with multiple integrators

We can now extend the results obtained for driftless systems to generalized higher-order integrators.

Proposition 16: If system (5.6) is controllable, then the infimum of (5.3) is achieved with the same constant control input $\bar{u}^* \in \bar{\mathcal{U}}$ as T_N^* in (5.8), and $T_{k,N}^*(d) = \sqrt[k]{k!\, T_N^*(d)}$ for all $d \in \mathbb{R}^n$.

Proof. If $d = 0$, then $T_{k,N}^*(d) = 0 = T_N^*(d)$, so the result holds. Let $d \neq 0$. By assumption, system $\dot{y}(t) = \bar{B}\bar{u}(t)$ with $y(0) = 0$ is controllable. Following Lemma 1 there exists a constant optimal control $\bar{u} \in \bar{\mathcal{U}}$ such that $y\big(T_N^*(d)\big) - y(0) = d = \bar{B}\bar{u}T_N^*(d)$, with $T_N^*(d) > 0$. Then, applying the control input $\bar{u}$ to (5.1) on the time interval $[0,\, t_1]$ leads to

$$x(t_1) - x_0 = \int_0^{t_1}\!\!\int_0^{t_2}\!\!\dots\int_0^{t_k} x^{(k)}(t_{k+1})\, dt_{k+1}\dots dt_2 = \int_0^{t_1}\!\!\int_0^{t_2}\!\!\dots\int_0^{t_k} \bar{B}\bar{u}\, dt_{k+1}\dots dt_2 = \bar{B}\bar{u}\frac{t_1^k}{k!} = \frac{d}{T_N^*(d)}\frac{t_1^k}{k!},$$

since $x^{(l)}(0) = 0$ for $l \in [\![1, k-1]\!]$ and $\bar{B}\bar{u} = \frac{d}{T_N^*(d)} \in \mathbb{R}^n$ is constant. By taking $t_1 = \sqrt[k]{k!\, T_N^*(d)}$, we obtain $x(t_1) - x_0 = d$. Thus, the state x_{goal} is reachable in finite time t_1, so the system (5.1) is controllable and $T_{k,N}^*(d) \leq t_1$.

Assume for contradiction purposes that there exists $\tilde{u} \in \bar{\mathcal{U}}$ such that the state of (5.1) can reach x_{goal} in a time $\tau < t_1$. Since $\tilde{u}$ can be time-varying, we build

$$\hat{u} := \frac{k!}{\tau^k}\int_0^{\tau}\dots\int_0^{t_k} \tilde{u}(t_{k+1})\, dt_{k+1}\dots dt_2.$$

Since $\tilde{u} \in \bar{\mathcal{U}}$, $\tilde{u}_i(t) \in [\bar{u}_i^{min}, \bar{u}_i^{max}]$ for all $i \in [\![1, m+p]\!]$ and $t \in [0, \tau]$. Because $\bar{u}_i^{min}$ and $\bar{u}_i^{max}$ are constant, one can easily obtain through k successive integrations that $\hat{u}_i \in [\bar{u}_i^{min}, \bar{u}_i^{max}]$ for all $i \in [\![1, m+p]\!]$. Thus, $\hat{u}$

is an admissible constant control input. Then, we apply $\tilde{u}$ to (5.1) on the time interval $[0,\tau]$ and we obtain

$$x(\tau) - x_0 = d = \int_0^\tau \dots \int_0^{t_k} \bar{B}\tilde{u}(t_{k+1})\, dt_{k+1} \dots dt_2 = \bar{B}\hat{u}\frac{\tau^k}{k!},$$

so $\bar{B}\hat{u} = \frac{k!d}{\tau^k}$. Applying the control input $\hat{u}$ to the system $\dot{y}(t) = \bar{B}\bar{u}(t)$ on the interval $[0,T]$ with $T := \frac{\tau^k}{k!}$ leads to

$$y(T) = \int_0^T \dot{y}(t)\, dt = \int_0^T \bar{B}\hat{u}\, dt = \bar{B}\hat{u}T = \frac{k!d}{\tau^k}\frac{\tau^k}{k!} = d.$$

Thus, y can reach d in a time $T = \frac{\tau^k}{k!} < \frac{t_1^k}{k!} = T_N^*(d)$, which contradicts the optimality of $T_N^*(d)$. In other words, t_1 is the minimal time for the state of (5.1) to reach x_{goal}. Therefore, the infimum of (5.3) is achieved with the same constant input $\bar{u} \in \bar{\mathcal{U}}$ as $T_N^*(d)$ in (5.8), and $T_{k,N}^*(d) = \sqrt[k]{k!\ T_N^*(d)}$. □

Remark: The proof of Proposition 16 went smoothly because the initial condition had zero derivatives. We will study a simple case with non-zero initial condition and show that the calculation of $T_{k,N}^*$ can be difficult, if not impossible. Let $k = 2$ and denote $v := \dot{x}$. Assume that system $\dot{v} = \bar{B}\bar{u}$ is controllable, and $\dot{x}(0) = v(0) = v_0 \neq 0$. For any $v_1 \in \mathbb{R}^n_*$ there exists an optimal control $\bar{u} \in \bar{\mathcal{U}}$ such that $v\big(T_N^*(v_1)\big) - v_0 = v_1 = \bar{B}\bar{u}T_N^*(v_1)$, with $T_N^*(v_1) > 0$, since $v_1 \neq 0$. Then,

$$v(t) - v_0 = \int_0^T \bar{B}\bar{u}\, d\tau = \int_0^T \frac{v_1}{T_N^*(v_1)}\, d\tau = \frac{tv_1}{T_N^*(v_1)},$$

$$x(T) - x_0 = \int_0^T \frac{tv_1}{T_N^*(v_1)} + v_0\, dt = \frac{T^2 v_1}{2T_N^*(v_1)} + v_0 T.$$

Taking $x(T) = x_{goal}$ leads to a quadratic polynomial in $\mathbb{R}^n$: $\frac{v_1}{2T_N^*(v_1)}T^2 + v_0T - d = 0$. These are n scalar equations for $n+1$ unknowns: v_1 and T. Because $T_N^*(v_1)$ depends on v_1, the equations are not independent and thus might not have a solution. In this case, the optimal control input is not constant, making $T_{2,N}^*$ much harder to obtain, even for this seemingly simple case. Time-varying inputs also ruin the geometric approach of Theorem 11, preventing to solve the fractional optimization of r_q over $d \in \mathbb{S}$.

A result similar to Proposition 16 holds for the malfunctioning reach time of order k.

Proposition 17: If system (5.7) is resilient, then system (5.2) is resilient for all $k \in \mathbb{N}$. The supremum and infimum of (5.4) are achieved with the same constant inputs $u^* \in \mathcal{U}$ and $w^* \in \mathcal{W}$ as T_M^* in (5.9), and $T_{k,M}^*(d) = \sqrt[k]{k!\ T_M^*(d)}$ for $d \in \mathbb{R}^n$.

Proof. We use the same calculations as in Proposition 16 but with $Bu^*(w) + Cw$ instead of $\bar{B}\bar{u}$ and $T_M(w,d)$ instead of $T_N^*(d)$. Then, u^* from Lemma 2 produces the best control input $u^*(w)$ for any $w \in \mathcal{W}$ for system (5.2).

We go again through the proof of Proposition 16, but this time we use $Bu^*(w^*) + Cw^*$ and $T_M^*(d)$. We conclude that $T_{k,M}^*(d) = \sqrt[k]{k!\ T_M^*(d)}$ and that w^* from Proposition 11 is also the worst undesirable input for system (5.2). □

We can now evaluate the quantitative resilience of order k.

Theorem 13: If system (5.6) is resilient, then for all $k \in \mathbb{N}$ system (5.1) is resilient and $r_{k,q} = \sqrt[k]{r_q}$.

Proof. Based on Propositions 16 and 17, $\frac{T_{k,M}^*(d)}{T_{k,N}^*(d)} = \frac{\sqrt[k]{k!\ T_M^*(d)}}{\sqrt[k]{k!\ T_N^*(d)}} = \sqrt[k]{\frac{T_M^*(d)}{T_N^*(d)}}$, so $r_{k,q} = \sqrt[k]{r_q}$. □

For a resilient system $r_q \in (0,1]$, then $r_{k,q} \geq r_q$. Thus, adding integrators to a resilient system increases its quantitative resilience. By studying $\dot{x}(t) = \bar{B}\bar{u}(t)$ we can then calculate the quantitative resilience of any system of the form $x^{(k)}(t) = \bar{B}\bar{u}(t)$ for $k \in \mathbb{N}$. We will now apply our theory to two numerical examples.

5.7 Numerical examples

Our first example considers a linearized model of a low-thrust spacecraft performing orbital maneuvers. We study the resilience of the spacecraft with respect to the loss of control over some thrust frequencies. Our second example features an octocopter UAV (Unmanned Aerial Vehicle) enduring a loss of control authority over some of its propellers.

5.7.1 Linear quadratic trajectory dynamics

We study a low-thrust spacecraft in orbit around a celestial body. Because of the complexity of nonlinear low-thrust dynamics the work [115] established a linear model for the spacecraft dynamics using Fourier thrust acceleration components. Given an initial state and a target state, the model simulates the trajectory of the spacecraft in different orbit maneuvers, such as an orbit raising or a plane change. The states of this linear model are the orbital elements $x := \big(a,\ e,\ i,\ \Omega,\ \omega,\ M\big)$ whose names are listed in Table 5.1.

Because of the periodic motion of the spacecraft, the thrust acceleration vector F can be expressed in terms of its Fourier coefficients α and β:

$$F = F_R\hat{r} + F_W\hat{w} + F_S(\hat{w} \times \hat{r}) \quad \text{with} \quad F_{R,W,S} = \sum_{k=0}^{\infty} \big(\alpha_k^{R,W,S} \cos kE + \beta_k^{R,W,S} \sin kE\big),$$

where F_R is the radial thrust acceleration, F_W is the circumferential thrust acceleration, F_S is the normal thrust acceleration and E is the eccentric anomaly. The work [116] determined that only 14 Fourier coefficients affect the average trajectory, and we use those coefficients as the input $\bar{u}$:

$$\bar{u} = \begin{bmatrix} \alpha_0^R & \alpha_1^R & \alpha_2^R & \beta_1^R & \alpha_0^S & \alpha_1^S & \alpha_2^S & \beta_1^S & \beta_2^S & \alpha_0^W & \alpha_1^W & \alpha_2^W & \beta_1^W & \beta_2^W \end{bmatrix}^\top.$$

The Fourier coefficients considered in [116] are chosen in $\big[-2.5 \times 10^{-7}, 2.5 \times 10^{-7}\big]$, so we can safely assume that for our case the Fourier coefficients all belong to $[-1,1]$. Following [115], the state-space form of the system dynamics is $\dot{x} = \bar{B}(x)\bar{u}$. We calculate $\bar{B}(x)$ using the averaged variational equations for the orbital elements given in [116]

$$\bar{B}(x) := \sqrt{\frac{a}{\mu}} \begin{bmatrix} 0_{2,3} & B_1(x) & 0_{2,2} & 0_{2,5} \\ 0_{2,3} & 0_{2,4} & 0_{2,2} & B_2(x) \\ B_3(x) & 0_{2,4} & B_4(x) & B_5(x) \end{bmatrix} \in \mathbb{R}^{6\times 14},$$

with $0_{i,j}$ the null matrix of i rows and j columns. We calculate the submatrices using the averaged variational equations for the orbital elements from [116]:

$$B_1(x) = \begin{bmatrix} ae & 2a\sqrt{1-e^2} & 0 & 0 \\ \frac{1}{2}(1-e^2) & -\frac{3}{2}e\sqrt{1-e^2} & \sqrt{1-e^2} & -\frac{1}{4}e\sqrt{1-e^2} \end{bmatrix}$$

$$B_2(x) = \begin{bmatrix} \cos\omega & 0 \\ 0 & \sin\omega\csc i \end{bmatrix} \begin{bmatrix} \frac{-3e}{2\sqrt{1-e^2}} & \frac{(1+e^2)}{2\sqrt{1-e^2}} & \frac{-e}{4\sqrt{1-e^2}} & -\frac{1}{2}\tan\omega & \frac{1}{4}e\tan\omega \\ \frac{-3e}{2\sqrt{1-e^2}} & \frac{(1+e^2)}{2\sqrt{1-e^2}} & \frac{-e}{4\sqrt{1-e^2}} & \frac{1}{2}\cot\omega & -\frac{1}{4}e\cot\omega \end{bmatrix}$$

$$B_3(x) = \begin{bmatrix} \sqrt{1-e^2} & -\frac{1}{2e}\sqrt{1-e^2} & 0 \\ -3 & \frac{3e}{2} + \frac{1}{2e} & -\frac{1}{2}e^2 \end{bmatrix} B_4(x) = \begin{bmatrix} \frac{1}{2e}(2-e^2) & -\frac{1}{4} \\ -\frac{1}{2e}(2-e^2)\sqrt{1-e^2} & \frac{1}{4}\sqrt{1-e^2} \end{bmatrix}$$

$$B_5(x) = \cos i \csc i \begin{bmatrix} \frac{3}{2}e\frac{\sin\omega}{\sqrt{1-e^2}} & -\frac{1}{2}(1+e^2)\frac{\sin\omega}{\sqrt{1-e^2}} & \frac{1}{4}e\frac{\sin\omega}{\sqrt{1-e^2}} & -\frac{1}{2} & \frac{1}{4}e \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with $\mu = 3.986 \times 10^{14}\,\mathrm{m}^3\mathrm{s}^{-2}$ being the standard gravitational parameter of the Earth.

We implement the orbit raising scenario presented in [115], with the orbital elements of the initial and target orbits listed in Table 5.1.

Table 5.1: Initial and Target States for Raising Maneuver

Name of the Orbital Elements	Parameters		initial	target
semi-major axis	a	[km]	6678	7345
eccentricity	e	[-]	0.67	0.737
inclination	i	[degrees]	20	22
longitude of the ascending node	Ω	[degrees]	20	22
argument of perigee	ω	[degrees]	20	22
mean anomaly	M	[degrees]	20	20

We approximate $\bar{B}(x)$ as a constant matrix $\bar{B}$ taken at the initial state. The resulting matrix is:

$$\bar{B} = 10^{-6} \times \begin{bmatrix} 0 & 0 & 0 & 18314 & 40583 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.1 & -3.4 & 2.3 & -0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.2 & 3.8 & -0.9 & -0.7 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.5 & 4 & -0.9 & 5.6 & -1.9 \\ 3 & -2.7 & 0 & 0 & 0 & 0 & 0 & 4.7 & -1 & 5.2 & -3.8 & 1.3 & -5.6 & 1.9 \\ -12.3 & 7.2 & -0.9 & 0 & 0 & 0 & 0 & -3.5 & 0.8 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Coefficients $\bar{B}_{1,4}$ and $\bar{B}_{1,5}$ are significantly larger than all the other coefficients of $\bar{B}$ because the semi-major axis is larger than any other element, as can be seen in Table 5.1. Losing control over one of the 14 Fourier coefficients means that a certain frequency of the thrust vector cannot be controlled. Since the coefficients $\bar{B}_{1,5}$ and $\bar{B}_{6,1}$ have a magnitude significantly larger than coefficients of respectively the first and last row of $\bar{B}$, we have the intuition that the system is not resilient to the loss of the 1^{st} or the 5^{th} Fourier coefficient.

The matrix $\bar{B}$ is full rank, so $\dot{x} = \bar{B}\bar{u}$ is controllable. We denote with r_{min} and r_q the vectors whose components are respectively $r_{min}(j)$ and $r_q(j)$ for the loss of the frequency $j \in [\![1, 14]\!]$,

$$r_{min} = \begin{bmatrix} -0.2 & 0.34 & 0.9 & -0.004 & -0.38 & 0.15 & 0.83 & -0.32 & 0.71 & -0.06 & 0.24 & 0.2 & -0.5 & 0.5 \end{bmatrix}.$$

Since the 1^{st}, 4^{th}, 5^{th}, 8^{th}, 10^{th}, and 13^{th} values of r_{min} are negative, according to Corollary 3 the system is not resilient to the loss of control over any one of these six corresponding frequencies. Their associated r_q is zero. This result validates our intuition about the 1^{st} and 5^{th} frequencies. Corollary 3 also states the resilience of the spacecraft to the loss over any one of the 2^{nd}, 3^{rd}, 6^{th}, 7^{th}, 9^{th}, 11^{th}, 12^{th} and 14^{th} frequency because their r_{min} belongs to $(0, 1]$. Indeed, the input bounds are symmetric, so we can use the results from [32] stating that $r(C) = r(-C) = r_{min}$. Then, using Theorem 12 we deduce that

$$r_q = \begin{bmatrix} 0 & 0.34 & 0.9 & 0 & 0 & 0.15 & 0.83 & 0 & 0.71 & 0 & 0.24 & 0.2 & 0 & 0.5 \end{bmatrix}.$$

Since $r_q(3)$, $r_q(7)$ and $r_q(9)$ are close to 1, the loss of one of these three frequency would not delay significantly the system. The lowest positive value of r_q occurs for the 6^{th} frequency, $r_q(6) = 0.15$. Its inverse, $\frac{1}{r_q(6)} = 6.8$ means that the malfunctioning system can take up to 6.8 times longer than the initial system to reach a target.

The maneuver described in Table 5.1 yields $d = x_{goal} - x_0 = \big(667,\ 0.067,\ 2,\ 2,\ 2,\ 2\big)$. We compute the associated time ratios $t(d)$ using (5.11) and (5.15) for the loss over each column of $\bar{B}$:

$$t(d) = \begin{bmatrix} 1.1 & 1.2 & 1.1 & 1 & \infty & 1 & 151.1 & \infty & 151.1 & \infty & 151.1 & 151.1 & \infty & 151.1 \end{bmatrix}. \tag{5.18}$$

Then, losing control over one of the first four frequencies will barely increase the time required for the malfunctioning system to reach the target compared with the initial system. However, after the loss over the 7^{th}, 9^{th}, 11^{th}, 12^{th}, or the 14^{th} frequency of the thrust vector, the undesirable input can multiply the maneuver time by a factor of up to 151.1. If one of the 5^{th}, 8^{th}, 10^{th}, or the 13^{th} frequency is lost, then some undesirable inputs can render the maneuver impossible to perform.

When computing r_q, we have seen that the system is not resilient to the loss of the 1^{st} or the 4^{th} frequency. Yet, the specific target described in Table 5.1 happens to be reachable for the same loss since the 1^{st} and 4^{th} components of $t(d)$ in (5.18) are finite. Indeed, r_q speaks only about a target for which the undesirable inputs cause maximal possible delay.

5.7.2 Resilience of an octocopter

Resilience of unmanned aerial vehicles (UAV) to propeller failure is crucial to their operations over populated areas [8]. Because quadcopters have 4 inputs for 6 degrees of freedom, they are underactuated and thus cannot be resilient to the loss of control authority over one of their propellers [8]. Instead, we consider the octocopter from [28] represented on Fig. 5.1. Its design decouples the rotational and the translational dynamics, allowing to keep a payload horizontal, which is crucial for pizza delivery for instance.

In Sections 5.7.2 and 5.7.2, we will first quantify the resilience of this UAV model. Since propellers cannot operate in a bang-bang fashion, we will then add propellers' dynamics to the UAV model in Section 5.7.2. Because of this modification the UAV dynamics are not driftless. Hence, most of our theory does not apply but still provides good intuition on the quantitative resilience of this octocopter model.

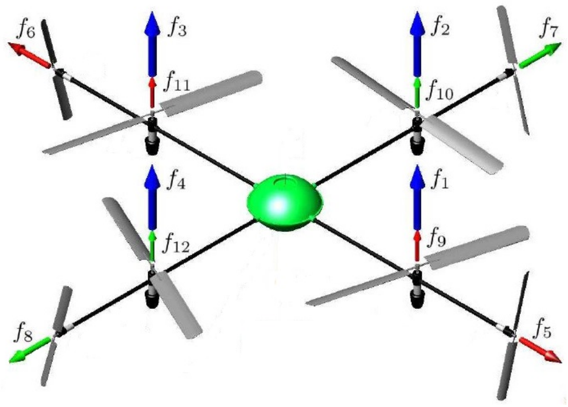


Figure 5.1: Octocopter layout, image modified from [28].

Rotational dynamics

The roll, pitch and yaw angles of the octocopter are gathered in $Y := (\phi, \theta, \psi)$. The propeller $i \in [\![1, 8]\!]$ spinning at an angular velocity ω_i produces a force $f_i = k\omega_i^2$, with the thrust coefficient k. The airflow created by the lateral rotors produces the extra vertical forces $f_9, \ldots, f_{12}$ on Fig. 5.1. From [28], $f_{9+i} = bf_{5+i}$ for $i \in [\![0, 3]\!]$ with the coupling constant $b = 0.64$. Relying on [110] and [28], the rotational dynamics of the octocopter are

$$\begin{aligned}
\ddot{\phi} &= \frac{I_y - I_z}{I_x}\dot{\theta}\dot{\psi} + \frac{lk}{I_x}\Big(\omega_3^2 - \omega_1^2 + b\big(\omega_7^2 - \omega_8^2\big)\Big) - \frac{I_{rotor}}{I_x}\dot{\theta}\gamma \\
\ddot{\theta} &= \frac{I_z - I_x}{I_y}\dot{\theta}\dot{\phi} + \frac{lk}{I_y}\Big(\omega_2^2 - \omega_4^2 + b\big(\omega_5^2 - \omega_6^2\big)\Big) + \frac{I_{rotor}}{I_y}\dot{\phi}\gamma \\
\ddot{\psi} &= \frac{I_x - I_y}{I_z}\dot{\theta}\dot{\phi} + \frac{d}{I_z}\Big(\omega_2^2 + \omega_4^2 - \omega_1^2 - \omega_3^2\Big),
\end{aligned}$$

with $\gamma = \omega_2 + \omega_4 - \omega_1 - \omega_3$. The rotational equations are linearized around $\dot{Y} = 0$ and become $\ddot{Y} = \bar{B}_r\Omega$, with $\Omega \in \mathbb{R}^8$ gathering the squared angular velocities of the propellers $\omega_1^2, \ldots, \omega_8^2$ and

$$\bar{B}_r = \begin{bmatrix} \frac{lk}{I_x} & 0 & 0 \\ 0 & \frac{lk}{I_y} & 0 \\ 0 & 0 & \frac{d}{I_z} \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & b & -b \\ 0 & 1 & 0 & -1 & b & -b & 0 & 0 \\ -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since the input sets are nonsymmetric: $\bar{u}_i := \omega_i^2 \in [0, \omega_{max}^2]$, and the dynamics are given by a double integrator, the theory of [32] cannot deal with this UAV model. Using Theorem 12 we calculate the quantitative resilience of the system $\dot{v}_Y(t) = \bar{B}_r\bar{u}(t)$ with $v_Y := \dot{Y}$ for the loss of control over each single propeller: $r_{min} = \big[0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1\big]$. Based on Corollary 3, the UAV is thus resilient to the loss of control over any single propeller in terms of angular velocity and $r_q = r_{min}$. Following Theorem 13 we deduce that $\ddot{Y}(t) = \bar{B}_r\bar{u}(t)$ is also resilient and $r_{2,q} = \sqrt{r_q} = \sqrt{0.1} = 0.32$. Then, $\frac{1}{r_{2,q}} = 3$ and $\frac{1}{r_q} = 10$ mean that after the loss of control over any single propeller the UAV might take as much as three times longer to reach a given orientation, while it might be ten times slower to reach a given angular velocity.

Translational dynamics

In the inertial frame the position of the UAV is $X := (x, y, z)$ and its orientation yields the rotation matrix $R(\psi, \theta, \phi)$. The translational equations of motion from [28] are $m\ddot{X} = R(\psi, \theta, \phi)\bar{B}_t k\Omega + G$, i.e.,

$$m\begin{bmatrix}\ddot{x}\\ \ddot{y}\\ \ddot{z}\end{bmatrix} = R(\psi,\theta,\phi)\begin{bmatrix}0&0&0&0&1&-1&0&0\\0&0&0&0&0&0&1&-1\\1&1&1&1&b&b&b&b\end{bmatrix}k\Omega + \begin{bmatrix}0\\0\\-mg\end{bmatrix}. \tag{5.19}$$

Because of the gravitation term G, the above dynamics are affine. We combine G with the input Ω to make the dynamics driftless using $R(\psi, \theta, \phi)^{-1} = R(-\psi, -\theta, -\phi)$,

$$\begin{aligned} m\ddot{X} &= R(\psi,\theta,\phi)\big(\bar{B}_t k\Omega + R(-\psi,-\theta,-\phi)G\big)\\ &= R(\psi,\theta,\phi)\begin{bmatrix} k(\omega_5^2-\omega_6^2) - mg(-c_\psi s_\theta c_\phi + s_\psi s_\phi)\\ k(\omega_7^2-\omega_8^2) - mg(s_\psi s_\theta c_\phi + c_\psi s_\phi)\\ k(\omega_1^2+\omega_2^2+\omega_3^2+\omega_4^2) + bk(\omega_5^2+\omega_6^2+\omega_7^2+\omega_8^2) - mgc_\theta c_\phi\end{bmatrix}.\end{aligned}$$

Since the rotational dynamics are resilient, we know that the controller can maintain the UAV horizontal even after the loss of control over a propeller. From now on, we will then assume $\theta = \phi = 0°$. To prevent obfuscating the following analysis, we assume that this orientation is maintained no matter the inputs u and w. Additionally, the yaw angle does not affect the translational dynamics, so we also take $\psi = 0°$. Then, the translational dynamics of the octocopter are equivalent to that of a point-mass model and they are fully decoupled from the rotational dynamics, as desired by design [28]. The position of the UAV is $X := (x, y, z)$ and satisfies

$$m\ddot{X} = \begin{bmatrix} k(\omega_5^2-\omega_6^2)\\ k(\omega_7^2-\omega_8^2)\\ k\sum_{i=1}^4 \omega_i^2 + bk\sum_{i=5}^8 \omega_i^2 - mg\end{bmatrix}.$$

The horizontal propellers $(\omega_1, \dots, \omega_4)$ are designed to sustain the weight of the drone while the lateral ones $(\omega_5, \dots, \omega_8)$ are smaller and should mostly be used for lateral displacements. Thus, we define the inputs $\bar{u}_i := k\omega_i^2 - \frac{mg}{4} \in [-\frac{mg}{4}, k\omega_{max}^2 - \frac{mg}{4}]$ for $i \in [\![1,4]\!]$ and $\bar{u}_i := k\omega_i^2 \in [0, k\omega_{max}^2]$ for $i \in [\![5,8]\!]$. Then, the translational dynamics become

$$\ddot{X}(t) = \bar{B}_t\bar{u}(t), \quad \dot{X}(0) = X(0) = 0 \in \mathbb{R}^3, \tag{5.20}$$

$$\text{with}\quad \bar{B}_t = \frac{1}{m}\begin{bmatrix}0&0&0&0&1&-1&0&0\\0&0&0&0&0&0&1&-1\\1&1&1&1&b&b&b&b\end{bmatrix}.$$

After the loss of control authority over a propeller, we split $\bar{B}_t$ and $\bar{u}$ into B, C and u, w as before. The initial state is the same and the malfunctioning dynamics are

$$\ddot{X}(t) = Bu(t) + Cw(t). \tag{5.21}$$

Matrix $\bar{B}_t$ has more columns than rows and the first four columns are identical so we expect the system to be resilient to the loss of any one of them. However, only column 5 can counteract column 6 and only column

7 can counteract column 8, and vice-versa. We thus have the intuition that the system is not resilient to the loss of any one of the last four columns.

For system $\dot{v} = \bar{B}_t \bar{u}$, with $v := \dot{X}$, Theorem 12 yields

$$r(C) = \begin{bmatrix} 0.766 & 0.766 & 0.766 & 0.766 & 0 & 0 & 0 & 0 \end{bmatrix},$$
$$r(-C) = \begin{bmatrix} 0.564 & 0.564 & 0.564 & 0.564 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then, according to Corollary 3 the system of dynamics $\dot{v} = \bar{B}_t \bar{u}$ is only resilient to the loss of any one of the first four propellers. Following Theorem 12, $r_q = \min\big\{r(C), r(-C)\big\} = \big[0.564\ 0.564\ 0.564\ 0.564\ 0\ 0\ 0\ 0\big]$. Since Theorem 13 only applies to resilient systems, we use it on the first four propellers $r_{2,q} = \sqrt{r_q} = \big[0.75\ 0.75\ 0.75\ 0.75\big]$. Then, $\frac{1}{r_q} = 1.77$ and $\frac{1}{r_{2,q}} = 1.33$ mean that the after the loss of a horizontal propeller, the UAV might need 1.77 times longer to reach a given velocity but only 1.33 times longer to reach a desired position.

Let us now evaluate how the loss of a propeller impacts the vertical velocity. We take $d = (0, 0, -1)$ and compute

$$t(d) = \big[1.77\ 1.77\ 1.77\ 1.77\ 2.26\ 2.26\ 2.26\ 2.26\big]. \tag{5.22}$$

The first four values are the same as $1/r_q$ because the direction the worst impacted by the loss of a horizontal propeller is along d, i.e., the loss of a horizontal propeller has the worst impact on the vertical velocity. We now simulate various loss of controls and aim to fly vertically the UAV along $d = (0, 0, -1)$.

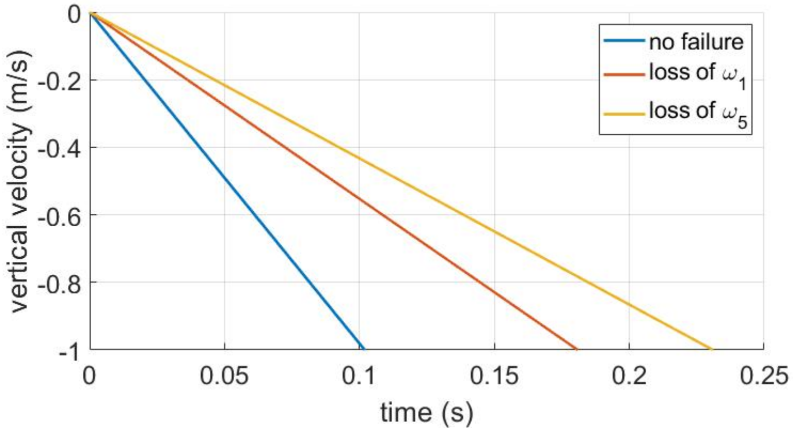


Figure 5.2: Time evolution of $\dot{z}$. For 'no failure', $\dot{v} = \bar{B}_t \bar{u}^{min}$. For 'loss of ω_1', $\dot{v} = Bu + Cw$ with C the 1^{st} column of $\bar{B}_t$, $w = k\omega_{max}^2 - mg/4$ and $u = u^{min}$. For 'loss of ω_5', $\dot{v} = Bu + Cw$ with C the 5^{th} column of $\bar{B}_t$, $w = k\omega_{max}^2$ and $u = u^{min}$ except $\bar{u}_6 = k\omega_{max}^2$ to keep the UAV on the z-axis.

As illustrated on Fig. 5.2, to reach the velocity $v = (0, 0, -1)$, the nominal system needs $0.102\,s$, while the malfunctioning ones need $0.181\,s$ and $0.231\,s$ after the loss of ω_1 and ω_5 respectively. Then, the reach times increased by factors 1.77 and 2.26, exactly the values calculated in (5.22) as the choice of inputs in the simulation is optimal.

We now study $T_N^*(d)$ and $T_M^*(d)$ for the velocity targets $d(\beta) = (0, \cos\beta, \sin\beta)$ for all $\beta \in [0, 2\pi]$. After the loss of ω_1, $\frac{1}{r_q} = 1.77$, so $T_M^*(d) \leq 1.77\ T_N^*(d)$ for any $d \in \mathbb{R}^n$, as illustrated on Fig. 5.3.

Note that $d(\frac{3\pi}{2}) = (0, 0, -1)$ and as calculated in (5.22) we have $T_M^*(d(\frac{3\pi}{2})) = 1.77\ T_N^*(d(\frac{3\pi}{2}))$ as shown on Fig. 5.3. If the inputs were symmetric we would have $T_M^*(\beta) = T_M^*(\beta + \pi)$ for all β as in Fig. 1 of [32].

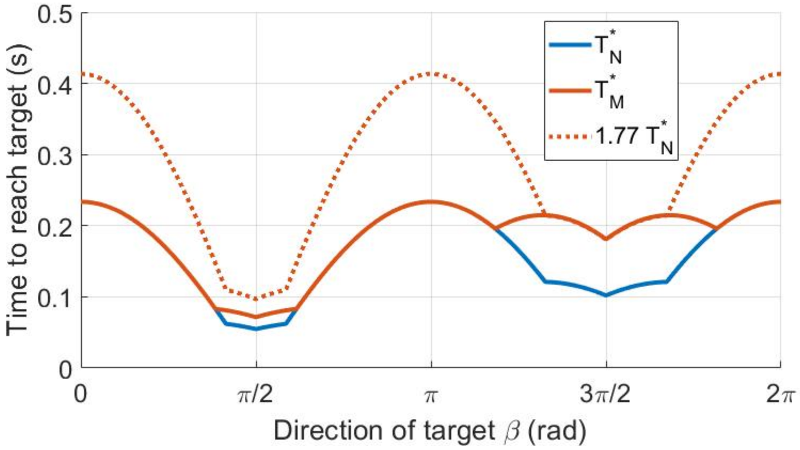


Figure 5.3: Evolution of $T_N^*(d)$ and $T_M^*(d)$ for a velocity target $d(\beta) = (0, \cos\beta, \sin\beta)$.

However, our inputs are not symmetric and thus $T_M^*(\frac{\pi}{2}) \neq T_M^*(\frac{3\pi}{2})$ as shown on Fig. 5.3. The lack of input symmetry results in $T_M^*(\beta) \neq T_M^*(\beta + \pi)$ as shown on Fig. 5.3. Such a situation could not be handled by the preliminary work [32].

High-fidelity dynamics of the propellers

So far in this work, all inputs were bang-bang because our definition of quantitative resilience asks for time-optimal transfers. The inputs of the translational dynamics (5.20) encode the propellers' angular velocities, which cannot physically change in a bang-bang fashion. Thus, in order to provide a more realistic model and display the capabilities of our work, we follow [107] and add first-order propellers' dynamics:

$$\ddot{X}(t) = \bar{B}_t \bar{u}(t), \quad \dot{\bar{u}}(t) = \frac{1}{\tau}\big(\bar{u}^c(t) - \bar{u}(t)\big), \tag{5.23}$$

with $\bar{u}^c \in \mathbb{R}^8$ a new, possibly bang-bang, command signal. System (5.23) is not driftless, hence preventing a direct application of our theory. Instead, we proceed heuristically, building on the intuition provided by our theory to tackle this high-fidelity model.

The time constant $\tau = 0.1\,s$ is chosen to match the propeller response in Fig. 3 of [92] which also corresponds to standard models in the literature [8], [28], [110] Because of these new dynamics, the system is not driftless anymore, but is modeled with

$$\begin{bmatrix} \dot{X} \\ \ddot{X} \\ \dot{\bar{u}} \end{bmatrix} = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & \bar{B}_t \\ 0 & 0 & -\frac{1}{\tau} I \end{bmatrix} \begin{bmatrix} X \\ \dot{X} \\ \bar{u} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\tau} I \end{bmatrix} \bar{u}^c. \tag{5.24}$$

After the loss of control over the first propeller, we split $\bar{B}_t$ and $\bar{u}$ as before such that

$$\ddot{X}(t) = Bu(t) + Cw(t), \quad \begin{cases} \dot{u}(t) = \frac{1}{\tau}\big(u^c(t) - u(t)\big), \\ \dot{w}(t) = \frac{1}{\tau}\big(w^c(t) - w(t)\big), \end{cases} \tag{5.25}$$

with the bang-bang command signals u^c and w^c. We will now study how the actuators' dynamics impact the resilience of the UAV in the vertical direction $d = (0, 0, 1)$.

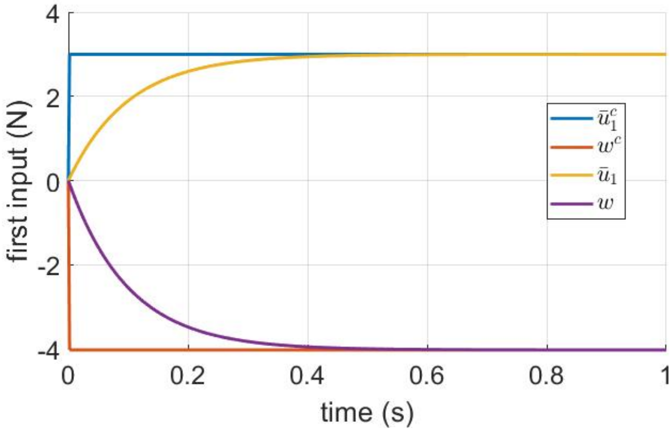


Figure 5.4: Exponential convergence of $\bar{u}_1$ and w to their bang-bang commands $\bar{u}_1^c = \bar{u}_1^{max} = k\omega_{max}^2 - \frac{mg}{4}$ and $w^c = \bar{u}_1^{min} = -\frac{mg}{4}$.

Since the inputs $\bar{u}$ in (5.23) and (u, w) in (5.25) have a non-zero rise time as shown on Fig. 5.4, the vertical velocities $\dot{z}_N$ of (5.23) and $\dot{z}_M$ of (5.25) react smoothly and slower than their bang-bang counterparts, as illustrated on Fig. 5.5. For $t \geq 0.4\,s$, $\bar{u}$ and (u, w) have converged to their commands $\bar{u}^c$ and (u^c, w^c), and thus the two slopes of $\dot{z}_N(t)$ in (5.20) and (5.23) are equal, as shown on Fig. 5.5, and so are that of $\dot{z}_M(t)$ in (5.21) and (5.25).

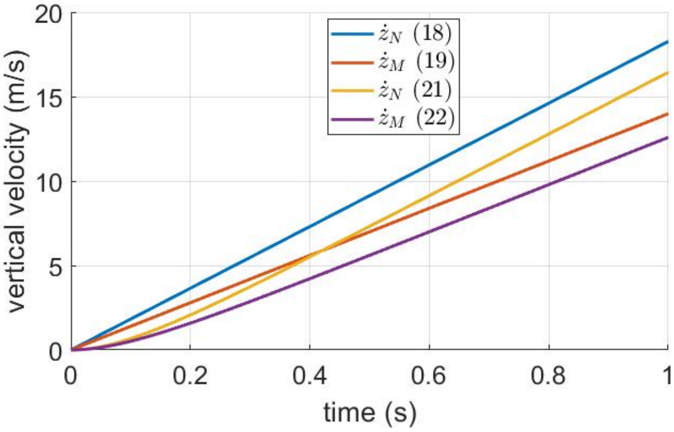


Figure 5.5: Vertical velocities $\dot{z}_N(t)$ and $\dot{z}_M(t)$ of the nominal and malfunctioning systems demonstrating the impact of the propellers' dynamics in (5.23) and (5.25).

The slower reaction time caused by the dynamics of the propellers is also reflected on the vertical positions z_N and z_M on Fig. 5.6.

Because of the specific geometry of the system, the optimal inputs for direction $d = (0, 0, 1)$ were trivial to determine. Then, we calculate the ratio of reach times for systems (5.23) and (5.25), $t(d) = 1.12$ and for systems (5.20) and (5.21), $t^c(d) = 1.14$. Hence, modeling the dynamics of the propellers increases slightly the resilience of the vertical dynamics.

However, the time-optimal commands $\bar{u}^c$ for (5.23) and (u^c, w^c) for (5.25) can be time-varying for other directions $d \in \mathbb{R}^3$ [78], and determining these optimal commands requires complex algorithms [30], [71]

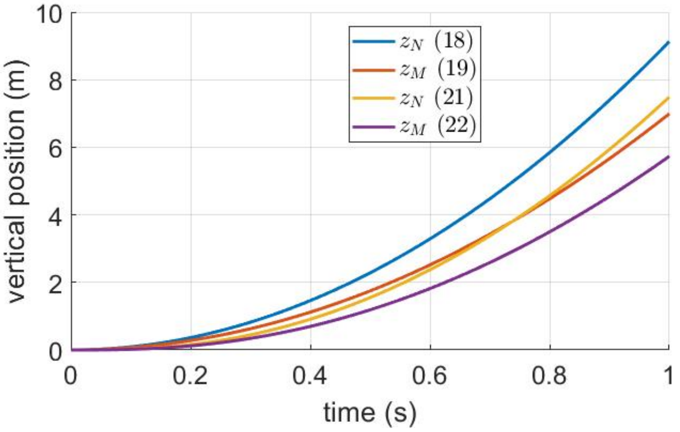


Figure 5.6: Vertical positions $z_N(t)$ and $z_M(t)$ of the nominal and malfunctioning systems demonstrating the impact of the propellers' dynamics in (5.23) and (5.25).

because the dynamics are not driftless anymore. Additionally, the Maximax-Minimax Quotient Theorem of [36] does not hold, which invalidates Theorem 11 and prevents the exact calculation of r_q without calculating $t(d)$ for all $d \in \mathbb{R}^3$. A stronger theory will be needed to tackle linear non-driftless systems.

5.8 Supporting Lemmata

Lemma 1: If system (5.6) is controllable, then for all $d = x_{goal} - x_0 \in \mathbb{R}^n$, the infimum $T_N^*(d)$ of (5.8) is a minimum achieved by a constant control input $\bar{u}^* \in \bar{\mathcal{U}}$.

Proof. According to Theorem 4.3 of [78] there exists a time optimal control $\bar{u}^* \in \mathcal{F}(\bar{\mathcal{U}})$. Following Pontryagin maximum principle [78], $\bar{u}^*$ is bang-bang but does not switch since the dynamics are driftless. Thus, the infimum T_N^* in (5.8) is a minimum achieved by a constant control input. □

Lemma 2: If system (5.7) is resilient, then for all $d \in \mathbb{R}^n_*$ and all $w \in \mathcal{F}(\mathcal{W})$, the infimum $T_M(w,d)$ of (5.9) is a minimum achieved by a constant control input $u^*(w) \in \mathcal{U}$ and

$$T_M(w,d) := \min_{u \in \mathcal{U}} \left\{ T \geq 0 : \int_0^T \big[Bu(t) + Cw(t)\big]\, dt = d \right\}. \tag{5.26}$$

Proof. The infimum of (5.9) is

$$T_M(w,d) = \inf_{u \in \mathcal{F}(\mathcal{U})} \left\{ T \geq 0 : \int_0^T Bu(t)\, dt = z \right\}, \qquad \text{with} \qquad z := d - \int_0^T Cw(t)\, dt \in \mathbb{R}^n,$$

a constant vector for w fixed. Since system (5.7) is resilient, any $z \in \mathbb{R}^n$ is reachable. Following Lemma 1 and Theorem 4.3 of [78], a constant time-optimal control exists and the infimum of (5.9) is a minimum. □

Lemma 3: For a resilient system following (5.7), function $T_M(w,d) := \min_{u \in \mathcal{U}} \big\{ T \geq 0 : (Bu + Cw)T = d \big\}$ is continuous in $w \in \mathcal{W}$ and $d \in \mathbb{R}^n_*$.

Proof. With $\mathcal{X} := \big\{Cw : w \in \mathcal{W}\big\}$, $\mathcal{Y} := \big\{Bu : u \in \mathcal{U}\big\}$ and $\lambda = \frac{1}{T}$ we obtain

$$T_M(x,d) = \frac{1}{\underset{y\,\in\,\mathcal{Y}}{\max}\big\{\lambda \geq 0 : x + y = \lambda d\big\}}.$$

Since $\|d\| > 0$ and $\lambda \geq 0$, we have $\lambda = \frac{\|\lambda d\|}{\|d\|} = \frac{\|x+y\|}{\|d\|}$. Let $d_1 := \frac{d}{\|d\|}$, then

$$T_M(x,d) = \frac{\|d\|}{\underset{y\,\in\,\mathcal{Y}}{\max}\big\{\|x+y\| : x + y \in \mathbb{R}^+ d_1\big\}}$$

and Lemma 5.2 of [36] states that T_M is continuous in w and d. □

5.9 Summary

This chapter introduced the notion of quantitative resilience for linear systems with multiple integrators and nonsymmetric input sets. Relying on bang-bang control theory and on two specific optimization results, we transformed a nonlinear problem consisting of four nested optimizations into a single linear optimization. This simplification leads to a computationally efficient method for verifying the resilience and calculating the quantitative resilience of driftless systems with multiple integrators.

Chapter 6

The Maximax Minimax Quotient Theorem

6.1 Introduction

This chapter establishes the Maximax Minimax Quotient Theorem and is adapted from our work [36]. This theorem focuses on the specific fractional optimization problem introduced in Chapter 5 and composed of four nested optimization problems. For this reason, a search algorithm would have a high computational cost and would be especially wasteful since an analytical solution exists.

Our ratio of interest features a max-min optimization [117] belonging to the setting of semi-infinite programming [118]. Because of the infinite number of constraints, it is not possible to immediately apply the classical results of linear max-min theory [119] stating that the maximum is attained on the boundary of the constraint set. Nonetheless, thanks to the specific geometry of our problem we are able to prove a very similar result.

Armed with this preliminary result on max-min programming, we formulate and establish the Maximax Minimax Quotient Theorem. This result concerns the maximization of a ratio of a maximum and a minimax over two polytopes. In the special case where these polytopes are symmetric, this result reduces to Theorem 3.2 of [32], whose the proof was again omitted for length concerns.

The remainder of this chapter is organized as follows. Section 6.2 establishes the existence of the Maximax Minimax Quotient and proves a preliminary optimization result. Section 6.3 states our central theorem and provides its proof. Section 6.4 gathers all the lemmas involved in the proof of the Maximax Minimax Quotient Theorem. Section 6.5 justifies the continuity of two maxima functions used during the proof of our main result. Finally, Section 6.6 illustrates the proof of our theorem on a simple example.

6.2 Preliminaries

Definition 11: A *polytope* in $\mathbb{R}^n$ is a compact intersection of finitely many half-spaces.

Thus, this chapter only considers convex polytopes. If $\mathcal{X}$ and $\mathcal{Y}$ are two nonempty polytopes in $\mathbb{R}^n$ with

$-\mathcal{X} \subset \text{int}(\mathcal{Y})$, and $d \in \mathbb{S}$, we define the *Maximax Minimax Quotient* as

$$r_{\mathcal{X},\mathcal{Y}}(d) := \frac{\underset{x \,\in\, \mathcal{X},\ y \,\in\, \mathcal{Y}}{\max} \big\{ \|x+y\| : x+y \in \mathbb{R}^+ d \big\}}{\underset{x \,\in\, \mathcal{X}}{\min} \big\{ \underset{y \,\in\, \mathcal{Y}}{\max} \big\{ \|x+y\| : x+y \in \mathbb{R}^+ d \big\} \big\}}. \tag{6.1}$$

The objective of the Maximax Minimax Quotient Theorem is to determine the direction d that maximizes $r_{\mathcal{X},\mathcal{Y}}(d)$. Note that in the numerator of (6.1), x and y are chosen together to satisfy the constraint $x+y \in \mathbb{R}^+ d$, while in the denominator this constraint only applies to y. Before starting the actual proof of this theorem, we first need to justify the existence of the minimum and the maxima appearing in (6.1).

Proposition 18: Let $\mathcal{X}$, $\mathcal{Y}$ be two nonempty polytopes in $\mathbb{R}^n$ with $-\mathcal{X} \subset \text{int}(\mathcal{Y})$, $\dim \mathcal{Y} = n$ and $d \in \mathbb{S}$. Then,

(a) $\underset{x \,\in\, \mathcal{X},\, y \,\in\, \mathcal{Y}}{\max} \big\{ \|x+y\| : x+y \in \mathbb{R}^+ d \big\}$ exists,

(b) $\lambda^*(x,d) := \underset{y \,\in\, \mathcal{Y}}{\max} \big\{ \|x+y\| : x+y \in \mathbb{R}^+ d \big\}$ exists for all $x \in \mathcal{X}$,

(c) $\underset{x \,\in\, \mathcal{X}}{\min} \big\{ \lambda^*(x,d) \big\}$ exists,

(d) and $\underset{x \,\in\, \mathcal{X}}{\min} \big\{ \lambda^*(x,d) \big\} > 0$.

Proof. (a) Let $S := \big\{ (x,y) \in \mathcal{X} \times \mathcal{Y} : x+y \in \mathbb{R}^+ d \big\}$. Set S is a closed subset of the compact set $\mathcal{X} \times \mathcal{Y}$, so S is compact. Since $\mathcal{X}$ is nonempty, we take $x \in \mathcal{X}$. Using $-\mathcal{X} \subset \mathcal{Y}$ we have $-x \in \mathcal{Y}$ and $x + (-x) = 0 \in \mathbb{R}^+ d$. Then, $(x, -x) \in S$, so S is nonempty. Function $f : S \to \mathbb{R}$ defined as $f(x,y) := \|x+y\|$ is continuous, so it reaches a maximum over S.

(b) For $x \in \mathcal{X}$ define $S(x) := \big\{ y \in \mathcal{Y} : x+y \in \mathbb{R}^+ d \big\}$. Since $S(x)$ is a closed subset of the compact set $\mathcal{Y}$, $S(x)$ is compact. Since $-\mathcal{X} \subset \mathcal{Y}$, we have $-x \in S(x)$ and so $S(x) \neq \emptyset$. Function $f_x : S(x) \to \mathbb{R}$ defined as $f_x(y) := \|x+y\|$ is continuous, so it reaches a maximum over $S(x)$, i.e., λ^* exists.

(c) For $x \in \mathcal{X}$ and $d \in \mathbb{S}$, the argument of $\lambda^*(x,d)$ is uniquely defined as $y^*(x,d) := \lambda^*(x,d)d - x$ since $\|d\| = 1$ and

$$y^*(x,d) = \arg \underset{y \,\in\, \mathcal{Y}}{\max} \big\{ \|x+y\| : x+y \in \mathbb{R}^+ d \big\}. \tag{6.2}$$

Lemma 15 shows that λ^* is continuous in x and d, so y^* is also continuous in x and d. Then, function $f : \mathcal{X} \to \mathbb{R}$ defined as $f(x) := \|x + y^*(x,d)\|$ is continuous, so it reaches a minimum over the compact and nonempty set $\mathcal{X}$.

(d) Note that $y^*(x,d) \in \partial \mathcal{Y}$ for all $x \in \mathcal{X}$. Indeed, assume for contradiction purposes that there exists $\varepsilon > 0$ such that $B_\varepsilon\big(y^*(x,d)\big) \in \mathcal{Y}$. We required $\dim \mathcal{Y} = n$ to make this ball of full dimension, so that $z := y^*(x,d) + \varepsilon d \in \mathcal{Y}$. Then, $x + z = \big(\lambda^*(x,d) + \varepsilon\big)d \in \mathbb{R}^+ d$ and $\|x+z\| = \lambda^*(x,d) + \varepsilon > \lambda^*(x,d)$ contradicting the optimality of λ^*. Thus, $y^*(x,d) \in \partial \mathcal{Y}$. Since $-\mathcal{X} \subset \text{int}(\mathcal{Y})$, we have $\|x + y^*(x,d)\| > 0$ for all $x \in \mathcal{X}$.

□

Then, with the assumptions of Proposition 18 the Maximax Minimax Quotient is well-defined. The proof of our main theorem relies on another optimization result stating that the argument of the minimum in (6.1) lies at a vertex of $\mathcal{X}$.

Definition 12: A *vertex* of a set $\mathcal{X} \subset \mathbb{R}^n$ is a point $x \in \mathcal{X}$ such that if there are $x_1 \in \mathcal{X}$, $x_2 \in \mathcal{X}$ and $\lambda \in [0,1]$ with $x = \lambda x_1 + (1-\lambda)x_2$, then $x = x_1 = x_2$.

With this definition, a vertex of a polytope corresponds to the usual understanding of a vertex of a polytope.

Theorem 14: Let $d \in \mathbb{S}$, $\mathcal{X}$ and $\mathcal{Y}$ two polytopes of $\mathbb{R}^n$ with $-\mathcal{X} \subseteq \mathcal{Y}$ and $\dim \mathcal{Y} = n$. Then, there exists a vertex v of $\mathcal{X}$ where $\min\limits_{x \in \mathcal{X}} \big\{\lambda^*(x,d)\big\}$ is reached.

Proof. According to Proposition 18 the minimum of λ^* exists. Then, let $x^* \in \mathcal{X}$ such that

$$\lambda^*(x^*,d) = \min_{x \in \mathcal{X}} \big\{\lambda^*(x,d)\big\}, \qquad \text{i.e.,} \qquad \|y^*(x^*) + x^*\| = \min_{x \in \mathcal{X}} \|y^*(x) + x\|.$$

Since $-x^*$ must minimize the distance between itself and $y^*(x^*) \in \partial\mathcal{Y}$, with $-\mathcal{X} \subset \mathcal{Y}$ obviously $x^* \in \partial\mathcal{X}$. Assume now that x^* is not on a vertex of $\partial\mathcal{X}$. Let S_x be the surface of lowest dimension in $\partial\mathcal{X}$ such that $x^* \in S_x$ and $\dim S_x \geq 1$.

Let v be a vertex of S_x and $x(\alpha) := x^* + \alpha(v - x^*)$ for $\alpha \in \mathbb{R}$. Notice that $x(0) = x^*$ and $x(1) = v$. Due to the choice of v, the convexity of S_x and x^* not being a vertex, there exists $\varepsilon > 0$ such that $x(\alpha) \in S_x$ for all $\alpha \in [-\varepsilon, 1]$. We also define the lengths $L(\alpha) := \|y^*\big(x(\alpha)\big) + x(\alpha)\|$ and $L^* := L(0)$.

Since $\|d\| = 1$ and $y^*\big(x(\alpha)\big) + x(\alpha) \in \mathbb{R}^+ d$, we have $L(\alpha) = \langle y^*\big(x(\alpha)\big) + x(\alpha), d\rangle$. By definition of x^*, we know that $L^* \leq L(\alpha)$ for all $\alpha \in [-\varepsilon, 1]$. For contradiction purposes assume that there exists $\alpha_0 \in (0,1]$ such that $L^* < L(\alpha_0)$. We introduce the convexity coefficient $\beta := \frac{\alpha_0}{\alpha_0 + \varepsilon} > 0$ and then

$$\begin{aligned} L^* &= \beta L^* + (1-\beta)L^* < \beta L(-\varepsilon) + (1-\beta)L(\alpha_0) \\ &< \beta\langle y^*\big(x(-\varepsilon)\big) + x(-\varepsilon), d\rangle + (1-\beta)\langle y^*\big(x(\alpha_0)\big) + x(\alpha_0), d\rangle = \langle z + x^*, d\rangle, \end{aligned}$$

with $z := \beta y^*\big(x(-\varepsilon)\big) + (1-\beta)y^*\big(x(\alpha_0)\big)$. Indeed, note that $\beta x(-\varepsilon) + (1-\beta)x(\alpha_0) = x^*$, and $z + x^* \in \mathbb{R}^+ d$. Note that

$$L^* = \max_{y \in \mathcal{Y}} \big\{\langle x^* + y, d\rangle : x^* + y \in \mathbb{R}^+ d\big\}, \qquad \text{but} \qquad L^* < \langle x^* + z, d\rangle.$$

Given that $z \in \mathcal{Y}$ by convexity of $\mathcal{Y}$ and $x^* + z \in \mathbb{R}^+ d$, we have reached a contradiction. Thus, there is no $\alpha_0 \in (0,1]$ such that $L^* < L(\alpha_0)$. Therefore, for all $\alpha \in [0,1]$, $L(\alpha) = L^*$. By taking $\alpha = 1$, we have $x(\alpha) = v$, so the minimum L^* is also reached on the vertex v of $\mathcal{X}$. □

We have now all the preliminary results necessary to state our central theorem.

6.3 The Maximax Minimax Quotient Theorem

Theorem 15 (Maximax Minimax Quotient Theorem)**:** If $\mathcal{X}$ and $\mathcal{Y}$ are two polytopes in $\mathbb{R}^n$ with $-\mathcal{X} \subseteq \text{int}(\mathcal{Y})$, $\dim \mathcal{X} = 1$, $\partial\mathcal{X} = \{x_1, x_2\}$ with $x_2 \neq 0$ and $\dim \mathcal{Y} = n$, then $\max\limits_{d \in \mathbb{S}} r_{\mathcal{X},\mathcal{Y}}(d) = \max\big\{r_{\mathcal{X},\mathcal{Y}}(x_2), r_{\mathcal{X},\mathcal{Y}}(-x_2)\big\}$.

Proof. Since $\dim \mathcal{X} = 1$, its extremities x_1 and x_2 are different, so at least one of them is nonzero. Then, imposing $x_2 \neq 0$ does not restrain the generality of our result. Following Proposition 18, $r_{\mathcal{X},\mathcal{Y}}$ is well-defined. Reusing y^* defined in (6.2), we introduce

$$x_M^*(d) := \arg\min_{x \in \mathcal{X}} \big\{\|x + y^*(x,d)\|\big\} \qquad \text{and} \qquad x_N^*(d) := \arg\max_{x \in \mathcal{X}} \big\{\|x + y^*(x,d)\| : x + y^*(x,d) \in \mathbb{R}^+ d\big\}.$$

According to Theorem 14, $x_M^*(d) \in \partial\mathcal{X}$ for all $d \in \mathbb{S}$ and following Lemma 16, x_N^* is a continuous function of d. For some $d \in \mathbb{S}$ the arg min and arg max in the definitions of x_M^* and x_N^* might not be unique; if so we take the arguments ensuring that $x_M^*(d) \in \partial\mathcal{X}$ and that x_N^* is continuous. We also define $y_N^*(d) := y^*\big(x_N^*(d), d\big)$ and $y_M^*(d) := y^*\big(x_M^*(d), d\big)$. Then,

$$r_{\mathcal{X},\mathcal{Y}}(d) = \frac{\max\limits_{y \,\in\, \mathcal{Y}} \big\{ \|y + x_N^*(d)\| : y + x_N^*(d) \in \mathbb{R}^+ d \big\}}{\max\limits_{y \,\in\, \mathcal{Y}} \big\{ \|y + x_M^*(d)\| : y + x_M^*(d) \in \mathbb{R}^+ d \big\}} = \frac{\|x_N^*(d) + y_N^*(d)\|}{\|x_M^*(d) + y_M^*(d)\|}.$$

Since $\dim \mathcal{X} = 1$, we can take $\mathcal{P}$ to be a two-dimensional plane containing $\mathcal{X}$. Then, we will study how $r_{\mathcal{X},\mathcal{Y}}(d)$ varies when d takes values in $\mathbb{S} \cap \mathcal{P}$. We introduce the signed angles $\alpha := \widehat{d, \partial\mathcal{Y}}$ and $\beta := \widehat{x_2, d}$. These angles are represented on Figure 6.1 and they take value in $[0, 2\pi)$. We parametrize all directions $d \in \mathbb{S} \cap \mathcal{P}$ by the angle β. Then, we will study how $r_{\mathcal{X},\mathcal{Y}}(d)$ varies when $\beta \in [0, 2\pi)$.

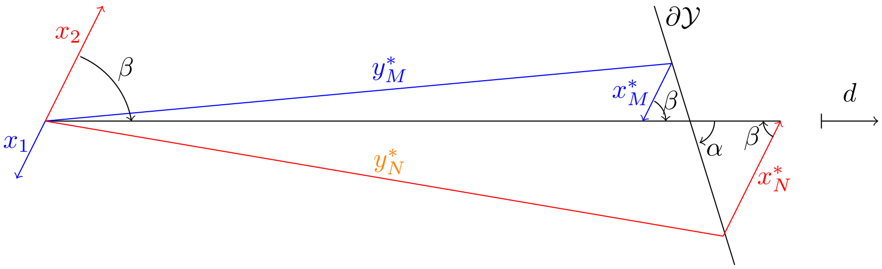


Figure 6.1: Illustration of y_N^*, x_N^*, y_M^* and x_M^* for a direction d parametrized by β.

We first establish in Lemma 4 that $x_N^*(d)$ and $x_M^*(d)$ are constant, different and both belong in $\partial\mathcal{X}$ when $y_M^*(d)$, d and $y_N^*(d)$ all intersect the same face of $\partial\mathcal{Y}$, as illustrated on Figure 6.1. In these situations, Lemma 5 shows that the ratio $r_{\mathcal{X},\mathcal{Y}}$ is constant. Thus, $r_{\mathcal{X},\mathcal{Y}}$ can only change when one of the three rays intersects a different face of $\partial\mathcal{Y}$ than the other two. We refer to these situations as vertex crossings. Lemma 6 introduces the vertices v_π and $v_{2\pi}$.

We study the crossing of vertices before v_π in Lemma 7. During these crossings Lemma 8 shows that $r_{\mathcal{X},\mathcal{Y}}$ decreases as β increases. Lemma 9 states that $r_{\mathcal{X},\mathcal{Y}}$ reaches a local minimum during the crossing of v_π. As β increases between v_π and π, Lemmas 10 and 11 prove that $r_{\mathcal{X},\mathcal{Y}}$ increases during the crossing of vertices. Finally, Lemma 12 completes the revolution by showing that $r_{\mathcal{X},\mathcal{Y}}$ decreases after $\beta = \pi$ until a local minimum at $v_{2\pi}$ and then increases again until $\beta = 2\pi$. Thus, the directions $d \in \mathcal{P} \cap \mathbb{S}$ maximizing $r_{\mathcal{X},\mathcal{Y}}(d)$ are collinear with the set $\mathcal{X}$. Note that Figure 6.1 implicitly assumes that $0 \in \mathcal{X}$. Lemma 13 proves that even if $0 \notin \mathcal{X}$ all above results still hold. Therefore, $\max\limits_{d \,\in\, \mathbb{S}} r_{\mathcal{X},\mathcal{Y}}(d) = \max\limits_{\mathcal{P}} \big\{ \max\limits_{d \,\in\, \mathcal{P} \cap \mathbb{S}} r_{\mathcal{X},\mathcal{Y}}(d) \big\} = \max \big\{ r_{\mathcal{X},\mathcal{Y}}(x_2), r_{\mathcal{X},\mathcal{Y}}(-x_2) \big\}$. $\square$

In the special case where $\mathcal{X}$ and $\mathcal{Y}$ are symmetric polytopes, this result reduces to Theorem 3.2 of [32]. Indeed, $r_{\mathcal{X},\mathcal{Y}}$ becomes an even function which leads to $r_{\mathcal{X},\mathcal{Y}}(x_2) = r_{\mathcal{X},\mathcal{Y}}(-x_2)$.

6.4 Supporting lemmata

In this section we establish all the lemmas involved in the proof of the Maximax Minimax Quotient Theorem.

Lemma 4: If d, $y_N^*(d)$ and $y_M^*(d)$ all intersect the same face of $\partial\mathcal{Y}$, then $x_N^*(d)$ and $x_M^*(d)$ are constant, different and both belong to $\partial\mathcal{X}$.

Proof. We introduce the angles $\beta_M := \widehat{x_2, y_M^*}$ and $\beta_N := \widehat{x_2, y_N^*}$. Let α_0 be the value of α when $\beta = 0$, i.e., when d is positively collinear with x_2.

We say that y_N^* is *leading* and y_M^* is *trailing* when $\beta_M < \beta_N$, and conversely when $\beta_N < \beta_M$, we say that y_M is *leading* and y_N is *trailing*.

For each $d \in \mathbb{S} \cap \mathcal{P}$ we define $D(d) := \max\limits_{y \in \mathcal{Y}} \big\{ \|y\| : y \in \mathbb{R}^+ d \big\}$, whose existence is justified by the compactness of $\mathcal{Y}$.

We say that y_N^* or y_M^* is *outside* when $\|y_N^* + x_N^*\| > D$ or $\|y_M^* + x_M^*\| > D$ respectively. Otherwise, y_N^* or y_M^* is *inside*. Directly related to the previous definition, we introduce

$$\delta_M(d) := D(d) - \|x_M^*(d) + y_M^*(d)\| \quad \text{and} \quad \delta_N(d) := \|x_N^*(d) + y_N^*(d)\| - D(d). \tag{6.3}$$

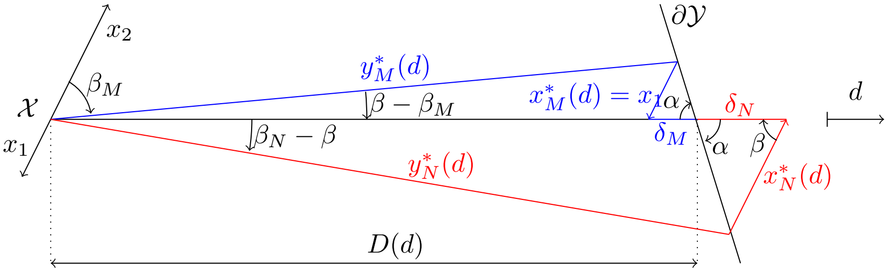


Figure 6.2: Illustration of $y_N^*(d)$ leading and outside, while $y_M^*(d)$ is trailing and inside the same face of $\partial\mathcal{Y}$.

We know from Theorem 14 that $x_M^*(d) \in \partial\mathcal{X}$ for all $d \in \mathbb{S}$. In the case illustrated on Figure 6.2, $x_M^*(d) = x_1$ because it maximizes δ_M.

If $\alpha + \beta \in \{\pi, 2\pi\}$, then $\mathcal{X}$ is parallel with a face of $\partial\mathcal{Y}$ making x_N^* and x_M^* not uniquely defined. Regardless, we can still take $x_N^*(d) \neq x_M^*(d)$, with $x_N^*(d) \in \partial\mathcal{X}$ and $x_M^*(d) \in \partial\mathcal{X}$. Otherwise, x_N^* and x_M^* are uniquely defined. Since $x_N^*(d) \in \mathcal{X}$, $x_M^*(d) \in \mathcal{X}$ for all $d \in \mathbb{S}$ and $\dim \mathcal{X} = 1$, vectors $x_N^*(d)$ and $x_M^*(d)$ are always collinear. We then use Thales's theorem and obtain $\delta_N(d) = \delta_M(d) \frac{\|x_N^*(d)\|}{\|x_M^*(d)\|}$. Since $x_N^*(d)$ is chosen to maximize δ_N and is independent from δ_M, it must have the greatest norm, so $x_N^*(d) \in \partial\mathcal{X}$. In the case where $\alpha + \beta \notin \{\pi, 2\pi\}$, $\|x + y\|$ depends on the value of x. Because $x_N^*(d)$ is chosen to maximize $\|x + y\|$ while $x_M^*(d)$ is minimizing it, we have $x_N^*(d) \neq x_M^*(d)$.

Since x_N^* is continuous according to Lemma 16 and $x_N^*(d) \in \big\{x_1, x_2\big\}$, then $x_N^*(d)$ is constant on the faces of $\partial\mathcal{Y}$. Because $x_M^*(d) \in \partial\mathcal{X}$ too, it must also be constant. □

Lemma 5: When d, $y_N^*(d)$ and $y_M^*(d)$ all intersect the same face of $\partial\mathcal{Y}$, the ratio $r_{\mathcal{X},\mathcal{Y}}(d)$ is constant.

Proof. Based on Figure 6.2 we apply the sine law in the triangle bounded by $\partial\mathcal{Y}$, δ_M and x_M^*

$$\frac{\|x_M^*(d)\|}{\sin\alpha} = \frac{\delta_M(d)}{\sin(\pi - \alpha - \beta)} = \frac{\delta_M(d)}{\sin(\alpha + \beta)}, \quad \text{so} \quad \frac{\delta_M(d)}{D(d)} = \frac{\|x_M^*(d)\| \sin(\alpha + \beta)}{D(d)\sin\alpha}.$$

Similarly for the triangle bounded by $\partial\mathcal{Y}$, δ_N and x_N^*, the law of sines yields

$$\frac{\|x_N^*(d)\|}{\sin\alpha} = \frac{\delta_N(d)}{\sin(\pi - \alpha - \beta)} = \frac{\delta_N(d)}{\sin(\alpha + \beta)}, \quad \text{so} \quad \frac{\delta_N(d)}{D(d)} = \frac{\|x_N^*(d)\| \sin(\alpha + \beta)}{D(d)\sin\alpha}.$$

Even if the two equations above were derived for the specific situation of Figure 6.2, they hold as long as y_N^*,

D and y_M^* intersect the same face of $\partial\mathcal{Y}$. Based on (6.3) we have

$$r_{\mathcal{X},\mathcal{Y}}(d) = \frac{D(d) + \delta_N(d)}{D(d) - \delta_M(d)} = \frac{1 + \frac{\delta_N}{D}}{1 - \frac{\delta_M}{D}}. \tag{6.4}$$

We will now prove that the ratios δ_N/D and δ_M/D do not change on a face of $\partial\mathcal{Y}$. Let $d_1 \in \mathcal{P} \cap \mathbb{S}$ and $d_2 \in \mathcal{P} \cap \mathbb{S}$ such that $D(d_1)$, $D(d_2)$, $y_M^*(d_1)$, $y_M^*(d_2)$, $y_N^*(d_1)$ and $y_N^*(d_2)$ all intersect the same face of $\partial\mathcal{Y}$, as illustrated on Figure 6.3.

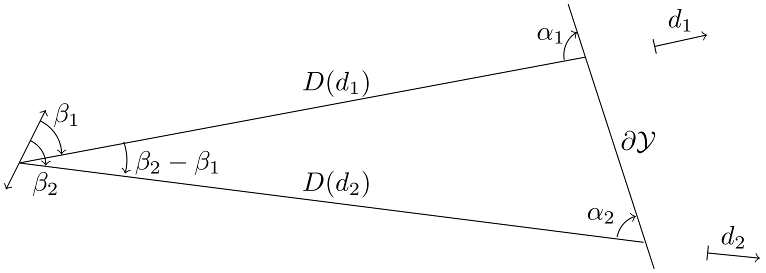


Figure 6.3: Ratio $r_{\mathcal{X},\mathcal{Y}}(d)$ is constant on a face of $\partial\mathcal{Y}$.

The sum of the angles of the triangle in Figure 6.3 is

$$(\beta_2 - \beta_1) + \alpha_2 + (\pi - \alpha_1) = \pi \qquad \text{so} \qquad \beta_2 + \alpha_2 = \beta_1 + \alpha_1. \tag{6.5}$$

Therefore, $\alpha + \beta$ is constant on faces of $\partial\mathcal{Y}$. We use the sine law in the triangle in Figure 6.3 and obtain

$$\frac{D(d_1)}{\sin\alpha_2} = \frac{D(d_2)}{\sin(\pi - \alpha_1)} = \frac{D(d_2)}{\sin\alpha_1}, \quad so, \quad D(d_1)\sin\alpha_1 = D(d_2)\sin\alpha_2.$$

According to Lemma 4 we also know that $x_N^*(d_1) = x_N^*(d_2)$, thus

$$\frac{\delta_N(d_1)}{D(d_1)} = \frac{\|x_N^*(d_1)\| \sin(\alpha_1 + \beta_1)}{D(d_1)\sin\alpha_1} = \frac{\|x_N^*(d_2)\| \sin(\alpha_2 + \beta_2)}{D(d_2)\sin\alpha_2} = \frac{\delta_N(d_2)}{D(d_2)}.$$

The same holds for δ_M/D. Hence, (6.4) yields $r_{\mathcal{X},\mathcal{Y}}(d_1) = r_{\mathcal{X},\mathcal{Y}}(d_2)$. □

Lemma 6: There are two vertices of $\mathcal{Y} \cap \mathcal{P}$, namely v_π and $v_{2\pi}$ whose crossing by d makes the angle $\alpha + \beta$ become greater than π and 2π respectively.

Proof. We have taken the convention that the angles are positively oriented in the clockwise orientation. According to (6.5), the angle $\alpha + \beta$ is constant on a face of $\partial\mathcal{Y}$. When d crosses a vertex of external angle ε as represented on Figure 6.5, the value of α has a discontinuity of $+\varepsilon$. Let q be the number of vertices of $\partial\mathcal{Y}$ and ε_i the external angle of the i^{th} vertex v_i. Since $\mathcal{Y} \cap \mathcal{P}$ is a polygon, $\sum_{i=1}^{q} \varepsilon_i = 2\pi$. We can then represent the evolution of $\alpha + \beta$ as a function of β with Figure 6.4. Instead of labeling the horizontal axis with the values taken by β as the corresponding vector $d(\beta)$ crosses the vertex v_i, we directly use v_i with a slight abuse of notation.

Recall that α_0 is the value of α when $\beta = 0$. After a whole revolution $\alpha + \beta = \alpha_0 + 2\pi$. So there are two vertices v_π and $v_{2\pi}$ where $\alpha + \beta$ first crosses π and then 2π. In the eventuality that $\alpha + \beta = \pi$ or 2π on a face of $\partial\mathcal{Y}$, we define v_π or $v_{2\pi}$ as the vertex preceding the face. □

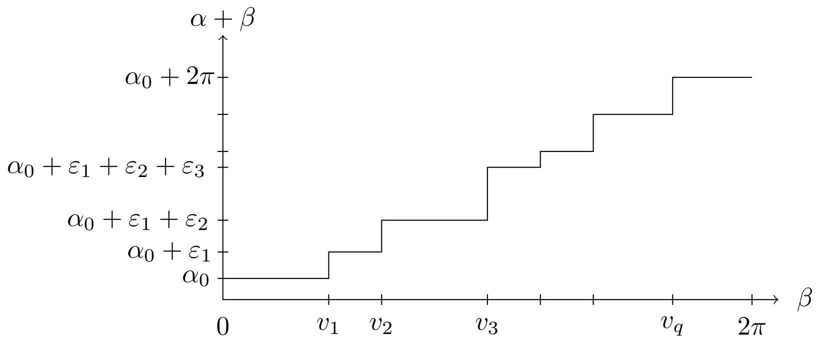


Figure 6.4: Evolution of $\alpha + \beta$ with β increasing clockwise in $[0, 2\pi)$.

Lemma 7: During the crossing of vertices before v_π as β increases, $x^*_N(d) = x_2$ and $x^*_M(d) = x_1$. They are constant, different and both belong in $\partial\mathcal{X}$.

Proof. We study the crossing of a vertex v of angle ε between the faces F_1 and F_2 of $\partial\mathcal{Y}$. For each vertex v we introduce x_v the vector collinear with $\mathcal{X}$, going from v to the ray directed by d, as illustrated on Figure 6.5 and we say that the crossing of v is ongoing as long as $\|x_v\| < \max\{\|x_1\|, \|x_2\|\}$. We also define $\delta_v := \|v + x_v\| - D$.

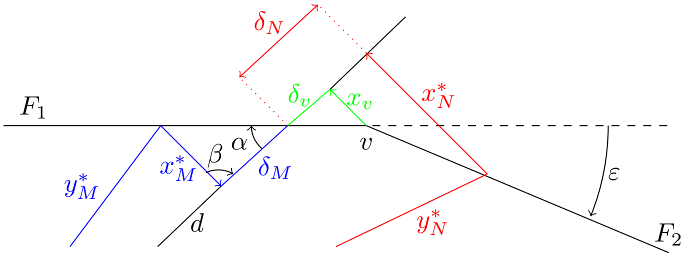


Figure 6.5: Illustration of x_v during the crossing of a vertex v, with y^*_N leading.

Before starting the crossing of v_π we have $\alpha + \beta \in (\alpha_0, \pi)$. This situation is depicted on Figure 6.2, where y^*_N is leading and outside, so y^*_N reaches the vertex before y^*_M and d. The length of $x^*_N(d)$ can vary to maximize δ_N, so y^*_N could still intersect F_1, even if the crossing is ongoing. We have seen in Lemma 4 that if y^*_N is still on F_1, then it must be the furthest possible to maximize δ_N, in that case $y^*_N = v$. Otherwise, y^*_N intersects F_2. We want to establish a criterion to distinguish these two possible scenarios.

We first consider the scenario where $y^*_N = v$ and $x^*_N(d) = x_v$. We take $y \in F_2 \backslash \{v\}$ such that $x_2 + y \in \mathbb{R}^+ d$ as represented on Figure 6.6 and we define $\delta := \|x_2 + y\| - D$.

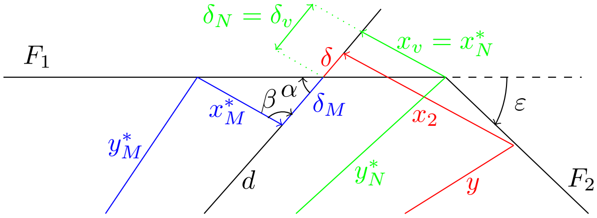


Figure 6.6: Illustration of the crossing scenario where $y^*_N = v$.

Since δ_N must be maximized by the choice of y^*_N and $y \neq y^*_N$, we have $\delta < \delta_N = \delta_v$. But $\|x_2\| > \|x_v\|$, so the line segment corresponding to x_2 crosses the interior of $\mathcal{Y}$. Focusing on this part of Figure 6.6 we obtain Figure 6.7.

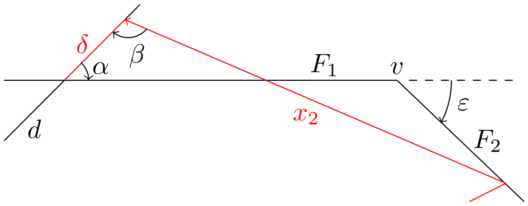


Figure 6.7: Illustration of the line segment corresponding to x_2 crossing the interior of $\mathcal{Y}$ in Figure 6.6.

Two of the angles of the triangle delimited by F_1, F_2 and x_2 are $\pi - \alpha - \beta$ and $\pi - \varepsilon$. Therefore, their sum is in $(0, \pi)$ and thus $\alpha + \beta + \varepsilon > \pi$. Since we assumed that $\alpha + \beta \in (\alpha_0, \pi)$, the vertex v must in fact be v_π for this scenario to happen.

Thus, the crossing of a vertex preceding v_π follows the second scenario as depicted on Figure 6.5 with $y_N^* \in F_2$. We study Figure 6.8 which is a more detailed view of Figure 6.5, with δ_0 depending solely on d and ε.

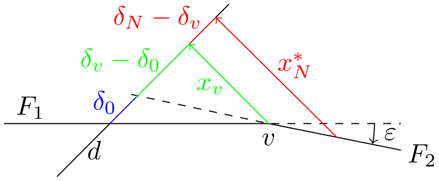


Figure 6.8: Illustration of x_v and x_N^* in Figure 6.5.

Since x_v and $x_N^*(d)$ are collinear, we can apply Thales's theorem in Figure 6.8 and obtain that

$$\delta_N - \delta_0 = (\delta_v - \delta_0)\frac{\|x_N^*(d)\|}{\|x_v(d)\|}.$$

Then, δ_N is maximized when $\|x_N^*(d)\|$ is maximal, so $x_N^*(d) = x_2$ during the crossing. We know from Theorem 14 that $x_M^*(d) \in \partial\mathcal{X}$ for all $d \in \mathbb{S}$. Then, as in Lemma 4, x_N^* and x_M^* are constant and different since x_N^* is continuous in d, so $x_M^*(d) = x_1$. □

Lemma 8: During the crossing of vertices before v_π as β increases, $r_{\mathcal{X},\mathcal{Y}}(d)$ decreases.

Proof. The leading vector y_N^* is outside and crosses a vertex v between the faces F_1 and F_2 of $\partial\mathcal{Y}$ while β increases. We separate the vertex crossing into two parts: when only $y_N^* \in F_2$, and when both $d \in F_2$ and $y_N^* \in F_2$. Let $\varepsilon > 0$ be the external angle of the vertex as shown on Figure 6.9.

According to Lemma 5, $r_{\mathcal{X},\mathcal{Y}}$ is constant on faces of $\partial\mathcal{Y}$ and we call r_{F_1} its value on the face F_1. If F_1 was prolonged past v with a straight line (dashed line on Figure 6.9), then we would have $y_N^*(d) \in F_1$ and $r_{\mathcal{X},\mathcal{Y}}(d) = r_{F_1}$. But, $y_N^*(d) \in F_2$ as proven in Lemma 7 because the crossing occurs before v_π. We call l the resulting difference in δ_N as illustrated on Figure 6.9. Notice that the two green segments of length l in Figure 6.9 are parallel. We parametrize the position of y_N^* on F_2 with the length m as defined on Figure 6.9.

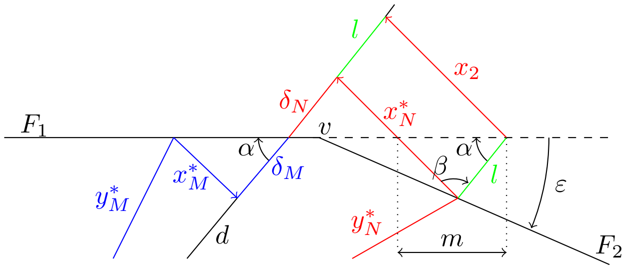


Figure 6.9: Part I of the crossing of vertex v by y_N^* leading and outside as β increases.

When $y_N^* = v$, $m = 0$, and m increases with β. Using the sine law we obtain

$$\frac{m}{\sin\beta} = \frac{l}{\sin(\pi - \alpha - \beta)} = \frac{l}{\sin(\alpha + \beta)}. \tag{6.6}$$

Then,

$$r_{\mathcal{X},\mathcal{Y}}(d) = \frac{D + \delta_N}{D - \delta_M} = \frac{D + \delta_N + l}{D - \delta_M} - \frac{l}{D - \delta_M} = r_{F_1} - \frac{m\sin(\alpha + \beta)}{(D - \delta_M)\sin(\beta)}. \tag{6.7}$$

By definition the length m is positive. Since $-x_M^* \in \operatorname{int}(\mathcal{Y})$ but $y_M^* \in \partial\mathcal{Y}$, we have $D - \delta_M = \|y_M^* + x_M^*\| > 0$. Before v_π we have $\alpha + \beta \in (\alpha_0, \pi)$. In that case $\sin(\alpha+\beta) > 0$ and $\sin(\beta) > 0$. Therefore, the term subtracted from r_{F_1} is positive, i.e., $r_{\mathcal{X},\mathcal{Y}}(d) < r_{F_1}$.

We can now tackle the second part of the crossing, when y_N^* and d both have crossed the vertex as illustrated on Figure 6.10.

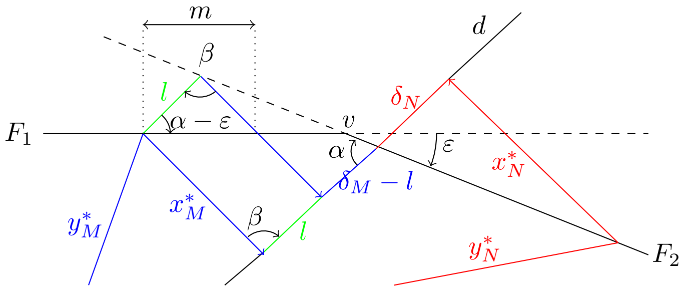


Figure 6.10: Part II of the crossing of vertex v by y_N^* leading and outside as β increases.

If F_2 was prolonged with a straight line before v and $y_M^* \in F_2$, then we would have $r_{\mathcal{X},\mathcal{Y}}(d) = r_{F_2}$, value of $r_{\mathcal{X},\mathcal{Y}}$ on F_2. But that is not the case, $y_M^*(d) \in F_1$ and the resulting difference in δ_M is called l. Using the sine law in Figure 6.10, we can relate l to m

$$\frac{m}{\sin\beta} = \frac{l}{\sin(\pi - \beta - \alpha + \varepsilon)} = \frac{l}{\sin(\alpha + \beta - \varepsilon)}. \tag{6.8}$$

We have $\alpha + \beta \in (\alpha_0, \pi)$, so $\sin(\beta) > 0$. If α was still measured between d and F_1, then its value would be $\alpha_{F_1} = \alpha - \varepsilon$. Since we are before the crossing of v_π, $\alpha_{F_1} + \beta \in (\alpha_0, \pi)$, i.e., $\alpha + \beta - \varepsilon \in (\alpha_0, \pi)$. This yields $\sin(\alpha + \beta - \varepsilon) > 0$, which makes $l > 0$, because the length m is positive by definition. Then,

$$r_{F_2} = \frac{D + \delta_N}{D - (\delta_M - l)} = \frac{D + \delta_N}{D - \delta_M + l} < \frac{D + \delta_N}{D - \delta_M} = r_{\mathcal{X},\mathcal{Y}}(d). \tag{6.9}$$

Thus, the ratio $r_{\mathcal{X},\mathcal{Y}}$ decreases during the crossing of a vertex before v_π. $\square$

Lemma 9: During the crossing of v_π, the ratio $r_{\mathcal{X},\mathcal{Y}}(d)$ reaches a local minimum.

Proof. Recall that before the crossing, $x_N^*(d) = x_2$ and $x_M^*(d) = x_1$. During the crossing of v_π, i.e., when $\|x_{v_\pi}\| < \max\{\|x_1\|, \|x_2\|\}$, we have $\alpha + \beta \leq \pi$ but $\alpha + \beta + \varepsilon > \pi$. The situation is illustrated on Figure 6.11. We showed in Lemma 7 that $y_N^* = v_\pi$ and $x_N^*(d) = x_{v_\pi}$.

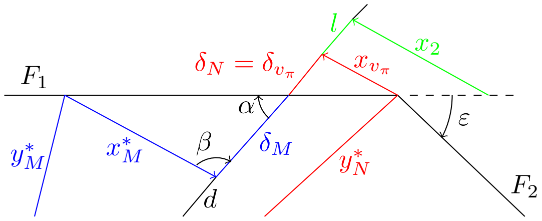


Figure 6.11: Crossing of v_π, with $y_N^* = v_\pi$.

If F_1 was prolonged with a straight line (dashed line of Figure 6.11), we would have $y_N^* \neq v_\pi$, $x_N^*(d) = x_2$ and the ratio would be $r_{F_1} = \frac{D+\delta_{v_\pi}+l}{D-\delta_M}$, which is the value of $r_{\mathcal{X},\mathcal{Y}}$ on F_1. Since d has not yet crossed v_π, $\alpha + \beta < \pi$ and thus (6.6) and (6.7) still hold, leading to $r_{\mathcal{X},\mathcal{Y}}(d) < r_{F_1}$.

Once d has crossed v_π, we still have $y_N^* = v_\pi$ to maximize δ_N. Then, the equality $x_N^*(d) = x_{v_\pi}$ holds during the whole crossing, i.e., as x_{v_π} goes from x_2 to x_1. The second part of the crossing is illustrated on Figure 6.12.

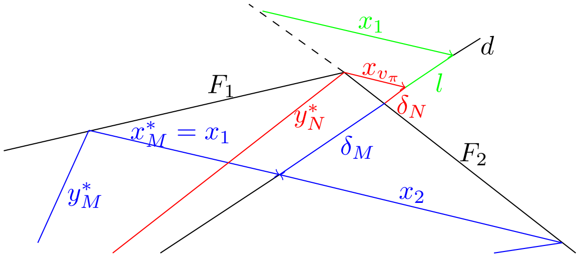


Figure 6.12: Illustration of the endpoint of y_M^* switching from F_1 to F_2 during the crossing of v_π.

Assume that during the entire crossing of v_π, $x_M^*(d) = x_1$. Then, at the end of the crossing we will have $y_M^* = v_\pi$ and $x_M^*(d) = x_{v_\pi} = x_N^*(d)$, which contradicts the definitions of $x_M^*(d)$ and $x_N^*(d)$, they must be different. Thus, $x_M^*(d)$ does not remain equal to x_1 during the entire crossing. Since $x_M^* \in \{x_1, x_2\}$, at some point x_M^* switches to x_2 as y_M^* switches from F_1 to F_2. This switching point is illustrated on Figure 6.12, and y_M^* becomes the leading vector.

After this switch, $y_M^* \in F_2$ and $x_M^*(d) = x_2$. If F_2 was prolonged with the dashed line on Figure 6.12, we would have $x_N^* = x_1$ instead of x_{v_π} with a gain of l for δ_N making the ratio equal to $r_{F_2} = \frac{D+\delta_N+l}{D-\delta_M}$, value of $r_{\mathcal{X},\mathcal{Y}}$ on F_2. But $x_N^* = x_{v_\pi}$ and $l > 0$, thus $r_{F_2} > \frac{D+\delta_N}{D-\delta_M} = r_{\mathcal{X},\mathcal{Y}}(d)$. Therefore, $r_{\mathcal{X},\mathcal{Y}}$ reaches a local minimum during the crossing of v_π. $\square$

Lemma 10: During the crossing of vertices after v_π as β increases until π, $x_N^*(d) = x_1$ and $x_M^*(d) = x_2$. They are constant different and both in $\partial\mathcal{X}$.

Proof. After the crossing of v_π, $\alpha + \beta \in (\pi, \alpha_0 + \pi)$ and y_M^* is leading and inside as established in Lemma 9. Thus, y_M^* is the first to reach vertex v. Since $x_M^* \in \{x_1, x_2\}$ we cannot have $x_M^* = x_v$ during the entire crossing because x_v is a continuous function of β. Thus y_M^* passes v and belongs to F_2. In Lemma 16 we showed that x_N^* is continuous in d. Thus, $x_N^*(d)$ cannot switch like $x_M^*(d)$ did around v_π to take the lead. Instead, $x_N^*(d)$ is trailing as illustrated on Figure 6.13.

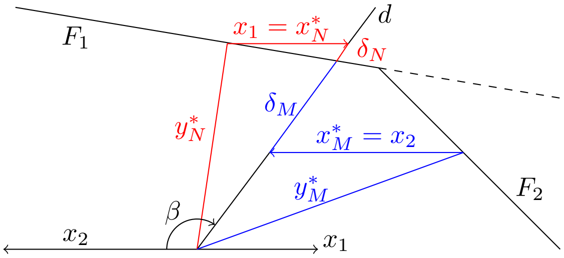


Figure 6.13: Crossing of a vertex v after v_π.

Since $y_N^* \in F_1$ during the crossing, we can apply Thales's theorem on Figure 6.13 and obtain that for a fixed d, δ_N is proportional to $\|x_N^*(d)\|$. Thus, to maximize δ_N we have $x_N^*(d) \in \partial\mathcal{X}$ and, since y_N^* is trailing, we have $x_N^*(d) = x_1$ during the entire crossing. By the definitions of $x_N^*(d)$ and $x_M^*(d)$, we have $x_N^*(d) \neq x_M^*(d)$. Since both $x_N^*(d)$ and $x_M^*(d)$ belong to $\partial\mathcal{X} = \{x_1, x_2\}$, then $x_M^*(d) = x_2$ during the entire crossing. □

Lemma 11: During the crossing of vertices after v_π as β increases until π, $r_{\mathcal{X},\mathcal{Y}}(d)$ increases.

Proof. The leading vector y_M^* is inside and crosses a vertex v between faces F_1 and F_2 as β increases. We define $\beta' := \pi - \beta$. Then, reversing the crossing illustrated on Figure 6.13 is exactly the crossing illustrated on Figure 6.9 with β' increasing and x_1 and x_2 exchanged. According to Lemma 8, in that reversed crossing $r_{\mathcal{X},\mathcal{Y}}$ is decreasing. Therefore, $r_{\mathcal{X},\mathcal{Y}}$ increases during the crossing of vertices after v_π as β increases until π. □

Lemma 12: For $\beta > \pi$, $r_{\mathcal{X},\mathcal{Y}}(d)$ decreases until $v_{2\pi}$ where it reaches a local minimum. After $v_{2\pi}$ as β increases until 2π, $r_{\mathcal{X},\mathcal{Y}}(d)$ increases.

Proof. Let us change the angle convention, so that angles are now positively oriented in the counterclockwise orientation. The vertex that was previously labeled as $v_{2\pi}$ becomes the new v_π. Then, we only need to apply Lemmas 7, 8, 9, 10 and 11 to this new configuration to conclude the proof. □

Lemma 13: All above results hold even if $0 \notin \mathcal{X}$.

Proof. In all the figures we made the implicit assumption that $0 \in \mathcal{X}$, so that x_1 and x_2 were negatively collinear. Let x_1 be positively collinear with x_2 and $\|x_2\| > \|x_1\|$.

On Figure 6.2, we would now have $y_N^*(d)$ and $y_M^*(d)$ both outside. Then, the definition of δ_M should be adapted. Let $\delta_M(d) := \|x_M^*(d) + y_M^*(d)\| - D(d)$ and then $r_{\mathcal{X},\mathcal{Y}}(d) = \frac{D+\delta_N}{D+\delta_M}$. Except for this modification, we would still have $x_N^*(d) = x_2$ and $x_M^*(d) = x_1$. Thales theorem can be used similarly to show that $x_N^*(d) \in \partial\mathcal{X}$. Therefore, Lemma 4 holds.

In the proof of Lemma 5 we still have δ_N/D and δ_M/D invariant with respect to d on a given face of $\partial\mathcal{Y}$, so $r_{\mathcal{X},\mathcal{Y}}$ is still constant on faces. Lemma 6 is not affected at all. The first part of the crossing of a vertex before v_π as β increases is illustrated by Figure 6.14.

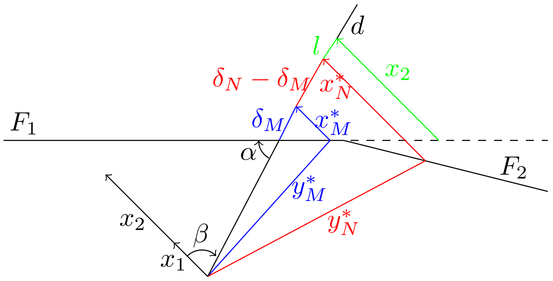


Figure 6.14: Part I of the crossing of a vertex before v_π with $0 \notin \mathcal{X}$.

For δ_M to be minimized and δ_N to be maximized, the Thales theorem clearly proves that $x_M^* \in \partial\mathcal{X}$ and $x_N^* \in \partial\mathcal{X}$ during the crossing. We still have $x_N^*(d) = x_2$ and $x_M^*(d) = x_1$, so Lemma 7 holds.

Following the reasoning in Lemma 8, we have $l > 0$, which leads to

$$r_{F_1} = \frac{D + \delta_N + l}{D + \delta_M} > \frac{D + \delta_N}{D + \delta_M} = r_{\mathcal{X},\mathcal{Y}}(d).$$

During the second part, both $y_N^* \in F_2$ and $y_M^* \in F_2$ but $d \in F_1$. This situation is illustrated on Figure 6.15.

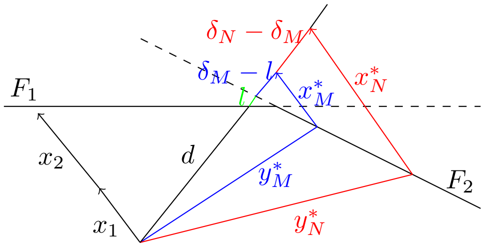


Figure 6.15: Part II of the crossing of a vertex before v_π with $0 \notin \mathcal{X}$.

We compare the current value of $r_{\mathcal{X},\mathcal{Y}}(d)$ with r_{F_2}, its value on F_2:

$$r_{F_2} = \frac{D + (\delta_N - l)}{D + (\delta_M - l)} \qquad \text{and} \qquad r_{\mathcal{X},\mathcal{Y}}(d) = \frac{D + \delta_N}{D + \delta_M}.$$

Since $l > 0$ and $\delta_N > \delta_M$, a simple calculation shows that $r_{\mathcal{X},\mathcal{Y}}(d) < r_{F_2}$. Therefore, $r_{\mathcal{X},\mathcal{Y}}$ is decreasing during the crossing of a vertex before v_π as β increases, Lemma 8 holds.

During the crossing of v_π, $y_N^* = v_\pi$ and $x_N^* = x_{v_\pi}$ with its norm decreasing continuously until $x_N^* = x_1$, while x_M^* will switch to x_2 in order to minimize δ_M. This is the same process as described in Lemma 9, so $r_{\mathcal{X},\mathcal{Y}}$ also reaches a local minimum.

Because all the results studied so far still hold, then Lemmas 10, 11 and 12 hold too because they rely on those earlier results. □

We have now established all the lemmas directly involved in the proof of the Maximax Minimax Quotient Theorem, but we still have a few claims of continuity to prove.

6.5 Continuity of extrema

In Proposition 18 (iii) we needed the continuity of λ^* to prove it has a minimum and in Lemma 4 we used the continuity of x_N^* and y_N^*. In this section we will thus prove the continuity of these two maxima functions relying on the Berge Maximum Theorem [1].

Lemma 14: Let $\mathcal{X}$ and $\mathcal{Y}$ be two nonempty polytopes in $\mathbb{R}^n$ with $-\mathcal{X} \subset \mathcal{Y}$. Then, the set-valued function $\varphi : \mathcal{X} \times \mathbb{S} \rightrightarrows \mathcal{Y}$ defined as $\varphi(x,d) := \mathcal{Y} \cap \big\{\lambda d - x : \lambda \geq 0\big\}$ satisfies Definition 17.2 of [1].

Proof. We define $\Omega := \mathcal{X} \times \mathbb{S}$, so that $\varphi : \Omega \rightrightarrows \mathcal{Y}$. On the space Ω we introduce the norm $\|\cdot\|_\Omega$ as $\|(x,d)\|_\Omega = \|x\| + \|d\|$. Since $\|\cdot\|$ is the Euclidean norm, $\|\cdot\|_\Omega$ is a norm on Ω. By Definition 17.2 of [1], we need to prove that φ is both upper and lower hemicontinuous at all points of Ω.

First, using Lemma 17.5 of [1] we will prove that φ is lower hemicontinuous by showing that for an open subset $\mathcal{A}$ of $\mathcal{Y}$, $\varphi^l(\mathcal{A})$ is open. The lower inverse image of $\mathcal{A}$ is defined in [1] as

$$\begin{aligned}\varphi^l(\mathcal{A}) &:= \big\{\omega \in \Omega : \varphi(\omega) \cap \mathcal{A} \neq \emptyset\big\} \\ &= \big\{(x,d) \in \mathcal{X} \times \mathbb{S} : \mathcal{Y} \cap \{\lambda d - x : \lambda \geq 0\} \cap \mathcal{A} \neq \emptyset\big\} \\ &= \big\{(x,d) \in \mathcal{X} \times \mathbb{S} : \{\lambda d - x : \lambda \geq 0\} \cap \mathcal{A} \neq \emptyset\big\},\end{aligned}$$

because $\mathcal{A} \subseteq \mathcal{Y}$. Let $\omega = (x,d) \in \varphi^l(\mathcal{A})$. Then, there exists $\lambda \geq 0$ such that $\lambda d - x \in \mathcal{A}$. Since $\mathcal{A}$ is open, there exists $\varepsilon > 0$ such that the ball $B_\varepsilon(\lambda d - x) \subset \mathcal{A}$. Now let $\omega_1 = (x_1, d_1) \in \Omega$ and denote $\varepsilon_x := \|x_1 - x\|$ and $\varepsilon_d := \|d_1 - d\|$. Then,

$$\|\lambda d_1 - x_1 - (\lambda d - x)\| = \|\lambda(d_1 - d) - (x_1 - x)\| \leq \lambda \varepsilon_d + \varepsilon_x.$$

Since $\lambda \geq 0$ is fixed, we can choose ε_d and ε_x positive and small enough so that $\lambda\varepsilon_d + \varepsilon_x \leq \varepsilon$. Then, we have showed that for all $\omega_1 = (x_1, d_1) \in \Omega$ such that $\|\omega - \omega_1\|_\Omega \leq \min(\varepsilon_d, \varepsilon_x)$, i.e., such that $\|x_1 - x\| \leq \varepsilon_x$ and $\|d_1 - d\| \leq \varepsilon_d$, we have $\lambda d_1 - x_1 \in B_\varepsilon(\lambda d - x) \subset \mathcal{A}$, i.e., $\omega_1 \in \varphi^l(\mathcal{A})$. Therefore, $\varphi^l(\mathcal{A})$ is open, and so φ is lower hemicontinuous.

To prove the upper hemicontinuity of φ, we will use Lemma 17.4 of [1] and prove that for a closed subset $\mathcal{A}$ of $\mathcal{Y}$, the lower inverse image of $\mathcal{A}$ is closed. Let $\{\omega_k\}$ be a sequence in $\varphi^l(\mathcal{A})$ converging to $\omega = (x,d) \in \Omega$. We want to prove that the limit $\omega \in \varphi^l(\mathcal{A})$.

For $k \geq 0$, we have $\omega_k = (x_k, d_k)$ and define $\Lambda_k := \big\{\lambda_k \geq 0 : \lambda_k d_k - x_k \in \mathcal{A}\big\} \neq \emptyset$. Since $\mathcal{A}$ is a closed subset of the compact set $\mathcal{Y}$, then $\mathcal{A}$ is compact. Thus Λ_k has a minimum and a maximum; we denote them by λ_k^{min} and λ_k^{max} respectively.

Since sequences $\{d_k\}$ and $\{x_k\}$ converge, they are bounded. The set $\mathcal{A}$ is also bounded, thus sequence $\{\lambda_k^{max}\}$ is bounded. Let $\lambda^{max} := \sup\limits_{k \geq 0} \lambda_k^{max} > 0$.

For $k \geq 0$, we define segments $S_k := \big\{\lambda d_k - x_k : \lambda \in [0, \lambda^{max}]\big\}$, and $S := \big\{\lambda d - x : \lambda \in [0, \lambda^{max}]\big\}$. These segments are all compact sets. We also introduce the sequences $a_k := \lambda_k^{min} d_k - x_k \in \mathcal{A} \cap S_k$ and $b_k := \lambda_k^{min} d - x \in S$.

Take $\varepsilon > 0$. Since sequences $\{d_k\}$ and $\{x_k\}$ converge toward d and x respectively, there exists $N \geq 0$ such that for $k \geq N$, we have

$$\|d_k - d\| \leq \frac{\varepsilon}{2\lambda^{max}} \qquad \text{and} \qquad \|x_k - x\| \leq \frac{\varepsilon}{2}.$$

Then, for any $\lambda_k \in [0, \lambda^{max}]$ as

$$\|\lambda_k d_k - x_k - (\lambda_k d - x)\| = \|\lambda_k(d_k - d) - (x_k - x)\| \leq \lambda_k \frac{\varepsilon}{2\lambda^{max}} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Since $\lambda_k^{min} \in [0, \lambda^{max}]$, we have $\|a_k - b_k\| \xrightarrow[k\to\infty]{} 0$. We define the distance between the sets $\mathcal{A}$ and S

$$dist(\mathcal{A}, S) := \min\big\{\|a - s_\lambda\| : a \in \mathcal{A},\ s_\lambda \in S\big\}.$$

The minimum exists because $\mathcal{A}$ and S are both compact and the norm is continuous. Since $a_k \in \mathcal{A}$ and $b_k \in S$, we have $dist(\mathcal{A}, S) \leq \|a_k - b_k\|$ for all $k \geq 0$. Therefore, $dist(\mathcal{A}, S) = 0$. So, $\mathcal{A} \cap S \neq \emptyset$, leading to $\omega \in \varphi^l(\mathcal{A})$. Then, $\varphi^l(\mathcal{A})$ is closed and so φ is upper hemicontinuous. □

Lemma 15: Let $\mathcal{X}$ and $\mathcal{Y}$ be two nonempty polytopes in $\mathbb{R}^n$ with $-\mathcal{X} \subset \mathcal{Y}$. Then,

$$\lambda^*(x, d) := \max_{y \,\in\, \mathcal{Y}} \big\{\|x + y\| : x + y \in \mathbb{R}^+ d\big\}$$

is continuous in $x \in \mathcal{X}$ and $d \in \mathbb{S}$.

Proof. According to Proposition 18 (ii), whose proof does not rely on the current lemma, λ^* is well-defined. We introduce the set-valued function $\varphi : \mathcal{X} \times \mathbb{S} \rightrightarrows \mathcal{Y}$ defined by

$$\varphi(x, d) := \big\{y \in \mathcal{Y} : x + y \in \mathbb{R}^+ d\big\} = \mathcal{Y} \cap \big(\mathbb{R}^+ d - \{x\}\big),$$

where $\mathbb{R}^+ d - \{x\} = \big\{\lambda d - x : \lambda \geq 0\big\}$.

We define the graph of φ as $\operatorname{Gr}\varphi := \big\{(x, d, y) \in \mathcal{X} \times \mathbb{S} \times \mathcal{Y} : y \in \varphi(x, d)\big\}$, and the continuous function $f : \operatorname{Gr}\varphi \to \mathbb{R}^+$ as $f(x, d, y) = \|x + y\|$. Set $\mathcal{X} \times \mathbb{S}$ is compact and nonempty. Since $\mathcal{Y}$ is compact and $\mathbb{R}^+ d - \{x\}$ is closed, their intersection $\varphi(x, d)$ is compact. Because $-\mathcal{X} \subset \mathcal{Y}$, for all $x \in \mathcal{X}$ we have $-x \in \varphi(x, d)$, so $\varphi(x, d) \neq \emptyset$. According to Lemma 14, φ satisfies Definition 17.2 of [1]. Then, we can apply the Berge Maximum Theorem [1] and conclude that λ^* is continuous in x and d. □

Lemma 16: Let $\mathcal{X}$ and $\mathcal{Y}$ be two nonempty polytopes in $\mathbb{R}^n$ with $-\mathcal{X} \subset \mathcal{Y}$. Then, the functions

$$\big(x_N^*, y_N^*\big)(d) = \arg \max_{x \,\in\, \mathcal{X},\, y \,\in\, \mathcal{Y}} \big\{\|x + y\| : x + y \in \mathbb{R}^+ d\big\}$$

are continuous in $d \in \mathbb{S}$.

Proof. Let $\mathcal{Z} := \mathcal{X} \oplus \mathcal{Y} = \big\{x + y : x \in \mathcal{X},\ y \in \mathcal{Y}\big\}$. Then $\mathcal{Z}$ is the Minkowski sum of two polytopes, so it is also a polytope [120]. According to Proposition 18 (i), whose proof does not rely on the current lemma, $\max\limits_{x \,\in\, \mathcal{X},\, y \,\in\, \mathcal{Y}} \big\{\|x + y\| : x + y \in \mathbb{R}^+ d\big\}$ exists and thus $\max\limits_{z \,\in\, Z} \big\{\|z\| : z \in \mathbb{R}^+ d\big\}$ is also well-defined.

Since $-\mathcal{X} \subset \mathcal{Y}$, for all $x \in \mathcal{X}$, $-x \in \mathcal{Y}$ and thus $0 \in \mathcal{Z}$. Then, $\{0\}$ and $\mathcal{Z}$ are two polytopes in $\mathbb{R}^n$ with $\pm 0 \in \mathcal{Z}$. According to Lemmma 15 the function $\lambda^*(0, d) := \max\limits_{z \,\in\, \mathcal{Z}} \big\{\|z + 0\| : z + 0 \in \mathbb{R}^+ d\big\}$ is continuous in $d \in \mathbb{S}$.

Then, we define the continous function $z(d) := \lambda^*(0, d) d \in \mathcal{Z}$ for $d \in \mathbb{S}$. Note that $z(d) = \arg\max\limits_{z \,\in\, \mathcal{Z}} \big\{\|z\| : z \in \mathbb{R}^+ d\big\} = \big(x_N^*, y_N^*\big)(d)$, so these functions are continuous. □

6.6 Illustration

We will now illustrate the Maximax Minimax Quotient Theorem on a simple example. We consider polygon $\mathcal{X}$ delimited by the vertices $x_1 = (0, -0.5)$ and $x_2 = (0, 1)$ in $\mathbb{R}^2$ and polygon $\mathcal{Y}$ with vertices $(\pm 1, \pm 2)$ and $(\pm 3, 0)$ as represented on Figure 6.16.

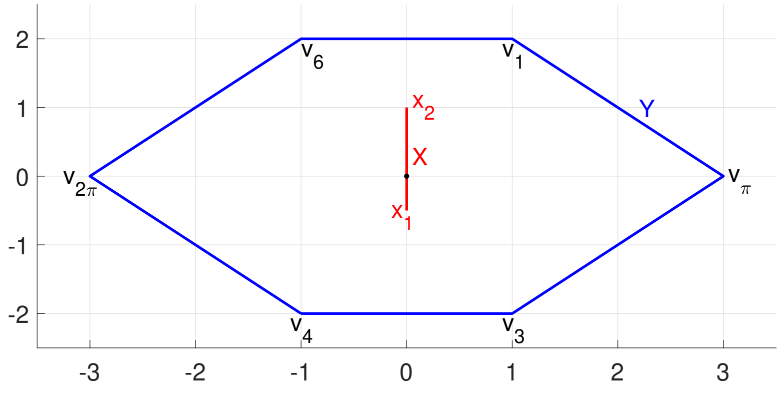


Figure 6.16: Illustration of polygons $\mathcal{X}$ and $\mathcal{Y}$.

Since $-\mathcal{X} \subset \text{int}(\mathcal{Y})$, $\dim \mathcal{X} = 1$, $x_2 \neq 0$ and $\dim \mathcal{Y} = 2$, the assumptions of the Maximax Minimax Quotient Theorem are satisfied. To illustrate the proof of the theorem, for all $d \in \mathbb{S}$ we define the angle $\beta := \widehat{x_2, d}$ positively oriented clockwise. We also enumerate the vertices in the clockwise direction and we note that $v_2 = v_\pi$ and $v_5 = v_{2\pi}$ as defined in Lemma 6. Then, we compute $r_{\mathcal{X},\mathcal{Y}}$ for $\beta \in [0, 2\pi)$ as shown on Figure 6.17. The red spikes denote when the ray $d(\beta)$ hits a vertex of $\mathcal{Y}$.

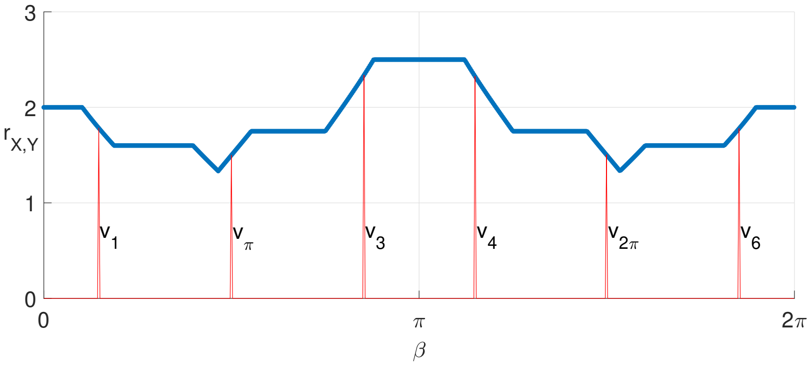


Figure 6.17: Graph of $r_{\mathcal{X},\mathcal{Y}}$ as a function of β.

As demonstrated by the Maximax Minimax Quotient Theorem, $r_{\mathcal{X},\mathcal{Y}}$ has two local maxima achieved at $\beta = 0$ and $\beta = \pi$. These two values are different because polygon $\mathcal{X}$ is not symmetric. Note also that the Maximax Minimax Quotient Theorem does not state that the maximum is *only* reached when $\beta \in \{0, \pi\}$. Indeed as shown in Figure 6.17 and established in Lemma 5, $r_{\mathcal{X},\mathcal{Y}}$ is constant on the faces of $\partial\mathcal{Y}$. Thus, the two local maxima are achieved on the faces $[v_1, v_6]$ and $[v_3, v_4]$. As proven in Lemma 9 and in Lemma 12, $r_{\mathcal{X},\mathcal{Y}}$ reaches a local minimum during the crossing of the vertices v_π and $v_{2\pi}$.

A video illustrating the Maximax Minimax Quotient Theorem on a different polytope can be found following the link here or in the footnote[1].

6.7 Summary

In this chapter we considered an optimization problem arising from optimal control and pertaining to both fractional programming and max-min programming. We first justified the existence of the Maximax Minimax Quotient. Then, relying on numerous geometrical arguments and on the continuity of two maxima functions we were able to establish the Maximax Minimax Quotient Theorem. This result provides an analytical solution to the maximization of a ratio of a maximum and a minimax over two polytopes. We illustrated our theorem and its proof on a simple example in $\mathbb{R}^2$. This work filled the theoretical gap left in Chapter 5, and because of our less restrictive assumptions we also open the way for a more general framework.

[1] https://www.youtube.com/watch?v=rjKzHyDJX40

Chapter 7

Resilience of Linear Systems with Bounded Amplitude

7.1 Introduction

This chapter focuses on extending the results on resilience and quantitative resilience from the driftless dynamics of Chapter 5 to full-fledged linear systems and is based on our works [33], [37]. The main contributions of this chapter are fourfold. Firstly, relying on the differential games theory of Hájek [23] and the controllability conditions of Brammer [91], we establish simple necessary and sufficient conditions to verify the resilient stabilizability of linear systems, i.e., whether the origin is resiliently reachable from any initial state. Secondly, we extend Hájek's duality theorem in order to study the resilient reachability of affine targets. Thirdly, we use zonotopic underapproximations of reachable sets [61], [82] to determine what states are guaranteed to be resiliently reachable. Finally, we employ Lyapunov theory [121] to establish analytical bounds on the quantitative resilience of linear systems.

This chapter is organized as follows. Section 7.2 introduces the system dynamics and the problems of interest. Section 7.3 provides background results. Section 7.4 establishes necessary and sufficient conditions for resilient stabilizability of linear systems. Section 7.5 extends these conditions to affine targets and describes zonotopic underapproximations of the resiliently reachable set of linear systems. Section 7.6 derives analytical bounds on the quantitative resilience of linear systems. Section 7.7 illustrates our theory on a fighter jet model and a temperature control system.

7.2 Problem statement

We consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + \bar{B}\bar{u}(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \qquad \bar{u}(t) \in \bar{\mathcal{U}}, \tag{7.1}$$

with constant matrices $A \in \mathbb{R}^{n \times n}$ and $\bar{B} \in \mathbb{R}^{n \times (m+p)}$. The admissible controls are assumed to be in $\bar{\mathcal{U}} := [-1, 1]^{m+p}$, in line with previous works [33], [71], [121].

After a loss of control authority over p of the $m + p$ actuators of system (7.1), the input signal $\bar{u}$ is split between the undesirable input signal $w \in \mathcal{F}(\mathcal{W})$, $\mathcal{W} := [-1, 1]^p$, and the controlled input signal $u \in \mathcal{F}(\mathcal{U})$,

$\mathcal{U} := [-1,1]^m$. Matrix $\bar{B}$ is accordingly split in $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{n \times p}$ so that the dynamics become

$$\dot{x}(t) = Ax(t) + Bu(t) + Cw(t), \qquad x(0) = x_0 \in \mathbb{R}^n. \tag{7.2}$$

We want to study how the partial loss of control authority affects the *stabilizability* and the *controllability* of the nominal dynamics.

Definition 13: System (7.1) is *stabilizable* (resp. *controllable*) if there exists an admissible control signal $\bar{u} \in \mathcal{F}(\bar{\mathcal{U}})$ driving the state of system (7.1) from any $x_0 \in \mathbb{R}^n$ to $0 \in \mathbb{R}^n$ (resp. to any $x_{tg} \in \mathbb{R}^n$).

To adapt these two properties to system (7.2), we first need the notion of *resilient reachability* introduced in [31].

Definition 14: A target $x_{tg} \in \mathbb{R}^n$ is *resiliently reachable* from $x_0 \in \mathbb{R}^n$ by system (7.2) if for all $w \in \mathcal{F}(\mathcal{W})$, there exists $T \geq 0$ and $u \in \mathcal{F}(\mathcal{U})$ such that $u(t)$ only depends on $w([0,t])$ and the solution to (7.2) exists, is unique, and $x(T) = x_{tg}$.

Note that $u(t)$ is allowed to depend on $w(t)$ thanks to real time sensors on all actuators of the system, even on the malfunctioning ones.

Definition 15: System (7.2) is *resiliently stabilizable* (resp. *resilient*) to the loss of the actuators corresponding to C if $0 \in \mathbb{R}^n$ (resp. every $x_{tg} \in \mathbb{R}^n$) is resiliently reachable from any $x_0 \in \mathbb{R}^n$ by system (7.2).

We are now led to our first problem.

Problem 10: Determine whether system (7.2) is resiliently stabilizable and/or resilient.

Even if system (7.2) is not resilient, it might still be able to resiliently reach some targets, just not all of $\mathbb{R}^n$.

Problem 11: Determine the states $x_{tg} \in \mathbb{R}^n$ that are resiliently reachable from a given $x_0 \in \mathbb{R}^n$ by system (7.2).

For time-constrained missions, resilience is not sufficient. We also need to quantify how much slower the malfunctioning system is compared to the nominal one. To do so, we follow Chapter 5 and recall the definitions of the *nominal reach time*

$$T_N^*(x_0, x_{tg}) := \inf_{\bar{u} \,\in\, \mathcal{F}(\bar{\mathcal{U}})} \left\{ T > 0 : x(T) = x_{tg} \text{ in system (7.1)} \right\}, \tag{7.3}$$

the *malfunctioning reach time*

$$T_M^*(x_0, x_{tg}) := \sup_{w \,\in\, \mathcal{F}(\mathcal{W})} \left\{ \inf_{u \,\in\, \mathcal{F}(\mathcal{U})} \left\{ T > 0 : x(T) = x_{tg} \text{ in system (7.2)} \right\} \right\}, \tag{7.4}$$

and the *quantitative resilience*

$$r_q(x_{tg}) := \inf_{x_0 \,\in\, \mathbb{R}^n} \frac{T_N^*(x_0, x_{tg})}{T_M^*(x_0, x_{tg})}. \tag{7.5}$$

If $x_0 = x_{tg}$, then $T_N^* = T_M^* = 0$ and we take the convention that their ratio is 1. If x_{tg} is reachable from x_0 by system (7.1), then Theorem 4.3 of [78] states that the inf in (7.3) becomes min since $\bar{\mathcal{U}}$ is compact and convex. Similarly, T_M^* in (7.4) is achieved by optimal signals $w^* \in \mathcal{F}(\mathcal{W})$ and $u^* \in \mathcal{F}(\mathcal{U})$ when system (7.2) is resilient.

The only way to calculate u^* without any future knowledge of w^* is to solve the intractable Isaac's main equation [86], which is the differential games counterpart of the Hamilton-Jacobi-Bellman (HJB) equation. According to [85], Isaac's main equation is even more difficult to solve than the HJB equation, which usually results in intractable partial differential equations [78]. Hence, [86] produces only suboptimal solutions, itself concluding that its practical contribution is minimal.

Instead of the setting of [86], we choose [30], where u^* and w^* are unique, *bang-bang* [80], and make a time-optimal transfer from x_0 to x_{tg}. The controller knows that w^* will be chosen to make T_M^* the longest. Thus, u^* is chosen to react optimally to this worst undesirable input. Then, w^* is chosen, and to make T_M^* the longest, it is the same as the controller had predicted, this is a Stakelberg optimum [87]. Hence, from an outside perspective it appears as if the controller built u^* knowing w^* in advance, as reflected by (7.4). Then, T_M^* is time-optimal and can be meaningfully compared with T_N^*, leading to the following problem.

Problem 12: Quantify the resilience of system (7.2).

We will now provide the background results upon which we build our theory.

7.3 Background results

We first introduce Hájek's differential games approach [23] which relies on dynamics

$$\dot{x}(t) = Ax(t) + z(t), \quad x(0) = x_0 \in \mathbb{R}^n, \quad z(t) \in \mathcal{Z}, \tag{7.6}$$

where $\mathcal{Z} \subseteq \mathbb{R}^n$ is the Minkowski difference between the set of admissible control inputs $B\mathcal{U} := \big\{Bu : u \in \mathcal{U}\big\}$ and the opposite of the set of undesirable inputs $C\mathcal{W} := \big\{Cw : w \in \mathcal{W}\big\}$, i.e.,

$$\mathcal{Z} := B\mathcal{U} \ominus (-C\mathcal{W}) = \big\{z \in B\mathcal{U} : z - Cw \in B\mathcal{U} \text{ for all } w \in \mathcal{W}\big\}.$$

Theorem 16 (Hájek's duality theorem [23])**:** The state of system (7.2) can be driven to $0 \in \mathbb{R}^n$ at time T for all $w \in \mathcal{F}(\mathcal{W})$ by control signal $u \in \mathcal{F}(\mathcal{U})$ if and only if the state of system (7.6) can be driven to 0 at time T by a control signal $z \in \mathcal{F}(\mathcal{Z})$, and $Bu(\cdot) = z(\cdot) - Cw(\cdot)$.

Informally, $\mathcal{Z}$ represents the control available after counteracting any undesirable input. Since $\bar{\mathcal{U}}$ is symmetric, compact, and convex, sets $B\mathcal{U}$ and $C\mathcal{W}$ also have these properties by linearity. According to [83], $\mathcal{Z}$ is then also symmetric, compact, and convex.

Theorem 16 transforms the resilient stabilizability of system (7.2) into the stabilizability of system (7.6). Because inputs are bounded, Kalman's stabilizability condition [90] do not apply, instead we employ Corollary 3.6 of [91].

Theorem 17 (Stabilizability condition [91])**:** If $\bar{\mathcal{U}} \cap \ker(\bar{B}) \neq \emptyset$ and $\text{int}(\text{co}(\bar{\mathcal{U}})) \neq \emptyset$, then system (7.1) is stabilizable if and only if $\text{rank}\big(\mathcal{C}(A, \bar{B})\big) = n$, $Re\big(\lambda(A)\big) \leq 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top \bar{B}\bar{u} \leq 0$ for all $\bar{u} \in \bar{\mathcal{U}}$.

The first condition of Theorem 17 ensures the existence of a control canceling $\bar{B}\bar{u}$ so that the state can be maintained at an equilibrium. The second condition verifies whether the set of admissible controls is wide enough. The rank condition is Kalman's [91] and the last two conditions guarantee that the drift term Ax does not prevent stabilization. If $\bar{\mathcal{U}} = \mathbb{R}^m$, Theorem 17 reduces to the usual stabilizability condition.

To verify controllability we use Corollary 3.7 of [91], which is very similar to Theorem 17 except that the eigenvalues of A must have a zero real part to avoid creating a drift preventing the reachability of affine targets.

Theorem 18 (Controllability condition [91])**:** If $\bar{\mathcal{U}} \cap \ker(\bar{B}) \neq \emptyset$ and $\text{int}(\text{co}(\bar{U})) \neq \emptyset$, then system (7.1) is controllable if and only if $\text{rank}\big(\mathcal{C}(A, \bar{B})\big) = n$, $Re\big(\lambda(A)\big) = 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top \bar{B}\bar{u} \leq 0$ for all $\bar{u} \in \bar{\mathcal{U}}$.

We now have all the background results to start solving Problem 10 by investigating resilient stabilizability.

7.4 Resilient stabilizability

In this section, we first establish a simple resilient stabilizability condition before deriving a more complex condition with a wider range of application.

Proposition 19: If $\text{int}(\mathcal{Z}) \neq \emptyset$, then system (7.2) is resiliently stabilizable if and only if $Re\big(\lambda(A)\big) \leq 0$.

Proof. According to Theorem 16, the resilient stabilizability of system (7.2) is equivalent to the stabilizability of system (7.6). We apply Theorem 17 and obtain that if $\mathcal{Z} \cap \ker(I) \neq \emptyset$ and $\text{int}(\text{co}(\mathcal{Z})) \neq \emptyset$ in $\mathbb{R}^n$, then system (7.6) is stabilizable if and only if $\text{rank}\big(\mathcal{C}(A, I)\big) = n$, $Re\big(\lambda(A)\big) \leq 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top I z \leq 0$ for all $z \in \mathcal{Z}$.

Because $\ker(I) = \{0\}$, the first condition becomes $0 \in \mathcal{Z}$. Since $\mathcal{Z}$ is convex, the second condition becomes $\text{int}(\mathcal{Z}) \neq \emptyset$, which is equivalent to $0 \in \text{int}(\mathcal{Z})$ according to Lemma 17 of Section 7.8. This second condition implies the first one, so we only keep $\text{int}(\mathcal{Z}) \neq \emptyset$.

We now assume that $\text{int}(\mathcal{Z}) \neq \emptyset$ and we simplify the last three conditions. Since $\text{rank}(I) = n$, the third condition is always true. Lemma 17 yields $0 \in \text{int}(\mathcal{Z})$. Thus, there exists $\varepsilon > 0$ such that $\mathbb{B}^n(0, \varepsilon) \subseteq \mathcal{Z}$. If $A^\top$ has no real eigenvector, the last condition is trivially true. Otherwise, for v be a real eigenvector of $A^\top$. Let $z = \varepsilon \frac{v}{\|v\|}$, then $z \in \mathbb{B}^n(0, \varepsilon)$, so $z \in \mathcal{Z}$ and $v^\top I z = \varepsilon \|v\| > 0$. □

Proposition 19 has a limited range of application because of its requirement $\text{int}(\mathcal{Z}) \neq \emptyset$ in $\mathbb{R}^n$, i.e., $\mathcal{Z}$ must be of dimension n. However, stabilizability does not require $B\mathcal{U}$ to be dimension n, so resilient stabilizability should not require that from $\mathcal{Z}$ either. We then want our condition to rely on the relative interior of $\mathcal{Z}$ instead of its interior.

Definition 16: The *relative interior* $\text{relint}(\mathcal{S})$ of a set $\mathcal{S}$ is the interior of $\mathcal{S}$ considered as a subset of its affine hull.

Definition 17: The *affine hull* of a set $\mathcal{S}$ is the largest subspace included in $\mathcal{S}$ with respect to inclusion.

If we apply Theorem 17 to system (7.6) as in Proposition 19, then $\text{int}(\mathcal{Z}) \neq \emptyset$ will appear. Instead, we first need to transport system (7.6) into a basis adapted to $\mathcal{Z}$. Let $r := \dim(\mathcal{Z}) \leq n$. If $\mathcal{Z} = \emptyset$, we take the convention that $r = -\infty$ and $Z := [\,] \in \mathbb{R}^{n \times 0}$, the empty matrix with $\text{Im}([\,]) = \emptyset$. Otherwise, according to Lemma 18 of Section 7.8, we have $0 \in \mathcal{Z}$. Then, $\text{span}(\mathcal{Z})$ is a vector space from which we take a basis $\{z_1, \ldots, z_r\}$ in $\mathbb{R}^n$. We define the matrix $Z := \big(z_1, \ldots, z_r\big) \in \mathbb{R}^{n \times r}$ with the convention that $Z = 0 \in \mathbb{R}^{n \times 1}$ if $r = 0$. Then, $\text{Im}(Z) = \text{span}(\mathcal{Z})$ and we can formulate a resilient stabilizability condition less restrictive than Proposition 19.

Proposition 20: If $\mathrm{relint}(\mathcal{Z}) \neq \emptyset$, then system (7.2) is resiliently stabilizable if and only if $\mathrm{rank}\big(\mathcal{C}(A, Z)\big) = n$, $Re\big(\lambda(A)\big) \leq 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$.

Proof. We apply Theorem 16 and work on system (7.6). Since $z_1, \dots, z_r$ are linearly independent, we complete this sequence into a basis of $\mathbb{R}^n$ with $V := (v_{r+1}, \dots, v_n)$ and obtain a transition matrix $T_z = (Z, V)$. We change basis in system (7.6) with $x = T_z^{-1} y$ so that

$$\dot{x}(t) = T_z^{-1}\dot{y}(t) = T_z^{-1}Ay(t) + T_z^{-1}z(t) = \hat{A}x(t) + s(t), \qquad s(t) \in \mathcal{S} := T_z^{-1}\mathcal{Z} = \big\{T_z^{-1}z : z \in \mathcal{Z}\big\},$$

with $\hat{A} = T_z^{-1}AT_z$. By definition, $z_i = T_z e_i$ and thus $\mathcal{S} \subseteq \mathrm{span}(\{e_1, \dots, e_r\})$ in $\mathbb{R}^n$. Let $s \in \mathcal{S}$. Then,

$$s = \begin{pmatrix} s_1 \\ \vdots \\ s_r \\ 0_{n-r,1} \end{pmatrix} = \begin{pmatrix} I_r \\ 0_{n-r,r} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_r \end{pmatrix} := \hat{B}\hat{s},$$

with $\hat{B} = T_z^{-1}Z \in \mathbb{R}^{n\times r}$ and $\hat{s} \in \mathbb{R}^r$, $\hat{s} \in \hat{\mathcal{S}} := \mathrm{proj}_r(\mathcal{S})$, the projection of $\mathcal{S}$ onto $\mathbb{R}^r$. Hence, the stabilizability of system (7.6) is equivalent to that of system

$$\dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + \hat{B}\hat{s}(t), \quad \hat{x}(0) = T_z^{-1}x_0, \quad \hat{s}(t) \in \hat{\mathcal{S}}. \tag{7.7}$$

Applying Theorem 17 to system (7.7) leads to the following stabilizability conditions: $\hat{\mathcal{S}} \cap \ker(\hat{B}) \neq \emptyset$, $\mathrm{int}(\mathrm{co}(\hat{\mathcal{S}})) \neq \emptyset$, $Re(\lambda(\hat{A})) \leq 0$, $\mathrm{rank}\big(\mathcal{C}(\hat{A}, \hat{B})\big) = n$, and there is no real eigenvector $\hat{v}$ of $\hat{A}^\top$ satisfying $\hat{v}^\top \hat{B}\hat{s} \leq 0$ for all $\hat{s} \in \hat{\mathcal{S}}$. We now simplify these five conditions.

1. Since $\hat{B} = \left(\begin{smallmatrix} I_r \\ 0 \end{smallmatrix}\right)$, $\mathrm{rank}(\hat{B}) = r$, and hence $\ker(\hat{B}) = \{0\}$ in $\mathbb{R}^r$. Then, $\hat{\mathcal{S}} \cap \ker(\hat{B}) \neq \emptyset$ is equivalent to $0 \in \hat{\mathcal{S}} = \mathrm{proj}_r(T_z^{-1}\mathcal{Z})$. In turn, this is equivalent to the existence of $v \in \mathbb{R}^{n-r}$ such that $T_z \left(\begin{smallmatrix} 0 \\ v \end{smallmatrix}\right) \in \mathcal{Z}$, i.e., $Vv \in \mathcal{Z}$. By definition of V, $\mathrm{Im}(V) \cap \mathrm{span}(\mathcal{Z}) = \{0\}$. Thus, $\hat{\mathcal{S}} \cap \ker(\hat{B}) \neq \emptyset$ is equivalent to $0 \in \mathcal{Z}$, i.e., $\mathrm{relint}(\mathcal{Z}) \neq \emptyset$ according to Lemma 18 of Section 7.8.

2. By definition of $\mathcal{S}$, $\mathrm{int}(\hat{\mathcal{S}}) \neq \emptyset$ in $\mathbb{R}^r$ is equivalent to $\mathrm{relint}(\mathcal{Z}) \neq \emptyset$ since T_z is invertible.

3. Because $\hat{A} = T_z^{-1}AT_z$, $\lambda(A) = \lambda(\hat{A})$, and thus the third condition becomes $Re(\lambda(A)) \leq 0$.

4. For $i \in [\![0, n{-}1]\!]$, $T_z\hat{A}^i\hat{B} = T_z\big(T_z^{-1}AT_z\big)^i\hat{B} = A^iT_z\hat{B} = A^iZ$ because $T_z\hat{B} = Z$. Hence, $\mathrm{Im}\,\big(T_z\mathcal{C}(\hat{A}, \hat{B})\big) = \mathrm{Im}\,\big(\mathcal{C}(A, Z)\big)$. The invertibility of T_z leads to $\mathrm{rank}\,\big(\mathcal{C}(\hat{A}, \hat{B})\big) = \mathrm{rank}\,\big(\mathcal{C}(A, Z)\big)$ [100].

5. Assume that $\hat{v}$ is a real eigenvector of $\hat{A}^\top$ associated to the eigenvalue $\hat{\lambda}$. Then, $v := T_z^{-\top}\hat{v}$ is an eigenvector of $A^\top$ associated to the same eigenvalue $\hat{\lambda}$ [100]. For $\hat{s} \in \hat{\mathcal{S}}$, we have $\hat{B}\hat{s} \in \mathcal{S}$ by definition. Hence, if we define $z := T_z\hat{B}\hat{s}$, we have $z \in \mathcal{Z}$. Then, $\hat{v}^\top\hat{B}\hat{s} = v^\top T_z\hat{B}\hat{s} = v^\top z$.

□

To further expand the applicability of our resilient stabilizability condition, we now remove the requirement $\mathrm{relint}(\mathcal{Z}) \neq \emptyset$ from Proposition 20 and obtain a necessary and sufficient condition.

Theorem 19 (Resilient stabilizability condition)**:** System (7.2) is resiliently stabilizable if and only if $\mathrm{rank}\,\big(\mathcal{C}(A, Z)\big) = n$, $Re\big(\lambda(A)\big) \leq 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$.

Proof. Let us define the three properties stated in Proposition 20 as $\mathcal{P}_1 :=$"relint$(\mathcal{Z}) \neq \emptyset$", $\mathcal{P}_2 :=$"System (7.2) is resiliently stabilizable", and $\mathcal{P}_3 :=$"rank$\big(\mathcal{C}(A, Z)\big) = n$, $Re\big(\lambda(A)\big) \leq 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$". Proposition 20 states that if $\mathcal{P}_1$ holds, then $\mathcal{P}_2$ is equivalent to $\mathcal{P}_3$. We will now show that when $\mathcal{P}_1$ is false, so are $\mathcal{P}_2$ and $\mathcal{P}_3$, which leads to $\mathcal{P}_2$ equivalent to $\mathcal{P}_3$ no matter the status of $\mathcal{P}_1$, which is exactly the statement of this theorem.

Assume that $\mathcal{P}_1$ is false. Then, according to Lemmas 18, 21, and 22 of Section 7.8, system (7.2) is not resiliently stabilizable, i.e., $\mathcal{P}_2$ is false. We took the convention that $Z = [\,]$ with rank$([\,]) = -\infty$, so $\mathcal{P}_3$ is false too. □

Note that the rank condition in Theorem 19 concerns the pair (A, Z) and not (A, B) as one might have wanted. For the stabilizability of these pairs to be equivalent, we need $\mathcal{Z}$ and $B\mathcal{U}$ to have the same dimension.

Corollary 4: If $\dim(\mathcal{Z}) = \text{rank}(B)$, then system (7.2) is resiliently stabilizable if and only if rank $\big(\mathcal{C}(A, B)\big) = n$, $Re\big(\lambda(A)\big) \leq 0$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$.

Proof. If $\mathcal{Z} = \emptyset$, then rank$(B) = -\infty$, i.e., $B = []$. Thus, system (7.2) is not resiliently stabilizable and rank $\big(\mathcal{C}(A, B)\big) \neq n$.

Now assume that $\mathcal{Z} \neq \emptyset$. From Lemma 20 of Section 7.8 we get Im$(B) =$ Im(Z). Then, Im $\big(\mathcal{C}(A, B)\big) =$ Im $\big(\mathcal{C}(A, Z)\big)$. In the proof of Proposition 20 we had Im $\big(\mathcal{C}(A, Z)\big) =$ Im $\big(T\mathcal{C}(\hat{A}, \hat{B})\big)$. Since T is invertible, we obtain rank $\big(\mathcal{C}(A, B)\big) =$ rank $\big(\mathcal{C}(\hat{A}, \hat{B})\big)$, and we conclude with the rest of the proof of Proposition 20. □

Notice how the three conditions listed in Corollary 4 are similar to the stabilizability conditions from Theorem 17. We are then led to the following result.

Corollary 5: If $\dim(\mathcal{Z}) = \text{rank}(B)$, then system (7.2) is resiliently stabilizable if and only if system (7.1) is stabilizable.

Proof. Let v be a real eigenvector of $A^\top$. Assume first that there exists $z \in \mathcal{Z}$ such that $v^\top z > 0$. By construction of B, $\mathcal{U}$, and $\mathcal{Z}$, we have $\mathcal{Z} \subseteq B\mathcal{U} \subseteq \bar{B}\bar{\mathcal{U}}$. Hence, there exists $\bar{u} \in \bar{\mathcal{U}}$ such that $z = \bar{B}\bar{u}$ and $v^\top \bar{B}\bar{u} > 0$.

On the other hand, assume that there exists $\bar{u} \in \bar{\mathcal{U}}$ such that $v^\top \bar{B}\bar{u} > 0$. According to Lemma 20, span$(\mathcal{Z}) =$ Im$(\bar{B})$. Then, the convexity of $\mathcal{Z}$ yields the existence of $\alpha \in \mathbb{R}$ and $z \in \mathcal{Z}$ such that $\bar{B}\bar{u} = \alpha z$. Note that $\alpha \neq 0$ by definition of $\bar{u}$. If $\alpha > 0$, we have $v^\top z > 0$. Otherwise, $\alpha < 0$ but we use the symmetry of $\mathcal{Z}$ to obtain $-z \in \mathcal{Z}$ and $v^\top(-z) > 0$.

Thus, the condition "there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$" is equivalent to "there is no real eigenvector v of $A^\top$ satisfying $v^\top \bar{B}\bar{u} \leq 0$ for all $\bar{u} \in \bar{\mathcal{U}}$" when $\dim(\mathcal{Z}) = \text{rank}(B)$. According to Lemma 20 of Section 7.8, Im$(B) =$ Im$(\bar{B})$. Hence, rank $\big(\mathcal{C}(A, B)\big) =$ rank $\big(\mathcal{C}(A, \bar{B})\big)$. Then, applying Corollary 4 to system (7.2) and Theorem 17 to system (7.1) concludes the proof. □

We have established several resilient stabilizability conditions, hence solving the first half of Problem 10. We will now tackle its second part concerning affine targets.

7.5 Resilient reachability

In this section we extend Hájek's duality theorem [23] to affine targets and study the resilience of linear systems.

Theorem 20 (Extended duality theorem)**:** The state of system (7.2) can be driven to $x_{tg} \in \mathbb{R}^n$ at time T for all $w \in \mathcal{F}(\mathcal{W})$ by control signal $u \in \mathcal{F}(\mathcal{U})$ if and only if the state of system (7.6) can be driven to x_{tg} at time T by a control signal $z \in \mathcal{F}(\mathcal{Z})$, and $Bu(\cdot) = z(\cdot) - Cw(\cdot)$.

Proof. Consider system (7.2) with a target state $x_{tg} \in \mathbb{R}^n$, $x_{tg} \neq 0$. Let $X(t) := \left(\begin{smallmatrix} x(t) - x_{tg} \\ Ax_{tg} \end{smallmatrix} \right) \in \mathbb{R}^{2n}$. Then,

$$\dot{X}(t) = A_2 X(t) + B_2 u(t) + C_2 w(t), \quad X(0) = X_0 \in \mathbb{R}^{2n}, \quad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W}, \tag{7.8}$$

$$\text{with} \quad A_2 = \begin{pmatrix} A & I_n \\ 0_{n,n} & 0_{n,n} \end{pmatrix}, \quad B_2 = \begin{pmatrix} B \\ 0_{n,m} \end{pmatrix}, \quad C_2 = \begin{pmatrix} C \\ 0_{n,p} \end{pmatrix} \quad \text{and} \quad X_0 = \begin{pmatrix} x_0 - x_{tg} \\ Ax_{tg} \end{pmatrix}.$$

Let the target set be $\mathcal{G} = \left\{ \left(\begin{smallmatrix} 0 \\ a \end{smallmatrix} \right) \in \mathbb{R}^{2n} \right\} = \{0\}^n \times \mathbb{R}^n$. Since $0 \in C_2\mathcal{W}$, we can apply Hájek's second duality theorem of [23] stating that $\mathcal{G}$ is resiliently reachable in time T from X_0 by system (7.8) if and only if $\mathcal{G}$ is reachable in time T from X_0 by the following system

$$\dot{X}(t) = A_2 X(t) + v_2(t), \quad X(0) = X_0, \quad v_2(t) \in \mathcal{V}_2 := B_2\mathcal{U} \cap \big[(B_2\mathcal{U} \oplus \mathcal{G}_{A_2}) \ominus (-C_2\mathcal{W}) \big] \subseteq \mathbb{R}^{2n}, \tag{7.9}$$

where $\mathcal{G}_{A_2}$ is the largest subspace of $\mathcal{G}$ invariant by A_2. Take $g = \left(\begin{smallmatrix} 0 \\ a \end{smallmatrix} \right) \in \mathcal{G}$, then

$$A_2 g = \begin{pmatrix} A & I_n \\ 0_{n,n} & 0_{n,n} \end{pmatrix} \begin{pmatrix} 0 \\ a \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix}.$$

Hence, $A_2 g \in \mathcal{G} \iff a = 0$, i.e., $\mathcal{G}_{A_2} = \{0\}^{2n}$. Thus,

$$\mathcal{V}_2 = \big\{ v \in B_2\mathcal{U} : v - C_2 w \in B_2\mathcal{U}, \text{for all } w \in \mathcal{W} \big\} = \mathcal{Z} \times \{0\}^n,$$

because of the architecture of B_2 and C_2. Then, system (7.9) is related to system (7.6) the same way that system (7.8) is related to system (7.2). Therefore, the following statements are equivalent:

- x_{tg} is resiliently reachable by system (7.2),
- $\mathcal{G}$ is resiliently reachable by system (7.8),
- $\mathcal{G}$ is reachable by system (7.9),
- x_{tg} is reachable by system (7.6).

□

Theorem 20 transforms resilience of system (7.2) into bounded controllability of system (7.6), which we verify with Theorem 18. We can easily adapt the results of Section 7.4 to the resilience case by reusing the same proofs, except that we use Theorems 20 and 18 instead of Theorems 16 and 17.

Proposition 21: If $\text{int}(\mathcal{Z}) \neq \emptyset$, then system (7.2) is resilient if and only if $Re(\lambda(A)) = 0$.

Corollary 6: If $\dim(\mathcal{Z}) = \text{rank}(B)$, then system (7.2) is resilient if and only if system (7.1) is controllable.

Theorem 21 (Resilience condition)**:** System (7.2) is resilient if and only if $Re\big(\lambda(A)\big) = 0$, rank $\big(\mathcal{C}(A, Z)\big) = n$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$.

We now have all the results necessary to solve Problem 10. However, the condition $Re\big(\lambda(A)\big) = 0$ in Theorem 21 is not satisfied by most systems, that are hence not resilient. This reasoning led us to Problem 11, i.e., the determination of the resiliently reachable set of system (7.2). Following Theorem 20, we will now study the reachable set of system (7.6) given by

$$R(T, x_0) := \left\{ e^{AT}\left(x_0 + \int_0^T e^{-At} z(t)\, dt \right) \quad \text{with } z(t) \in \mathcal{Z} \text{ for all } t \in [0, T] \right\}.$$

Because analytical study of $R(T, x_0)$ is difficult, most of the research tries to approximate it (see [61] and references therein). We want inner approximations of $R(T, x_0)$ in order to determine the states that are guaranteed to be resiliently reachable. We will then present a method of *zonotopic* underapproximation of $R(T, x_0)$ combining the approaches of [61] and [82].

Definition 18: A *zonotope* $\mathcal{S} \subseteq \mathbb{R}^n$ is a set parametrized by a center $c \in \mathbb{R}^n$ and generators $g_1, \ldots, g_q \in \mathbb{R}^n$ expressed as $\mathcal{S} := \{c + \sum_{i=1}^q \alpha_i g_i : \alpha_i \in [-1, 1]\}$ and is denoted $\mathcal{S} = (c, g_1, \ldots g_q)$.

Note that $B\mathcal{U}$ is a zonotope of center 0 and generators B_i, the columns of B. Similarly, $C\mathcal{W} = (0, C_1, \ldots, C_p)$. However, $\mathcal{Z}$ is not a zonotope in general since these sets are not closed under Minkowski difference except for some specific scenarios, as detailed in [82].

Following [82], we build an underapproximation of $\mathcal{Z}$ with a symmetric zonotope $\big(0, g_1, \ldots, g_r\big) \subseteq \mathcal{Z}$ by removing or contracting the generators of $B\mathcal{U}$. We apply the method described in [61] to compute efficiently an inner approximation of $R(T, x_0)$. For $N \in \mathbb{N}$, $N \geq 1$, we define

$$\delta t := \frac{T}{N}, \quad \Omega_0 := \{x_0\}, \quad V := \left\{ \int_0^{\delta t} e^{A(\delta t - t)} z(t)\, dt : z(t) \in \mathcal{Z} \text{ for } t \in [0, \delta t] \right\},$$

and the recursion $\Omega_{i+1} := e^{A\delta t}\Omega_i \oplus V$. Note that Ω_i is the exact reachable set $R(i\,\delta t, x_0)$.

However, V is not a zonotope and cannot be computed exactly. Thus, we define the zonotope

$$\tilde{V} := \left(0, \int_0^{\delta t} e^{A(\delta t - t)} g_1\, dt, \ldots, \int_0^{\delta t} e^{A(\delta t - t)} g_r\, dt \right),$$

and $\tilde{V} \subseteq V$ since $\tilde{V}$ corresponds to piecewise constant components of $z(t)$ in $\big(0, g_1, \ldots, g_r\big)$. Indeed, for $\tilde{v} \in \tilde{V}$ there exists $\alpha_1, \ldots, \alpha_r \in [-1, 1]$ such that

$$\tilde{v} = \sum_{i=1}^r \alpha_i \int_0^{\delta t} e^{A(\delta t - t)} g_i\, dt = \int_0^{\delta t} e^{A(\delta t - t)} \sum_{i=1}^r g_i \alpha_i\, dt.$$

Note that $\sum_{i=1}^r g_i\alpha_i \in \big(0, g_1, \ldots, g_r\big) \subseteq \mathcal{Z}$, so $\tilde{v} \in V$, i.e., $\tilde{V} \subseteq V$. Then, we build $\tilde{\Omega}_0 = \Omega_0 = \{x_0\}$ and $\tilde{\Omega}_{i+1} := e^{A\delta t}\tilde{\Omega}_i \oplus \tilde{V}$, which yields $\tilde{\Omega}_i \subseteq \Omega_i$ for all $i \geq 0$. Since linear maps and Minkowski sums are straightforward on zonotopes [61], [82], $\tilde{\Omega}_i$ is an easily computable inner approximation of the reachable set $R(i\,\delta t, x_0)$. Note that the precision of the approximation increases with N.

Before implementing this solution to Problem 11 in Section 7.7.1, we need to answer Problem 12 by quantifying the resilience of linear systems.

7.6 Quantitative resilience

Let us now investigate more complex missions where the target needs to be reached by a certain time. In such scenarios it is crucial to evaluate the maximal time penalty incurred by the malfunctioning system.

Unlike in the driftless case [32], the optimal reach times T_N^* (7.3) and T_M^* (7.4) cannot be reduced to a linear optimization and elude analytical expressions [29]. Following [71] and [30] we could numerically compute these reach times, but not the quantitative resilience r_q (7.5) since it would require computing $T_N^*(x_0)$ and $T_M^*(x_0)$ for all $x_0 \in \mathbb{R}^n$. Instead, using Lyapunov theory [121], we establish analytical bound on these two reach times for the target $x_{tg} = 0$ and analytically approximate r_q.

7.6.1 Nominal reach time

Assume that A is Hurwitz. Then, for any $Q \succ 0$ there exists $P \succ 0$ such that $PA + A^\top P = -Q$ [121]. Let us consider any such pair (P, Q). We define the Lyapunov function $V(x) := x^\top P x = \|x\|_P^2$ [122]. Then, for x following (7.1) we have

$$\dot{V}(x) = \dot{x}^\top P x + x^\top P \dot{x} = x^\top (A^\top P + PA)x + 2x^\top P\bar{B}\bar{u} = -x^\top Q x + 2x^\top P \bar{B}\bar{u}.$$

We will now bound $T_N^*(x_0)$.

Proposition 22: If system (7.1) is stabilizable and A is Hurwitz, then

$$T_N^*(x_0) \;\geq\; 2\frac{\lambda_{min}^P}{\lambda_{max}^Q} \ln\left(1 + \frac{\lambda_{max}^Q \|x_0\|_P}{2\lambda_{min}^P b_{max}^P}\right), \tag{7.10}$$

with $b_{max}^P := \max\big\{\|\bar{B}\bar{u}\|_P : \bar{u} \in \bar{\mathcal{U}}\big\}$.

Proof. Because $\bar{\mathcal{U}}$ is compact and convex, and system (7.1) is stabilizable, there exists a time-optimal control signal $\bar{u}^* \in \mathcal{F}(\bar{\mathcal{U}})$ driving the state from x_0 to the origin in a finite time $T_N^*(x_0)$ [78].

We now bound $\dot{V}$ using (7.10). Since $P \succ 0$, there exists $M \in \mathbb{R}^{n\times n}$ such that $P = M^\top M$ [100]. Then, $x^\top P\bar{B}\bar{u} = (Mx)^\top M\bar{B}\bar{u} \geq -\|Mx\|\|M\bar{B}\bar{u}\|$, by the Cauchy-Schwarz inequality [100]. Notice $\|Mx\|^2 = x^\top M^\top M x = x^\top P x = \|x\|_P^2$. Similarly, $\|M\bar{B}\bar{u}\| = \|\bar{B}\bar{u}\|_P$.

The maximum b_{max}^P exists since $\bar{\mathcal{U}}$ is compact and the map $\bar{u} \mapsto \|\bar{B}\bar{u}\|_P$ is continuous. Since $Q \succ 0$, we have $x^\top Q x \leq \lambda_{max}^Q \|x\|^2$ and $\|x\|^2 \leq \|x\|_P^2/\lambda_{min}^P$ because $P \succ 0$. For $x \neq 0$, we have now lower bounded (7.10)

$$\dot{V}(x) = \frac{d}{dt}\|x\|_P^2 \geq -\frac{\lambda_{max}^Q}{\lambda_{min}^P}\|x\|_P^2 - 2b_{max}^P\|x\|_P. \tag{7.11}$$

Let $y(t) := \|x(t)\|_P$, $\alpha := \frac{\lambda_{max}^Q}{2\lambda_{min}^P} > 0$, and $\beta := b_{max}^P > 0$. For $x \neq 0$ we divide (7.11) by $2y > 0$ so that $\dot{y} \geq f(y) := -\alpha y - \beta$. The solution of the differential equation $\dot{s}(t) = f\big(s(t)\big)$ with $s(0) = y(0)$ is given by $s(t) = e^{-\alpha t}\left(y(0) + \frac{\beta}{\alpha}\right) - \frac{\beta}{\alpha}$.

Since f is Lipschitz, we can apply the comparison lemma of [122] and we obtain $y(t) \geq s(t)$ for all $t \geq 0$. At time $T = \frac{1}{\alpha}\ln\left(1 + \frac{\alpha}{\beta}y(0)\right)$, we have $s(T) = 0$. Because $\|x(t)\|_P \geq s(t) > 0$ for all $t \in [0, T]$, we have $T_N^*(x_0) \geq T$. Substituting α and β yields (7.12). □

The proof of Propositions 22, as well as subsequent Propositions 23, 24, and 25, is shorter than first presented in [33] due to our use of the comparison lemma [122]. We now upper bound $T_N^*(x_0)$.

Proposition 23: If $\mathrm{rank}(\bar{B}) = n$ and A is Hurwitz, then

$$T_N^*(x_0) \;\leq\; 2\frac{\lambda_{max}^P}{\lambda_{min}^Q} \ln\left(1 + \frac{\lambda_{min}^Q \|x_0\|_P}{2\lambda_{max}^P b_{min}^P}\right), \tag{7.12}$$

with $b_{min}^P := \min\left\{\|\bar{B}\bar{u}\|_P : \bar{u} \in \partial\bar{\mathcal{U}}\right\}$.

Proof. The minimum b_{min}^P exists since map $\bar{u} \mapsto \|\bar{B}\bar{u}\|_P$ is continuous and $\partial\bar{\mathcal{U}}$ is compact. Because $\mathrm{rank}(\bar{B}) = n$, we can choose $\bar{u} \in \mathcal{F}(\bar{\mathcal{U}})$ such that $\bar{B}\bar{u}(t) = -\frac{x(t)}{\|x(t)\|_P} b_{min}^P$ for $x(t) \neq 0$. Indeed, assume for contradiction purposes that for some $\tau \geq 0$, $\bar{u}(\tau) \notin \bar{\mathcal{U}}$, i.e., $\|\bar{u}(\tau)\|_\infty > 1$. Let $\hat{u} := \frac{\bar{u}(\tau)}{\|\bar{u}(\tau)\|_\infty}$. Then, $\|\hat{u}\|_\infty = 1$, so $\hat{u} \in \partial\bar{\mathcal{U}}$, but $\|\bar{B}\hat{u}\|_P = \frac{\|\bar{B}\bar{u}(\tau)\|_P}{\|\bar{u}(\tau)\|_\infty} = \frac{b_{min}^P}{\|\bar{u}\|_\infty} < b_{min}^P$, which is a contradiction. Hence, the proposed control signal is admissible and we implement it in (7.10).

We obtain $2x^\top P\bar{B}\bar{u} = -2b_{min}^P\|x\|_P$, so that

$$\frac{d}{dt}\|x\|_P^2 = \dot{V}(x) \leq -\frac{\lambda_{min}^Q}{\lambda_{max}^P}\|x\|_P^2 - 2b_{min}^P\|x\|_P. \tag{7.13}$$

Let $y(t) := \|x(t)\|_P$, $\gamma := \frac{\lambda_{min}^Q}{2\lambda_{max}^P} > 0$, and $\kappa := b_{min}^P > 0$. For $x \neq 0$, dividing (7.13) by $2y > 0$, yields $\dot{y} \leq f(y) := -\gamma y - \kappa$. As in Proposition 22, the comparison lemma of [122] yields

$$y(t) \leq s(t) = e^{-\gamma t}\left(y(0) + \frac{\kappa}{\gamma}\right) - \frac{\kappa}{\gamma} \qquad \text{for all } t \geq 0 \qquad \text{as long as } y(t) > 0.$$

At time $T = \frac{1}{\gamma}\ln\left(1 + \frac{\gamma}{\kappa}y(0)\right)$, $s(T) = 0$. Since $y\big(T_N^*(x_0)\big) = 0$, $T_N^*(x_0) \leq T$. □

We now bound the malfunctioning reach time T_M^* following the same method applied to T_N^*.

7.6.2 Malfunctioning reach time

We use the same Lyapunov function as above, but with x following (7.2), so $\dot{V}(x) = -x^\top Qx + 2x^\top P(Bu + Cw)$. We can now lower bound T_M^* as we have done for T_N^*.

Proposition 24: If system (7.2) is resiliently stabilizable and A is Hurwitz, then

$$T_M^*(x_0) \;\geq\; 2\frac{\lambda_{min}^P}{\lambda_{max}^Q} \ln\left(1 + \frac{\lambda_{max}^Q \|x_0\|_P}{2\lambda_{min}^P z_{max}^P}\right), \tag{7.14}$$

with $z_{max}^P := \max\left\{\|z\|_P : z \in \mathcal{Z}\right\}$.

Proof. Since $B\mathcal{U}$ and $C\mathcal{W}$ are compact, $\mathcal{Z}$ is compact [83], so z_{max}^P exists. Since system (7.2) is resiliently stabilizable, $T_M^*(x_0)$ exists. Let $w^* \in \mathcal{F}(\mathcal{W})$ and $u^* \in \mathcal{F}(\mathcal{W})$ be the arguments of the optimizations in (7.4). By definition of $\mathcal{Z}$, $z = Cw^* + Bu^* \in \mathcal{F}(\mathcal{Z})$. Then, $\|Cw^*(t) + Bu^*(t)\|_P \leq z_{max}^P$, which yields

$$\dot{V}(x) \geq -\frac{\lambda_{max}^Q}{\lambda_{min}^P}\|x\|_P^2 - 2z_{max}^P\|x\|_P.$$

We now proceed as in the second half of the proof of Proposition 22 to obtain (7.14). □

Similarly, we upper bound the malfunctioning reach time.

Proposition 25: If $\mathrm{int}(\mathcal{Z}) \neq \emptyset$ and A is Hurwitz, then

$$T_M^*(x_0) \;\leq\; 2\frac{\lambda_{max}^P}{\lambda_{min}^Q} \ln\left(1 + \frac{\lambda_{min}^Q \|x_0\|_P}{2\lambda_{max}^P z_{min}^P}\right), \tag{7.15}$$

with $z_{min}^P := \min\big\{\|z\|_P : z \in \partial\mathcal{Z}\big\}$.

Proof. According to Proposition 19, system (7.2) is resiliently stabilizable, hence a finite T_M^* exists.

Since $\mathcal{Z}$ is compact, so is $\partial\mathcal{Z}$, and thus z_{min}^P exists. Because $\mathrm{int}(\mathcal{Z}) \neq \emptyset$, according to Lemma 17, $0 \in \mathrm{int}(\mathcal{Z})$. Then, the convexity of $\|\cdot\|_P$ yields $\big\{z \in \mathbb{R}^n : \|x\|_P \leq z_{min}^P\big\} \subseteq \mathcal{Z}$, so $z(t) := \frac{-x(t)}{\|x(t)\|_P} z_{min}^P \in \mathcal{Z}$.

Let $w^* \in \mathcal{F}(\mathcal{W})$ be the argument of the maximum in (7.4). Since $z(t) \in \mathcal{Z}$, there exists $u \in \mathcal{F}(\mathcal{U})$ such that $z(t) = Cw^*(t) + Bu(t)$. Then, applying w^* and u leads to an upper bound of T_M^* since u is not necessarily optimal, while w^* is optimal. Hence

$$\dot{V}(x) \leq -\frac{\lambda_{min}^Q}{\lambda_{max}^P}\|x\|_P^2 - 2z_{min}^P\|x\|_P.$$

We now proceed as in the second half of the proof of Proposition 23 to obtain (7.15). □

We can now bound $T_N^*(x_0)/T_M^*(x_0)$ for all $x_0 \in \mathbb{R}^n$ and hence obtain an approximate of quantitative resilience r_q which cannot be done with prior algorithms [30], [71] that only compute a single instance of $T_N^*(x_0)$ or $T_M^*(x_0)$.

7.6.3 Bounding quantitative resilience

If the system's quantitative resilience r_q is bounded by $\gamma \leq r_q$, then in the worst case, the malfunctioning system will take less than $1/\gamma$ times longer than the nominal system to reach the origin from the same initial state.

Theorem 22: If $\mathrm{int}(\mathcal{Z}) \neq \emptyset$ and A is Hurwitz, then

$$r_q \geq \max\left(\frac{\lambda_{min}^P \lambda_{min}^Q}{\lambda_{max}^P \lambda_{max}^Q}, \frac{z_{min}^P}{b_{max}^P}\right), \tag{7.16}$$

for any $P \succ 0$ and $Q \succ 0$ such that $A^\top P + PA = -Q$.

Proof. According to Proposition 19, system (7.2) is resiliently stabilizable. Since $\mathrm{int}(\mathcal{Z}) \neq \emptyset$, we have $\dim(\mathcal{Z}) = n$, and $\mathcal{Z} \subseteq B\mathcal{U} \subseteq \mathbb{R}^n$ yields $\mathrm{rank}(B) = n$. According to Corollary 5, system (7.1) is stabilizable, so we can use (7.10) and (7.15). We define the positive constants

$$a := \frac{\lambda_{min}^P \lambda_{min}^Q}{\lambda_{max}^P \lambda_{max}^Q}, \qquad b := \frac{\lambda_{max}^Q}{2\lambda_{min}^P b_{max}^P}, \qquad \text{and} \qquad c := \frac{\lambda_{min}^Q}{2\lambda_{max}^P z_{min}^P},$$

so that for $x_0 \in \mathbb{R}^n$, $x_0 \neq 0$, (7.10) and (7.15) yield

$$\frac{T_N^*(x_0)}{T_M^*(x_0)} \geq a\frac{\ln(1 + b\|x_0\|_P)}{\ln(1 + c\|x_0\|_P)} := f(\|x_0\|_P).$$

Then, according to (7.5), $r_q \geq \inf\limits_{x_0 \,\in\, \mathbb{R}^n} f(\|x_0\|_P)$.

If $b = c$, then $f(s) = a$ for all $s \geq 0$, so $r_q \geq a$. If $b > c$, then f is increasing, so $\inf \big\{ f(s) : s > 0 \big\} = \lim\limits_{s \to 0} f(s)$. L'Hôpital's Rule [123] yields

$$\lim_{s \to 0} f(s) = \lim_{s \to 0} a \frac{\ln(1 + bs)}{\ln(1 + cs)} = \lim_{s \to 0} a \frac{\frac{b}{1+bs}}{\frac{c}{1+cs}} = \frac{ab}{c}.$$

Then, $f(0) = \frac{ab}{c} = \frac{z_{min}^P}{b_{max}^P} > a$. If $c > b$, then f is decreasing, so $\inf \big\{ f(s) : s \geq 0 \big\} = \lim\limits_{s \to +\infty} f(s) = a$ by L'Hôpital's Rule [123]. To sum up, $\inf\limits_{s \,\geq\, 0} f(s) = \max \big(a,\, \frac{ab}{c} \big) \leq r_q$. □

We can upper bound r_q using a similar approach.

Theorem 23: If $\text{rank}(\bar{B}) = n$, A is Hurwitz, and system (7.2) is resiliently stabilizable, then

$$r_q \leq \max \left(\frac{\lambda_{max}^P \lambda_{max}^Q}{\lambda_{min}^P \lambda_{min}^Q}, \, \frac{z_{max}^P}{b_{min}^P} \right), \tag{7.17}$$

for any $P \succ 0$ and $Q \succ 0$ such that $A^\top P + PA = -Q$.

Proof. With our assumptions we are allowed to use Propositions 23 and 24. We define the positive constants

$$a := \frac{\lambda_{max}^P \lambda_{max}^Q}{\lambda_{min}^P \lambda_{min}^Q}, \qquad b := \frac{\lambda_{min}^Q}{2\lambda_{max}^P b_{min}^P}, \qquad \text{and} \qquad c := \frac{\lambda_{max}^Q}{2\lambda_{min}^P z_{max}^P},$$

so that for $x_0 \in \mathbb{R}^n$, $x_0 \neq 0$, (7.12) and (7.14) yield

$$\frac{T_N^*(x_0)}{T_M^*(x_0)} \leq a \frac{\ln(1 + b\|x_0\|_P)}{\ln(1 + c\|x_0\|_P)} := g(\|x_0\|_P).$$

Then, according to (7.5), $r_q \leq \inf\limits_{x_0 \,\in\, \mathbb{R}^n} g(\|x_0\|_P)$. This function g is similar to f in the proof of Theorem 22, and thus $r_q \leq \inf\limits_{x_0 \,\in\, \mathbb{R}^n} g(\|x_0\|_P) = \max \big(a,\, a\frac{b}{c} \big)$, yielding (7.17). □

Note that we used the same pair (P, Q) to bound both T_N^* and T_M^*. Employing different pairs (P_N, Q_N) and (P_M, Q_M) would make f depend on both $\|x_0\|_{P_N}$ and $\|x_0\|_{P_M}$. Then, we would need to take $x_0 \in \mathbb{R}^n$ instead of $\|x_0\|_P \in \mathbb{R}^+$ as the argument of f, which would significantly complicate the minimum search. We leave this more convoluted approach for possible future work.

Following the discussion at the beginning of this section 7.6, recall that an exact calculation of r_q is impossible. Hence, by deriving bounds on r_q with Theorems 22 and 23, we provided a considerable if not complete solution to Problem 12. We will now apply our theory to two numerical examples.

7.7 Numerical results

We will first study the resilient reachability of the ADMIRE fighter jet model [106], before quantifying the resilience of a temperature control system.

7.7.1 Resilient reachability of the ADMIRE fighter jet model

The ADMIRE model developed by the Swedish Armed Forces [106] has been used as an application case in numerous control frameworks [35], [108] and is illustrated on Fig. 7.1.

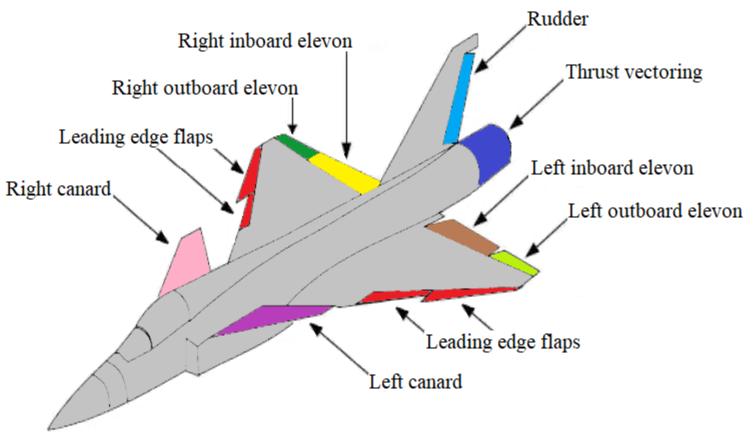


Figure 7.1: The ADMIRE fighter jet model. Image modified from [106] with a different color for each independent actuator.

Relying on the simulation package *Admirer4p1*[1] we run the ADMIRE simulation in MATLAB and obtain the linearized dynamics at Mach 0.3 and altitude $2000\,m$. We scale $\bar{B}$ so that the input set of each actuator from [106] is scaled to $[-1, 1]$. The states and matrices of the system $\dot{X}(t) = AX(t) + \bar{B}\bar{u}(t)$ are given below.

Consider a scenario in which, after sustaining damage, an actuator of the fighter jet starts producing uncontrolled and possibly undesirable inputs. By studying $\bar{B}$, we gain intuition on the resilience of the jet. The effect of the yaw (resp. pitch) thrust vectoring on the yaw (resp. pitch) rate is larger than that of all the other actuators combined, which gives the intuition that the jet is not resilient to the loss control over thrust vectoring. None of the other actuators produce such a dominant effect, hence giving the intuition that the jet is resilient to the loss of control over any one of the first eight actuators.

Following Lemma 22, we test our intuition by verifying whether $C\mathcal{W} \subseteq B\mathcal{U}$. These sets are zonotopes of dimension 9, represented in MATLAB using function *zonotope*$(\cdot)$ from the CORA package [62]. The associated function *in*$(\cdot)$ is employed to verify their inclusion. As expected, $C\mathcal{W} \subseteq B\mathcal{U}$ for the loss of control over any one actuator except for the thrust vectoring ones, as shown on Fig. 7.2. Note that for any projection $\text{proj}(\cdot)$, we have $\text{proj}(C\mathcal{W}) \nsubseteq \text{proj}(B\mathcal{U})$ implies $C\mathcal{W} \nsubseteq B\mathcal{U}$, but $\text{proj}(C\mathcal{W}) \subseteq \text{proj}(B\mathcal{U})$ does not yield $C\mathcal{W} \subseteq B\mathcal{U}$.

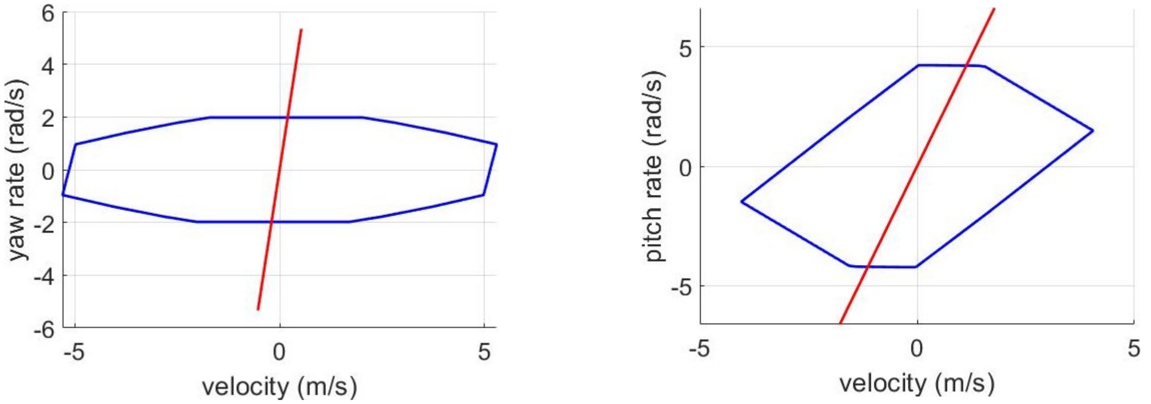
The eigenvalues of A are $\lambda(A) = \big\{-2.79, -1.58, -0.18 \pm 1.71i, -0.09, -0.02, 0, 0, 1.23\big\}$. Hence, none of the conditions $Re(\lambda(A)) \leq 0$ or $Re(\lambda(A)) = 0$ are verified. The system is neither resilient nor resiliently stabilizable. However, as anticipated with Problem 11, the linearized model is only valid locally and hence we should only study the resilient reachability of targets close to the linearization equilibrium

$$X_{eq} = \big(102.9\,m/s, 0.12\,rad, 0, 0, 0, 0, 0, 0.12\,rad, 0\big).$$

The state X represents small variations around X_{eq}, hence X is around 0. For $x_0 = 0$, the center of all zonotopes $\tilde{\Omega}_i$ stays at 0.

We follow the method detailed in Section 7.5 to approximate the resiliently reachable set of the malfunctioning system. Assume the pilot lost control over the right outboard elevon $\bar{u}_3$. We use the CORA [62] function *minus*$(\cdot,\cdot)$ to underapproximate the Minkowski difference $\mathcal{Z} = B\mathcal{U} \ominus C\mathcal{W}$ as a zonotope $(0, g_1, \ldots, g_9)$, following the method of [82]. We take $T = 0.2\,s$, $N = 5$, $\delta t = T/N$, $\Omega_0 = \{x_0\}$, and the

[1] https://github.com/Jean-BaptisteBouvier/ADMIRE



(a) Yaw thrust vectoring.

(b) Pitch thrust vectoring.

Figure 7.2: 2D projection of sets $B\mathcal{U}$ (blue) and $C\mathcal{W}$ (red) for the loss of control over the two thrust vectoring actuators.

$$X = \begin{pmatrix} v \\ \alpha \\ \beta \\ p \\ q \\ r \\ \psi \\ \theta \\ \varphi \end{pmatrix} \begin{array}{l} \text{velocity } (m/s), \\ \text{angle of attack } (rad), \\ \text{sideslip angle } (rad), \\ \text{roll rate } (rad/s), \\ \text{pitch rate } (rad/s), \\ \text{yaw rate } (rad/s), \\ \text{heading angle } (rad), \\ \text{pitch angle } (rad), \\ \text{roll angle } (rad), \end{array} \qquad A = \begin{pmatrix} -0.02 & -4.65 & 0.37 & 0 & -0.3 & 0 & 0 & -9.81 & 0 \\ 0 & -0.78 & 0.01 & 0 & 0.97 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.19 & 0.12 & 0 & -0.98 & 0 & 0 & 0.1 \\ 0 & 0 & -15.47 & -1.5 & 0 & 0.54 & 0 & 0 & 0 \\ 0 & 4.18 & -0.01 & 0 & -0.78 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.95 & -0.09 & 0 & -0.34 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0.12 & 0 & 0 & 0 \end{pmatrix}$$

$$\bar{B}^\top = \begin{pmatrix} -0.62 & 0 & 0 & 0.37 & 0.67 & -0.19 & 0 & 0 & 0 \\ -0.62 & 0 & 0 & -0.37 & 0.67 & 0.19 & 0 & 0 & 0 \\ -0.4 & -0.02 & 0 & -2.27 & -0.55 & -0.1 & 0 & 0 & 0 \\ -0.62 & -0.04 & 0.01 & -1.96 & -0.88 & -0.22 & 0 & 0 & 0 \\ -0.62 & -0.04 & -0.01 & 1.96 & -0.88 & 0.22 & 0 & 0 & 0 \\ -0.4 & -0.02 & 0 & 2.27 & -0.55 & 0.1 & 0 & 0 & 0 \\ -0.16 & 0 & 0.02 & 1.59 & 0 & -0.96 & 0 & 0 & 0 \\ 0.08 & 0 & 0 & 0 & -0.02 & 0 & 0 & 0 & 0 \\ -0.53 & 0 & 0.11 & -0.64 & 0.01 & -5.34 & 0 & 0 & 0 \\ -1.78 & -0.11 & 0 & 0 & -6.63 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{array}{l} \text{right canard,} \\ \text{left canard,} \\ \text{right outboard elevon,} \\ \text{right inboard elevon,} \\ \text{left inboard elevon,} \\ \text{left outboard elevon,} \\ \text{rudder,} \\ \text{leading edge flaps,} \\ \text{yaw thrust vectoring,} \\ \text{pitch thrust vectoring.} \end{array}$$

zonotope

$$\tilde{V} = \left(0, \int_0^{\delta t} e^{A(\delta t - t)} g_1 \, dt, \dots, \int_0^{\delta t} e^{A(\delta t - t)} g_9 \, dt\right).$$

Then, we underapproximate $R(T, x_0)$ with $\tilde{\Omega}_N$ using the recursion $\tilde{\Omega}_{i+1} = e^{A\delta t}\tilde{\Omega}_i \oplus \tilde{V}$ of Section 7.5.

Since the malfunctioning actuator $\bar{u}_3$ has a strong impact on the roll rate p of the jet, we want to see what range of roll rates is reachable. We compute $\tilde{\Omega}_1, \dots, \tilde{\Omega}_N$ and project them in 2D as shown on Fig. 7.3. Then, in time T the jet can change its roll rate up to $\pm 1.2\, rad/s$, despite the loss of control over the right outboard elevon.

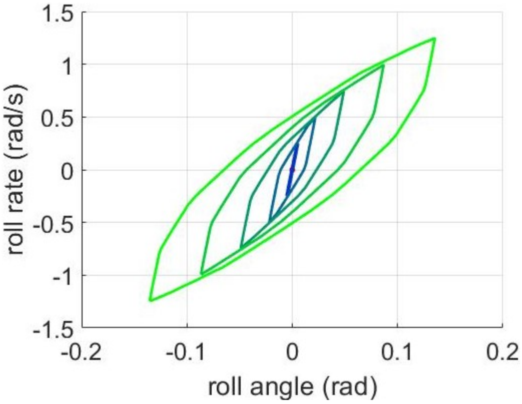


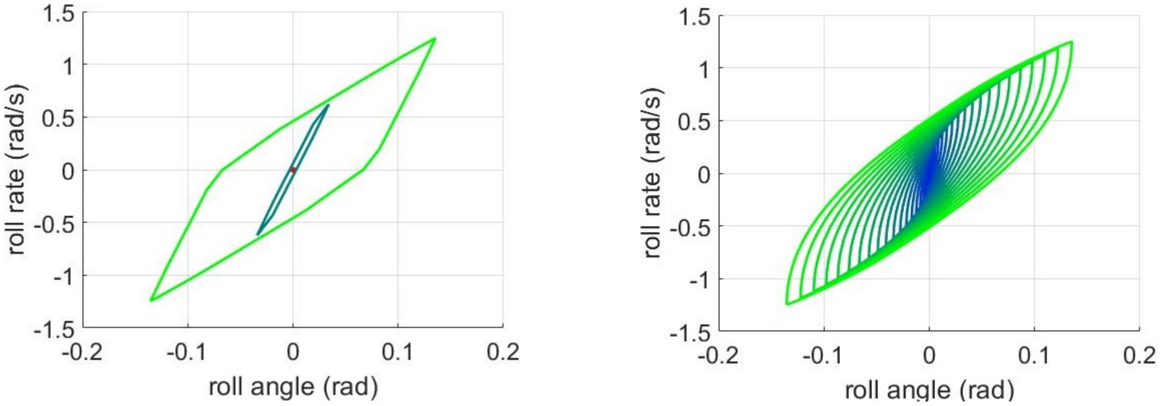
Figure 7.3: Projection of $\tilde{\Omega}_1, \dots, \tilde{\Omega}_5$ on the (ϕ, p) plane.

We now study the impact of N, i.e., of δt on the precision of $\tilde{\Omega}_N$ to approximate the real reachable set $R(T, x_0)$ when keeping T constant. Since $\dim\big(R(T, x_0)\big) = 9$, we will only study the impact on the range of roll rates reachable at roll angle $\phi = 0\, rad$. For $\delta t = 0.1\, s$ the reachable range of roll rate around c_N is $\pm 0.37\, rad/s = \pm 21.2°/s$, while for $\delta t = 0.04\, s$ it is $\pm 0.42\, rad/s = \pm 24.1°/s$, and $\pm 0.43\, rad/s = \pm 24.6°/s$ for $\delta t = 0.01\, s$. Hence, as explained in Section 7.5, decreasing δt gives a wider under approximation of the reachable set. For $N = 2$ the reachable range of roll rate is $\pm 0.37\, rad/s = \pm 21.2°/s$, while for $N = 5$ it is $\pm 0.42\, rad/s = \pm 24.1°/s$, and $\pm 0.43\, rad/s = \pm 24.6°/s$ for $N = 20$, as illustrated on Fig. 7.4. For $N = 2$ the reachable range of roll rates is $\pm 0.37\, rad/s$, while for $N = 5$ it is $\pm 0.42\, rad/s$, and $\pm 0.43\, rad/s$ for $N = 20$, as illustrated on Fig. 7.3 and 7.4. Hence, as explained in Section 7.5, increasing N raises nonlinearly the precision of $\tilde{\Omega}_N$ and increases linearly the computational cost since $\tilde{\Omega}_N$ is a zonotope with $9N$ generators, as the Minkowski addition of $\tilde{V}$ adds 9 generators to $\tilde{\Omega}_i$ at each iteration.

Now assume that the in-flight damage responsible for the loss of control over the elevon $\bar{u}_3$ also initially caused it to jerk resulting in a sudden jump in roll rate. Then, instead of $X(0) = 0$ we have $p(0) = 0.44\, rad/s$ and the goal is to stabilize the jet at the origin X_{tg}.

We can see on Fig. 7.5 that the target only enters the projection of the reachable set after 4 iterations of $\delta t = 0.04\, s$, i.e., for $t \geq 0.16\, s$. By choosing a smaller δt we can refine the precision on the minimal entering time. However, to calculate the reachable time $T_M^*(X_0, X_{tg})$ we need to use the CORA function $in(\cdot)$ to verify whether $X_{tg} \in \tilde{\Omega}_N$ since Fig. 7.5 is only a 2D projection of the 9D reachable set and could be deceiving. Indeed, for $p(0) = 0.5\, rad/s$, the 2D projection is similar to Fig. 7.5 with the red dot inside the projection of $\tilde{\Omega}_N$, but $X_{tg} \notin \tilde{\Omega}_N$.

We successfully demonstrated the developed resilience theory and the zonotopic method to underapproximate the resiliently reachable set of the ADMIRE jet model.



(a) $N = 2$.

(b) $N = 20$.

Figure 7.4: Projection of $\tilde{\Omega}_1, \ldots, \tilde{\Omega}_N$ on the (ϕ, p) plane for different values of N.

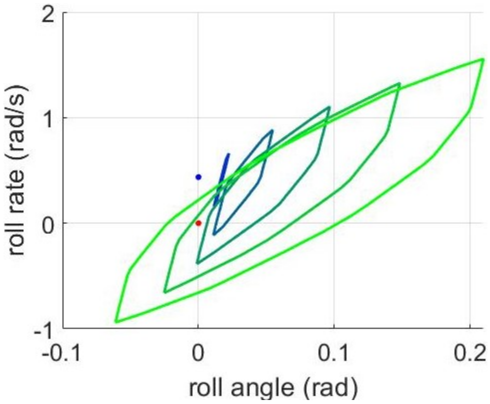


Figure 7.5: Projection of $\tilde{\Omega}_1, \ldots, \tilde{\Omega}_5$ on the (ϕ, p) plane. Initial state X_0 is the blue dot, target X_{tg} is the red dot, and $N = 5$.

7.7.2 Temperature control system

We now illustrate our quantitative resilience bounds on a temperature control system motivated by [124] and illustrated on Fig. 7.6.

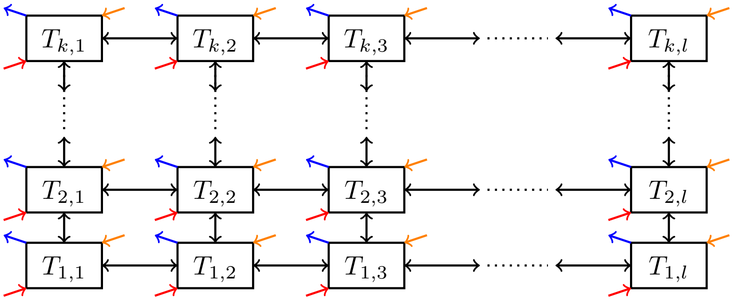


Figure 7.6: Heat exchange graph of an office building with k floors of l rooms, each at a temperature $T_{i,j}$.

We study a scenario where a worker remains in their office after hours and manually opens or closes their door and window, thus overriding the building heat controller which aims at maintaining a target temperature T_{tg}. After this loss of control, we will compare our analytical bounds on the nominal and malfunctioning reach times with the numerical results of [30], [71]. We will also bound the quantitative resilience of the system which could not be done with prior work and motivated the analytical bounds of Section 7.6.

The controller uses a central heater q_h, central AC q_{AC}, and incrementally opens doors q_d and windows q_w for room specific adjustments. The controller also takes advantage of solar heating q_S, heat losses through the outside wall q_l, and heat transfers between adjoining rooms q_{adj}. The temperature dynamics are then

$$mC_p\dot{T}_{i,j} = q_h - q_{AC} + q_{d_{i,j}} - q_{w_{i,j}} + q_{S_{i,j}} - q_{l_{i,j}} + \sum q_{adj}$$

with m the mass of air in each room, C_p its specific heat capacity, $q_{adj} = aU(T_{adj} - T_{i,j})$, with a the area of the wall between rooms, and U the overall heat transfer coefficient between adjoining rooms, which depends on the wall materials. To have symmetric inputs, we combine the heat transfers in pairs: $q_h - q_{AC} =: Q_{hAC}u_{hAC}$, $q_{d_{i,j}} - q_{w_{i,j}} =: Q_{dw}u_{dw}^{i,j}$, and $q_{S_{i,j}} - q_{l_{i,j}} =: Q_{Sl}u_{Sl}^{i,j}$ with u_{hAC}, $u_{dw}^{i,j}$, and $u_{Sl}^{i,j} \in [-1, 1]$.

We write the dynamics as $\dot{T} = AT + \bar{B}\bar{u}$, with

$$A = \frac{a}{mC_p}\begin{pmatrix} -2U & U & 0 & 0 & \dots & 0 & U & 0 & 0 & \dots \\ U & -3U & U & 0 & \dots & 0 & 0 & U & 0 & \dots \\ 0 & \ddots & & \ddots & & \ddots & & \ddots & & \ddots \end{pmatrix}, \qquad \bar{u} = \begin{pmatrix} u_{Sl}^{1,1} \\ \vdots \\ u_{Sl}^{k,l} \\ u_{dw}^{1,1} \\ \vdots \\ u_{dw}^{k,l} \\ u_{hAC} \end{pmatrix} \in \mathbb{R}^{2kl+1},$$

$$\bar{B} = \frac{1}{mC_p}\begin{pmatrix} Q_{Sl}I_{kl,kl} & Q_{dw}I_{kl,kl} & Q_{hAC}\mathbf{1}_{kl} \end{pmatrix}, \qquad \text{and} \qquad T = \begin{pmatrix} T_{1,1} \\ \vdots \\ T_{k,l} \end{pmatrix} \in \mathbb{R}^{kl}.$$

To perform numerical calculations, we restrict our building to $k = 1$ and $l = 3$, as schematized in Fig. 7.7.

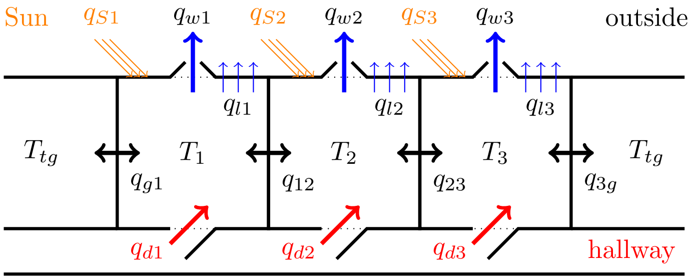


Figure 7.7: Scheme of the rooms and of the heat transfers. The heater q_h and AC transfers q_{AC} are not shown for clarity.

The objective is to make the rooms 1, 2, and 3 reach target temperature T_{tg}, which is the temperature of the neighboring rooms as shown on Fig. 7.7. We write the dynamics as $\dot{T} = AT + \bar{B}\bar{u} + DT_{tg}$, with

$$A = \frac{1}{mC_p}\begin{pmatrix} -U_{g1} - U_{12} & U_{12} & 0 \\ U_{12} & -U_{12} - U_{23} & U_{23} \\ 0 & U_{23} & -U_{23} - U_{3g} \end{pmatrix}, \qquad D = \frac{1}{mCp}\big(U_{g1}, 0, U_{3g}\big),$$

and $x = \big(x_1, x_2, x_3\big) = T - \mathbf{1}T_{tg}$. Then,

$$\dot{x} = \dot{T} = Ax + \bar{B}\bar{u} + DT_{tg} + A\mathbf{1}_3 T_{tg} = Ax + \bar{B}\bar{u},$$

and $x_{tg} = \big(0, 0, 0\big)$. Taking $x := T - T_{tg}$, the heat dynamics of the system illustrated on Fig. 7.7 are $\dot{x} = Ax + \bar{B}\bar{u}$ with $x_{tg} = 0$. Based on [124], we use the following values: $a = 12\,m^2$, $mC_p = 42186\,J/K$, $U_{g1} = 6.27\,W/K$, $U_{12} = 5.08\,W/K$, $U_{23} = 5.41\,W/K$, $U_{3g} = 6.27\,W/K$, $Q_{hAC} = 350\,W$, $Q_{dw} = 300\,W$, $Q_{Sl} = 200\,W$, and $T_{tg} = 293\,K$.

Since $\lambda(A) = \big\{-0.052, -0.033, -0.010\big\} \subseteq \mathbb{R}^-$, A is Hurwitz. Then, according to Theorem 21, the system is not resilient, but it might be resiliently stabilizable. For the loss of any one column C, $\text{rank}(B) = 3$ and we numerically verify that $-C\mathcal{W} \subseteq \text{int}(B\mathcal{U})$. Then, following Lemma 19, $\dim(\mathcal{Z}) = 3$, so $\text{int}(\mathcal{Z}) \neq \emptyset$. According to Proposition 19, the system is resiliently stabilizable.

The controller wants to cool the building overnight from an initial state $x_0^\top = \big(0.8°C,\ 0.7°C,\ 0.9°C\big)$. However, a worker is overriding u_{dw}^1 by manually opening the door and window in room 1. We now compare the analytical bounds on the nominal and malfunctioning reach times of Section 7.6 with the numerical results of [30], [71]. Our bounds require pairs $P \succ 0$ and $Q \succ 0$ solutions of $A^\top P + PA = -Q$. We generate randomly a thousand of such pairs (P, Q) and compute bounds on T_N^* with (7.10) and (7.12), and on T_M^* with (7.14) and (7.15). Another way of choosing P relies on the linearization of (7.14), which yields $T_M^* \geq \frac{\|x_0\|_P}{z_{max}^P}$. This bound is maximized when $P \succ 0$ is the tightest ellipsoidal approximation of $\mathcal{Z}$, which results in much tighter bound than stochastic P, as shown on Fig. 7.8.

For the given x_0 the best bounds on the reach times are

$$\begin{aligned} 35.5\,s \ &\leq\ T_N^*(x_0) = 42.5\,s \ \leq\ 54.1\,s, \\ 53\,s \ &\leq\ T_M^*(x_0) = 110.5\,s \ \leq\ 135\,s. \end{aligned}$$

Then, the rooms can take up to $T_M^*(x_0)/T_N^*(x_0) = 2.6$ times longer to all reach T_{tg} from the initial state $T_{tg} + x_0$ after the loss of control authority over u_{dw}^1, while our bounds predict a worst-case factor of 3.8.

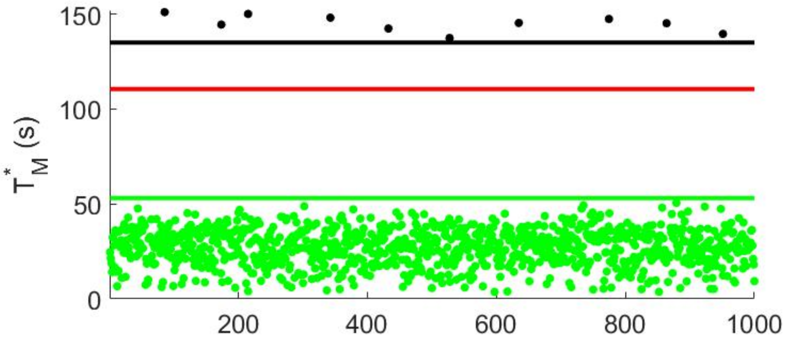


Figure 7.8: Bounds on the malfunctioning reach time $T_M^*(x_0)$ in red. The dots are the upper (7.15) and lower bounds (7.14) for 1000 stochastic pairs (P, Q). The tightest bounds in green and black result from the ellipsoidal approximations of $\mathcal{Z}$.

The resulting bounds on T_M^* correspond to the dots on Fig. 7.8. We obtained better bounds when choosing $P \succ 0$ as the tightest ellipsoidal approximation of $\mathcal{Z}$, as shown on Fig. 7.8. Another approach relies on the fact that for y small, $\ln(1+y) \approx y$. Thus, the lower bound of T_M^* in (7.14) can be approximated by $\frac{\|x_0\|_P}{z_{max}^P}$. To maximize this lower bound, we minimize $z_{max}^P = \max\big\{\|z\|_P : z \in \mathcal{Z}\big\}$, i.e., we choose $P \succ 0$ generating the tightest ellipsoid outer approximation of $\mathcal{Z}$. Similarly, to minimize the upper bound (7.15), we need P to generate the largest ellipsoid inside $\mathcal{Z}$. Then, we take $Q = -A^\top P - PA$, but there is no guarantee that $Q \succ 0$.

We were able to compute numerically $T_N^*(x_0)$ [71] and $T_M^*(x_0)$ [30], but accessing r_q can only be done analytically with Theorems 22 and 23. Over all $x_0 \in \mathbb{R}^3$, they predict $r_q \in [0.166, 0.979]$. Hence, the loss of control over u_{dw}^1 can render the damaged system up to $1/0.166 = 6$ times slower to reach the target temperature from any initial state. This information could not be obtained with prior work and is the motivation for our analytical bounds in Section 7.6.

If instead of losing control over u_{dw}^1 a disgruntled worker takes over the central heating/AC unit u_{hAC}, the rooms can take as much as $T_M^*(x_0)/T_N^*(x_0) = 4.7$ times longer to reach T_{tg} from the same initial temperature, while our bound predicts a max ratio of 9.3. These values are larger than for the loss of u_{dw}^1 because $Q_{hAC} > Q_{dw}$ and the central heating/AC affects directly all 3 rooms. Additionally, Theorem 22 yields $r_q \in [0.1, 0.37]$, so the malfunctioning controller can take between 2.7 and 10 times longer than nominally to enforce the target temperature from any initial condition.

7.8 Supporting lemmata

In this section we provide supporting results concerning sets $\mathcal{BU}$, $\mathcal{CW}$, and $\mathcal{Z}$ defined in Section 7.3.

Lemma 17: The interior of $\mathcal{Z}$ is non-empty if and only if $0 \in \text{int}(\mathcal{Z})$.

Proof. Since $\mathcal{Z}$ is convex and symmetric, so is its interior [125]. If $\text{int}(\mathcal{Z}) \neq \emptyset$, there exists $z \in \text{int}(\mathcal{Z})$, by symmetry $-z \in \text{int}(\mathcal{Z})$, and $0 \in \text{int}(\mathcal{Z})$ by convexity. The reverse implication is trivial. □

Lemma 18: The following statements are equivalent: (a) $0 \in \text{relint}(\mathcal{Z})$, (b) $0 \in \mathcal{Z}$, (c) $\mathcal{Z} \neq \emptyset$, (d) $\text{relint}(\mathcal{Z}) \neq \emptyset$.

Proof. Since $\text{relint}(\mathcal{Z}) \subseteq \mathcal{Z}$, we have $(a) \implies (b)$ and trivially, $(b) \implies (c)$. Since $\mathcal{Z}$ is a convex subset of $\mathbb{R}^n$, $(c) \implies (d)$ according to Lemma 7.33 of [1]. Because $\mathcal{Z}$ is convex and symmetric, so is its relative interior according to [125]. Then, the same proof as for Lemma 17 yields $(d) \implies (a)$ which completes the proof. □

Definition 19: The dimension of a compact set $\mathcal{S}$ is the dimension of the smallest affine subspace (with respect to inclusion) containing $\mathcal{S}$ [1].

Lemma 19: The relative interior of $B\mathcal{U}$ contains $-C\mathcal{W}$ if and only if $\dim(\mathcal{Z}) = \operatorname{rank}(B)$.

Proof. Let $q := \dim(B\mathcal{U}) \leq n$. Since $\mathcal{U} = [-1,1]^{m-p}$, its interior is not empty in $\mathbb{R}^{m-p}$ and thus $q = \operatorname{rank}(B)$. Take q linearly independent vectors of $B\mathcal{U}$ denoted by $B_q := (b_1, \dots, b_q)$ and pick $V := (v_{q+1}, \dots, v_n) \in \mathbb{R}^{n \times (n-q)}$ such that $T_b := (B_q, V)$ is invertible. Then, T_b is a transition matrix with $T_b e_i = b_i$ for $i \in [\![1, q]\!]$.

Assume first that $-C\mathcal{W} \subseteq \operatorname{relint}(B\mathcal{U})$. Then, there exists $\varepsilon > 0$ such that $T_b\big(\mathbb{B}^q(0,\varepsilon) \times \{0\}^{n-q}\big) \oplus -C\mathcal{W} \subseteq B\mathcal{U}$. Informally, $-C\mathcal{W}$ remains in $B\mathcal{U}$ when it is 'extended' by ε in all q dimensions of $B\mathcal{U}$. Because $\mathcal{Z} = \big\{z \in \mathbb{R}^n : \{z\} \oplus -C\mathcal{W} \subseteq B\mathcal{U}\big\}$, we have $T_b\big(\mathbb{B}^q(0,\varepsilon) \times \{0\}^{n-q}\big) \subseteq \mathcal{Z}$. Then, $q \leq \dim(\mathcal{Z})$. Since $0 \in -C\mathcal{W}$, $\mathcal{Z} \subseteq B\mathcal{U}$, and hence $\dim(\mathcal{Z}) \leq q$. Thus, $\dim(\mathcal{Z}) = q = \operatorname{rank}(B)$.

On the other hand, assume that $\dim(\mathcal{Z}) = q$. Since $0 \in -C\mathcal{W}$, $\mathcal{Z} \subseteq B\mathcal{U}$. Then, $\mathcal{Z}$ being of same dimension and included in $B\mathcal{U}$ yields that $(b_1, \dots, b_q)$ is also a basis of $\operatorname{span}(\mathcal{Z}) = \operatorname{Im}(B)$. Hence, T_b is a transition matrix from $\mathbb{R}^n$ to $\operatorname{span}(\mathcal{Z})$. According to Lemma 18, $0 \in \operatorname{relint}(\mathcal{Z})$, i.e, there exists $\delta > 0$ such that $T_b\big(\mathbb{B}^q(0,\delta) \times \{0\}^{n-q}\big) \subseteq \mathcal{Z}$. As above, the definition of $\mathcal{Z}$ yields $T_b\big(\mathbb{B}^q(0,\delta) \times \{0\}^{n-q}\big) \oplus (-C\mathcal{W}) \subseteq B\mathcal{U}$. Because $\dim(\mathbb{B}^q(0,\varepsilon)) = q = \dim(B\mathcal{U})$, we have $-C\mathcal{W} \subseteq \operatorname{relint}(B\mathcal{U})$. □

Lemma 20: If $\dim(\mathcal{Z}) = \operatorname{rank}(B)$, then $\operatorname{span}(\mathcal{Z}) = \operatorname{Im}(B) = \operatorname{Im}(\bar{B})$.

Proof. In the proof of Lemma 19 we showed that $\operatorname{span}(\mathcal{Z}) = \operatorname{Im}(B)$. The inclusion $-C\mathcal{W} \subseteq \operatorname{relint}(B\mathcal{U})$ holds according to Lemma 19 and yields $\operatorname{Im}(C) \subseteq \operatorname{Im}(B)$, and since $\bar{B} = [B\ C]$ after adequate column permutations, we have $\operatorname{Im}(\bar{B}) = \operatorname{Im}([B\ C]) = \operatorname{Im}(B)$. □

Lemma 21: Set $\mathcal{Z}$ is empty if and only if set $C\mathcal{W}$ is not entirely included in $B\mathcal{U}$, i.e., $\mathcal{Z} = \emptyset \iff C\mathcal{W} \nsubseteq B\mathcal{U}$.

Proof. If $\mathcal{Z} = \emptyset$, then by definition, for all $z \in B\mathcal{U}$, there exists $w \in \mathcal{W}$ such that $z - Cw \notin B\mathcal{U}$. Taking $z = 0$ yields $C\mathcal{W} \nsubseteq B\mathcal{U}$.

On the other hand, assume that there exists $w \in \mathcal{W}$ such that $Cw \notin B\mathcal{U}$. Assume for contradiction purposes that $\mathcal{Z} \neq \emptyset$. Then, we can take $z \in \mathcal{Z}$ and $z - Cw \in B\mathcal{U}$. Since $B\mathcal{U}$ is symmetric, we thus have $-z + Cw \in B\mathcal{U}$. Because $z \in \mathcal{Z}$ and $-w \in \mathcal{W}$, we also have $z + Cw \in B\mathcal{U}$. The convexity of $B\mathcal{U}$ yields $\frac{1}{2}(-z + Cw) + \frac{1}{2}(z + Cw) \in B\mathcal{U}$, i.e., $Cw \in B\mathcal{U}$ which contradicts our first assumption. Hence, $\mathcal{Z} = \emptyset$. □

Lemma 22: If $C\mathcal{W} \nsubseteq B\mathcal{U}$, then system (7.2) is not resiliently stabilizable.

Proof. Since $C\mathcal{W} \nsubseteq B\mathcal{U}$, there exists $w \in \mathcal{W}$ such that $Cw \notin B\mathcal{U}$. The sets $\{Cw\}$ and $B\mathcal{U}$ are nonempty, disjoint, convex, and compact, hence they are strongly separated according to Theorem 5.79 of [1]. Then, there exists $v \in \mathbb{R}^n$, $v \neq 0$, $c > 0$, and $\varepsilon > 0$ such that $\langle Cw, v\rangle \geq c + \varepsilon$, and for all $u \in \mathcal{U}$, $\langle Bu, v\rangle \leq c - \varepsilon$. Because $B\mathcal{U}$ and $C\mathcal{W}$ are symmetric, $\{-Cw\}$ and $B\mathcal{U}$ are also strongly separated by the symmetric hyperplane: $\langle -Cw, v\rangle \leq -c - \varepsilon$ and for all $u \in \mathcal{U}$, $\langle Bu, v\rangle \geq -c + \varepsilon$.

If $A \neq 0$, then $\|A\| > 0$. Since $v \neq 0$, we can define $r := \frac{\varepsilon}{\|v\|\,\|A\|} > 0$. We will show that if $x \in \mathbb{B}^n(0, r)$, then no controls $u \in \mathcal{U}$ can bring the state x closer to the origin. Let $x \in \mathbb{B}^n(0, r)$ and first assume that $\langle x, v\rangle \geq 0$. Then, we apply the undesirable input w and any control $u \in \mathcal{U}$ to system (7.2)

$$\langle \dot{x}, v\rangle = \langle Ax, v\rangle + \langle Bu, v\rangle + \langle Cw, v\rangle \geq -\|Ax\|\,\|v\| - c + \varepsilon + c + \varepsilon \geq -\|A\|\,\|x\|\,\|v\| + 2\varepsilon \geq \varepsilon,$$

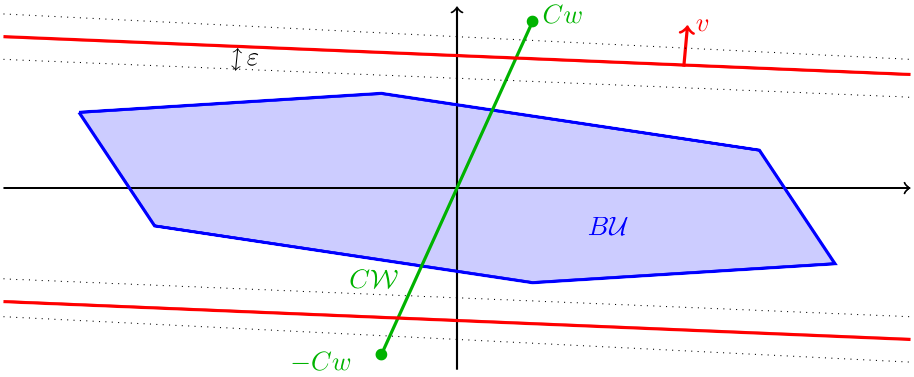


Figure 7.9: Illustration of the strong separation of sets $B\mathcal{U}$ (blue) and $\{\pm Cw\}$ (green) by symmetric hyperplanes.

where we used the Cauchy-Schwarz inequality [100], the definition of $\|A\|$ and $\|x\| \leq r$. Similarly, if $\langle x, v\rangle < 0$, we apply the undesirable input $-w$ and any control $u \in \mathcal{U}$ to system (7.2)

$$\langle \dot{x}, v\rangle = \langle Ax, v\rangle + \langle Bu, v\rangle + \langle -Cw, v\rangle \leq \|A\|\,\|x\|\,\|v\| + c - \varepsilon - c - \varepsilon \leq r\|A\|\,\|v\| - 2\varepsilon = -\varepsilon.$$

Thus, the state $x \in \mathbb{B}^n(0, r)$ can be pushed away from the origin along v. Hence, system (7.2) is not stabilizable.

If $A = 0$, we can take any $x \in \mathbb{R}^n$ such that $\langle x, v\rangle \geq 0$ (resp. ≤ 0) and obtain $\langle \dot{x}, v\rangle \geq 2\varepsilon$ (resp. $\leq -2\varepsilon$) so the same conclusion holds. □

7.9 Summary

This chapter established novel necessary and sufficient conditions for the resilient stabilizability of linear systems with drift. We also extended Hájek's duality theorem [23] in order to study the resilient reachability of affine targets. We then used zonotopic underapproximations of reachable sets to determine what states are guaranteed to be resiliently reachable. Finally, we employed Lyapunov theory [121] to quantify the resilience of control systems to the loss of authority over some of their actuators.

Chapter 8

Extensions of Resilience Theory

8.1 Introduction

This chapter is taken from our work [39] and is a first step towards Problem 4. In the previous chapters, we only considered systems with a resilience reachability mission. We also assumed that the controller had immediate knowledge of the undesirable inputs. Additionally, all the work accomplished so far only concerned linear dynamics. In this chapter, we will lift these three restrictive and simplifying assumptions in order to extend the scope of resilience theory as prescribed by Problem 4.

The contributions of this chapter are threefold. In Section 8.2, we extend resilience from the simple objective of resilient reachability to the more complex and more realistic aim of resilient trajectory tracking. In Section 8.3, we remove one of the main simplifying assumptions of resilience theory, namely the instantaneous knowledge of the undesirable input by the controller. Finally, in Section 8.4 we present a partial extension of resilience theory to nonlinear systems.

8.2 Resilient trajectory tracking

In this section, we assume that the mission of the nominal system under study is not to simply reach a target, but instead to follow a given reference trajectory $\mathcal{T}_{\text{ref}} := \big\{ x_{\text{ref}}(t),\ t \geq 0 \big\}$. We will then aim at deriving conditions under which a linear system affected by a partial loss of control authority over its actuators can resiliently follow trajectory $\mathcal{T}_{\text{ref}}$.

Following Chapter 7, we study a linear system of initial dynamics

$$\dot{x}(t) = Ax(t) + \bar{B}\bar{u}(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \qquad \bar{u}(t) \in \bar{\mathcal{U}}, \tag{8.1}$$

where $\bar{\mathcal{U}}$ is an hyperrectangle in $\mathbb{R}^{m+q}$ and $A \in \mathbb{R}^{n \times n}$, $\bar{B} \in \mathbb{R}^{n \times (m+q)}$ constant matrices. After some malfunction, system (8.1) suffers a loss of control authority over q of its $m+q$ actuators. We then split the signal $\bar{u}$ into its controlled part $u \in \mathcal{F}(\mathcal{U})$ and its uncontrolled part $w \in \mathcal{F}(\mathcal{W})$ with $\mathcal{U}$ and $\mathcal{W}$ hyperrectangles in $\mathbb{R}^m$ and $\mathbb{R}^q$ respectively. Matrix $\bar{B}$ is accordingly split into $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{n \times q}$ such that the dynamics of the malfunctioning system are

$$\dot{x}(t) = Ax(t) + Cw(t) + Bu(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \qquad u(t) \in \mathcal{U}, \qquad w(t) \in \mathcal{W}. \tag{8.2}$$

Following the method detailed in Chapter 7, we study the resilience of system (8.2) with the results built on Hájek's theory [23] and relying on the modified dynamics

$$\dot{x}(t) = Ax(t) + z(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \qquad z(t) \in \mathcal{Z} := B\mathcal{U} \ominus (-C\mathcal{W}). \tag{8.3}$$

Note that $\mathcal{Z}$ represents the amount of control authority remaining to the malfunctioning system (8.2) after counteracting the worst undesirable input. We now assume to be in possession of the reference inputs z_{ref} to track reference trajectory $\mathcal{T}_{\text{ref}}$ with dynamics (8.3), i.e., $\mathcal{T}_{\text{ref}} = \big\{x_{\text{ref}}(t) : \dot{x}_{\text{ref}}(t) = Ax_{\text{ref}}(t) + z_{\text{ref}}(t),\ t \geq 0\big\}$ with $z_{\text{ref}} \in \mathcal{F}(\mathcal{Z}_{\text{ref}})$.

The initial state of the malfunctioning system x_0 is most likely not exactly equal to $x_{\text{ref}}(0)$, the initial state of reference trajectory $\mathcal{T}_{\text{ref}}$, which was designed before the malfunction occurred. We then need to design a tracking controller with robustness to uncertainty on the initial state. Moreover, if the difference $x_0 - x_{\text{ref}}(0)$ is not actively reduced, it can grow exponentially with time [122]. Thus, we need some extra control capability to counteract $x(t) - x_{\text{ref}}(t)$.

Formally, we define $\mathcal{Z}_\varepsilon$ as a compact set of $\mathbb{R}^n$ satisfying $0 \in \text{relint}(\mathcal{Z}_\varepsilon)$ and $\dim(\mathcal{Z}_\varepsilon) = n$. Input set $\mathcal{Z}_\varepsilon$ will be tasked with counteracting $x(t) - x_{\text{ref}}(t)$. For the robust tracking of $\mathcal{T}_{\text{ref}}$ to be admissible, we then need $\mathcal{Z}_\varepsilon \oplus \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}$. We now introduce the dynamics tasked with counteracting the initial state error

$$\dot{y}(t) = Ay(t) + z_\varepsilon(t), \quad y(0) = x_0 - x_{\text{ref}}(0), \quad z_\varepsilon(t) \in \mathcal{Z}_\varepsilon. \tag{8.4}$$

Proposition 26: If $\mathcal{Z}_\varepsilon \oplus \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}$ and system (8.4) is stabilizable in a finite time t_f, then the reference trajectory $\mathcal{T}_{\text{ref}}$ can be tracked exactly by system (8.2) after time t_f, i.e., $x(t) = x_{\text{ref}}(t)$ for all $t \geq t_f$.

Proof. Since system (8.4) is stabilizable in a finite time t_f, there exists a signal $z_\varepsilon \in \mathcal{F}(\mathcal{Z}_\varepsilon)$ on $[0, t_f]$ yielding $y(t_f) = 0$ in system (8.4). Since $0 \in \mathcal{Z}_\varepsilon$, we extend the control signal to $z_\varepsilon(t) = 0$ for all $t > t_f$. We now define the control law $z_{track}(t) := z_\varepsilon(t) + z_{\text{ref}}(t)$. Since, $\mathcal{Z}_\varepsilon \oplus \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}$, we have $z_{track}(t) \in \mathcal{Z}$ for all $t \geq 0$.

Let $w \in \mathcal{F}(\mathcal{W})$ be any undesirable input signal. Then, by definition of $\mathcal{Z}$, there exists $u \in \mathcal{F}(\mathcal{U})$ such that $Bu(t) = z_{track}(t) - Cw(t)$ for all $t \geq 0$. We now implement this controller for $T \geq t_f$ in system (8.2):

$$\begin{aligned}
x(T) &= e^{AT}\left(x_0 + \int_0^T e^{-At}\big(Bu(t) + Cw(t)\big)\,dt\right) = e^{AT}\left(x_0 + \int_0^T e^{-At}\big(z_\varepsilon(t) + z_{\text{ref}}(t)\big)\,dt\right)\\
&= e^{AT}\left(x_0 + \int_0^T e^{-At} z_\varepsilon(t)\,dt + e^{-AT}x_{\text{ref}}(T) - x_{\text{ref}}(0)\right),
\end{aligned}$$

because $x_{\text{ref}}(T) = e^{AT}\left(x_{\text{ref}}(0) + \int_0^T e^{-At} z_{\text{ref}}(t) dt\right)$. Then,

$$x(T) - x_{\text{ref}}(T) = e^{AT}\left(x_0 - x_{\text{ref}}(0) + \int_0^T e^{-At} z_\varepsilon(t)\,dt\right) = e^{AT}\left(x_0 - x_{\text{ref}}(0) + \int_0^{t_f} e^{-At} z_\varepsilon(t)\,dt\right),$$

since $z_\varepsilon(t) = 0$ for $t > t_f$. By definition of z_ε,

$$y(t_f) = 0 = e^{At_f}\left(y(0) + \int_0^{t_f} e^{-At} z_\varepsilon(t)\,dt\right), \quad \text{i.e.,} \quad x_0 - x_{\text{ref}}(0) + \int_0^{t_f} e^{-At} z_\varepsilon(t)\,dt = 0.$$

Therefore, $x(T) = x_{\text{ref}}(T)$ for all $T \geq t_f$. □

Proposition 26 states that as long as $\mathcal{Z}_\varepsilon + \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}$, there exists a finite time t_f after which any trajectory $\mathcal{T}_{\text{ref}}$ can be tracked perfectly despite the loss of control authority over a thruster. Since $\mathcal{Z}_\varepsilon$ is the input set of system (8.4), the size of $\mathcal{Z}_\varepsilon$ is inversely correlated with the stabilization time t_f. Then, the constraint $\mathcal{Z}_\varepsilon + \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}$ yields that the smaller $\mathcal{Z}_{\text{ref}}$, the larger $\mathcal{Z}_\varepsilon$ and so the smaller t_f is. In other words, the smaller the inputs required to track the reference trajectory, the faster the system can resume perfect tracking after a loss of control authority.

Lemma 23: Let $Z_\varepsilon \in \mathbb{R}^{n \times d}$ be a matrix whose columns are d linearly independent vectors of $\mathcal{Z}_\varepsilon$, with $d = \dim(\mathcal{Z}) = \dim(\mathcal{Z}_\varepsilon)$. System (8.4) is stabilizable in a finite time if and only if $Re(\lambda(A)) \leq 0$, $\text{rank}(\mathcal{C}(A, Z_\varepsilon)) = n$ and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}_\varepsilon$.

Proof. By construction $\text{Im}(Z_\varepsilon) = \text{span}(\mathcal{Z}_\varepsilon)$, so that $\mathcal{C}(A, Z_\varepsilon)$ is a controllability matrix associated with system (8.4). We made the assumption that $0 \in \text{relint}(\mathcal{Z}_\varepsilon)$, so we can apply Proposition 20 which guarantees that system (8.4) is stabilizable in finite time. □

We now have a condition to verify whether resilient trajectory tracking is possible. This also allows to fulfill resilient reach-avoid objectives, where some areas of the state space are judged unsafe and should not be entered. With resilient reachability, we had no control over the trajectory of the malfunctioning system between the two end points, which could have resulted in entering unsafe areas. With resilient trajectory tracking, we only need to design a safe trajectory that the malfunctioning system can follow and it will avoid the unsafe regions. Let us now investigate how the system would perform if the controller could not react instantly to undesirable inputs.

8.3 Resilience in the presence of actuation delay

In order to account for the unavoidable sensors and actuators delays present on any system [25]–[27], we now assume that the controller of system (8.2) operates with a constant input delay $\tau > 0$. Then, the dynamics of this malfunctioning system become

$$\dot{x}(t) = Ax(t) + Cw(t) + Bu\big(t, x(t-\tau), w(t-\tau)\big), \quad x(0) = x_0 \in \mathbb{R}^n, \quad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W}. \tag{8.5}$$

The controller cannot react immediately to a change of the undesirable input $w(t)$ and cancel it instantaneously as in the previous chapters. Note that in (8.5), the controller has also a delayed knowledge of the state. Indeed, if the controller $u(t)$ knows $x(t)$, with its knowledge of A, B and C it would be able to reconstruct $w(t)$, which should not be possible. Therefore, the controller must have a delayed knowledge of the state.

8.3.1 Framework for actuation delay

Let us plant the framework to study resilient reachability by system (8.5).

Definition 20: A convex set $\mathcal{G} \subseteq \mathbb{R}^n$ is *resiliently reachable* at time T from x_0 by system (8.5) if for all undesirable inputs $w \in \mathcal{F}(\mathcal{W})$ there exists a control signal $u \in \mathcal{F}(\mathcal{U})$ such that $u(t) = u\big(t, x(t-\tau), w(t-\tau)\big)$ and $x(T) \in \mathcal{G}$.

Only starting at $t + \tau$ can the controller try to counteract $Cw(t)$. However, at time $t + \tau$ the effect of $w(t)$ on the state $x(t+\tau)$ has become $e^{A\tau}Cw(t)$. Hence, set $\mathcal{Z}$ introduced in (8.3) does not describe the remaining control authority anymore.

Generalizing works [23], [33] we introduce the family of sets $\mathcal{Z}_t := B\mathcal{U} \ominus (-e^{At}C\mathcal{W})$ for all $t \geq 0$ with $e^{At}C\mathcal{W} := \{e^{At}Cw : w \in \mathcal{W}\}$. We will show that $\mathcal{Z}_\tau$ is the set of actual control inputs of system (8.5), when u has canceled any undesirable input w with a delay τ. Set $\mathcal{Z}_\tau$ is the time delayed extension of $\mathcal{Z}$ from (8.3).

Definition 21: The *minimal correction time* T_c represents the minimal time after which any undesirable input can be counteracted, $T_c := \inf\{t \geq \tau : \mathcal{Z}_t \neq \emptyset\}$.

If $T_c = +\infty$, then the impact on the state of some undesirable inputs cannot be canceled by any control input after the actuation delay, i.e., there exists some $w \in \mathcal{W}$ such that $-e^{At}Cw \notin B\mathcal{U}$ for all $t \geq \tau$. In this case, resilient reachability is impossible according to Lemma 22. Let us now assume that T_c is finite in order to build the theory for resilience in the presence of actuation delay.

8.3.2 Resilient reachability despite actuation delay

We want to know whether a target set $\mathcal{G} \subseteq \mathbb{R}^n$ is resiliently reachable by system (8.5). Because of the actuation delay τ, the controller can only guarantee that $x(t)$ is in some neighborhood of $x(t-\tau)$, it cannot ensure an exact location. Then, set $\mathcal{G}$ needs a minimal radius $\rho > 0$ to be resiliently reachable. Inspired by Hájek's approach [23], we introduce system (8.6) as a counterpart to system (8.5), just like system (8.3) was the counterpart of system (8.2) for the spacecraft without actuation delay.

$$\dot{x}(t) = Ax(t) + z(t), \quad x(0) = e^{AT_c}x_0, \quad z(t) \in \mathcal{Z}_{T_c} := B\mathcal{U} \ominus (-e^{AT_c}C\mathcal{W}). \tag{8.6}$$

Note that the input z of system (8.6) does not suffer from actuation delay by design of $\mathcal{Z}_{T_c}$.

Theorem 24: [Resilient reachability with actuation delay] If there exists $x_g \in \mathcal{G}$ such that $\mathbb{B}(x_g, \rho) \subseteq \mathcal{G}$ and x_g is reachable in a finite time T by system (8.6), then $\mathcal{G}$ is resiliently reachable by system (8.5) in time $T + T_c$, with $\rho := \frac{c}{\mu(A)}\big(e^{\mu(A)T_c} - 1\big)$ and $c := \max\{\|Cw\| : w \in \mathcal{W}\}$.

Proof. We first introduce the log-norm of matrix A defined in [105] as $\mu(A) := \max\{\lambda((A + A^\top)/2)\}$. Then, $\|e^{At}\| \leq e^{\mu(A)t}$ for all $t \geq 0$ according to. Since T is the time at which system (8.6) can reach x_g from $e^{AT_c}x_0$, there exists $z(s) \in \mathcal{Z}_{T_c}$ for all $s \in [0, T]$ such that

$$x_g = x(T) = e^{AT}\left(e^{AT_c}x_0 + \int_0^T e^{-As}z(s)\,ds\right), \quad \text{i.e.,} \quad e^{AT}\int_0^T e^{-As}z(s)\,ds = x_g - e^{A(T+T_c)}x_0.$$

Let $w \in \mathcal{F}(\mathcal{W})$ be some undesirable input affecting system (8.5). We now define the corresponding control input $u \in \mathcal{F}(\mathcal{U})$ so that it satisfies: $Bu(t) = 0$ for $t \in [0, T_c]$ and $Bu(t) = z(t - T_c) - e^{AT_c}Cw(t - T_c)$ for

$t \in [T_c, T + T_c]$. Note that $u(t) \in \mathcal{U}$ by definition of $z(t) \in \mathcal{Z}_{T_c}$. We apply this control law to system (8.5):

$$\begin{aligned}
x(T+T_c) &= e^{A(T+T_c)}\left(x_0 + \int_0^{T+T_c} e^{-At}Cw(t)\,dt + \int_0^{T+T_c} e^{-At}Bu(t)dt\right)\\
&= e^{A(T+T_c)}\left(x_0 + \int_0^{T+T_c} e^{-At}Cw(t)\,dt + \int_{T_c}^{T+T_c} e^{-At}\big(z(t-T_c) - e^{AT_c}Cw(t-T_c)\big)dt\right)\\
&= e^{A(T+T_c)}\left(x_0 + \int_0^{T+T_c} e^{-At}Cw(t)\,dt + \int_0^{T} e^{-A(s+T_c)}\big(z(s) - e^{AT_c}Cw(s)\big)ds\right)\\
&= e^{A(T+T_c)}\left(x_0 + \int_T^{T+T_c} e^{-At}Cw(t)dt\right) + e^{AT}\int_0^T e^{-As}z(s)ds\\
&= e^{A(T+T_c)}x_0 + \int_0^{T_c} e^{As}Cw(T+T_c-s)ds + \big(x_g - e^{A(T+T_c)}x_0\big) = x_g + \int_0^{T_c} e^{As}Cw(T+T_c-s)ds,
\end{aligned}$$

thanks to the definition of $z(s) \in \mathcal{Z}_{T_c}$. Then, by subtracting x_g and using the triangle inequality we obtain

$$\big\|x(T+T_c) - x_g\big\| \leq \int_0^{T_c} \big\|e^{As}\big\|\,\big\|Cw(T+T_c-s)\big\|ds \leq \int_0^{T_c} e^{\mu(A)s}c\,ds = \frac{c}{\mu(A)}\big(e^{\mu(A)T_c} - 1\big) = \rho.$$

Since $\mathbb{B}(x_g, \rho) \subseteq \mathcal{G}$, we have $x(T+T_c) \in \mathcal{G}$. Hence, $\mathcal{G}$ is resiliently reachable by system (8.5) in time $T+T_c$. □

Note that the control $z(t - T_c)$ responsible for steering to x_g in Theorem 24 is in fact an *open loop* control. A *feedback control* may perform better in practice, but the saturation enforcing that such signal remains bounded in $\mathcal{Z}_{T_c}$ would lead to a substantial increase in complexity of the proof.

Theorem 24 provides a sufficient resilient reachability condition for delayed system (8.5) in terms of the reachability of $\mathcal{G}$ by system (8.6). In turn, a sufficient condition for this last property can be verified with the lemma below.

Lemma 24: Let $Z \in \mathbb{R}^{n\times d}$ be a matrix whose columns are d linearly independent vectors of $\mathcal{Z}_{T_c}$ with $d := \dim(\mathcal{Z}_{T_c}) \geq 0$. If $0 \in \text{int}(\mathcal{Z}_{T_c})$, then system (8.6) is controllable if and only if $Re(\lambda(A)) = 0$, $\text{rank}(\mathcal{C}(A, Z)) = n$, and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}_{T_c}$.

Proof. By construction, $\text{Im}(Z) = \text{span}(\mathcal{Z}_{T_c})$ so that $\mathcal{C}(A, Z)$ is a controllability matrix associated with system (8.6). Since $\mathcal{U}$ and $\mathcal{W}$ are convex, so are $B\mathcal{U}$ and $e^{AT_c}C\mathcal{W}$. Their Minkowski difference $\mathcal{Z}_{T_c}$ is then also convex [83]. Then, $0 \in \text{int}(\text{co}(\mathcal{Z}_{T_c}))$ and trivially $0 \in \text{Ker}(I_n) \cap \mathcal{Z}_{T_c}$. These two inclusions allow us to apply Corollary 3.7 of [91] which yields the controllability condition. □

Thus, combining Lemma 24 and Theorem 24 provides a sufficient condition for resilient reachability in the presence of actuation delay. We will now investigate the more complicated problem of resilient trajectory tracking despite actuation delay.

8.3.3 Resilient trajectory tracking despite actuation delay

We want system (8.5) to track the actuated reference trajectory $\mathcal{T}_{\text{ref}}$ designed for system (8.6) by $\mathcal{T}_{\text{ref}} := \big\{x_{\text{ref}}(t) : \dot{x}_{\text{ref}}(t) = Ax_{\text{ref}}(t) + z_{\text{ref}}(t),\ \text{for all } t \geq 0\big\}$ with $z_{\text{ref}} \in \mathcal{F}(\mathcal{Z}_{\text{ref}})$. We also define $\mathcal{Z}_\varepsilon$ as a compact set of $\mathbb{R}^n$ satisfying $0 \in \text{relint}(\mathcal{Z}_\varepsilon)$ and $\dim(\mathcal{Z}_\varepsilon) = n$. As in Section 8.2, we use input set $\mathcal{Z}_\varepsilon$ to counteract the error arising from $x_0 \neq x_{\text{ref}}(0)$ through the dynamics

$$\dot{y}(t) = Ay(t) + z_\varepsilon(t), \quad y(0) = e^{AT_c}\big(x_0 - x_{\text{ref}}(0)\big), \quad z_\varepsilon(t) \in \mathcal{Z}_\varepsilon. \tag{8.7}$$

We can then state our resilient trajectory tracking result.

Theorem 25: If $\mathcal{Z}_\varepsilon \oplus \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}_{T_c}$ and system (8.7) is stabilizable in a finite time t_f, then the reference trajectory $\mathcal{T}_{\text{ref}}$ can be tracked by system (8.5) with a precision ρ after time $t_f + T_c$, i.e., $\|x(T) - x_{\text{ref}}(T)\| \leq \rho$ for all $T \geq t_f + T_c$.

Proof. Since system (8.7) is stabilizable in a finite time t_f, there exists a signal $z_\varepsilon(t) \in \mathcal{Z}_\varepsilon$ for all $t \in [0, t_f]$ yielding $y(t_f) = 0$. Because $0 \in \mathcal{Z}_\varepsilon$, we can extend signal z_ε with $z_\varepsilon(t) = 0$ for all $t > t_f$. Let $w \in \mathcal{F}(\mathcal{W})$ be the undesirable input signal. Since $\mathcal{Z}_\varepsilon \oplus \mathcal{Z}_{\text{ref}} \subseteq \mathcal{Z}_{T_c} = B\mathcal{U} \ominus (-e^{AT_c}C\mathcal{W})$, there exists $u \in \mathcal{F}(\mathcal{U})$ such that

$$Bu(t) = z_{\text{ref}}(t) + z_\varepsilon(t - T_c) - e^{AT_c}Cw(t - T_c) \qquad \text{for all} \quad t \geq T_c.$$

Indeed, the reference trajectory is known ahead of time, so $u(t)$ has access to $z_{\text{ref}}(t)$. We define u to satisfy $Bu(t) = z_{\text{ref}}(t)$ for $t \in [0, T_c]$. We now implement this controller for $T \geq T_c$ in system (8.5)

$$\begin{aligned}
x(T) &= e^{AT}\left(x_0 + \int_0^{T_c} e^{-At}Bu(t)\,dt + \int_{T_c}^{T} e^{-At}Bu(t)\,dt + \int_0^T e^{-At}Cw(t)\,dt\right)\\
&= e^{AT}\left(x_0 + \int_0^{T_c} e^{-At}z_{\text{ref}}(t)\,dt + \int_{T_c}^{T} e^{-At}\big(z_{\text{ref}}(t) + z_\varepsilon(t - T_c) - e^{AT_c}Cw(t - T_c)\big)\,dt + \int_0^T e^{-At}Cw(t)\,dt\right)\\
&= e^{AT}\left(x_0 + \int_0^{T} e^{-At}z_{\text{ref}}(t)\,dt + \int_0^{T-T_c} e^{-As}\big(e^{-AT_c}z_\varepsilon(s) - Cw(s)\big)ds + \int_0^T e^{-At}Cw(t)\,dt\right)\\
&= e^{AT}\left(x_0 + e^{-AT}x_{\text{ref}}(T) - x_{\text{ref}}(0) + e^{-AT_c}\int_0^{T-T_c} e^{-As}z_\varepsilon(s)\,ds + \int_{T-T_c}^{T} e^{-At}Cw(t)\,dt\right),
\end{aligned}$$

by definition of z_{ref}. Then,

$$x(T) - x_{\text{ref}}(T) = e^{A(T-T_c)}\left(e^{AT_c}\big(x_0 - x_{\text{ref}}(0)\big) + \int_0^{T-T_c} e^{-As}z_\varepsilon(s)\,ds\right) + \int_0^{T_c} e^{As}Cw(T-s)\,ds. \tag{8.8}$$

Note that the last integral term is the same as in Theorem 24 and hence can be bounded similarly:

$$\left\|\int_0^{T_c} e^{As}Cw(T-s)\,ds\right\| \leq \int_0^{T_c}\|e^{As}\|\,\|Cw(T-s)\|\,ds \leq c\int_0^{T_c} e^{\mu(A)s}ds = \rho.$$

Since z_ε stabilizes system (8.7) in time t_f, we have $y(T) = 0$ for all $T \geq t_f$. In particular, for $T \geq t_f + T_c$ we obtain

$$y(T - T_c) = 0 = e^{A(T-T_c)}\left(y(0) + \int_0^{T-T_c} e^{-As}z_\varepsilon(s)\,ds\right).$$

Note that $y(T - T_c)$ is exactly the central term in (8.8), which finally yields $\big\|x(T) - x_{\text{ref}}(T)\big\| \leq \rho$. □

Without control signal z_ε, the tracking error would be $\|x(t) - x_{\text{ref}}(t)\| \leq \rho + \big\|e^{At}\big(x_0 - x_{\text{ref}}(0)\big)\big\|$, which can grow exponentially if $x_0 - x_{\text{ref}}(0)$ is collinear with a positive eigenvector of A. When $x_0 = x_{\text{ref}}(0)$, we do not need z_ε and the tracking can be performed with precision ρ from time T_c onward.

Theorem 25 provides a sufficient condition for resilient trajectory tracking by delayed system (8.5) in terms of the finite time stabilizability of system (8.7). In turn, this property can be verified with the lemma below.

Lemma 25: Let $Z_\varepsilon \in \mathbb{R}^{n \times d}$ be a matrix whose columns are d linearly independent vectors of $\mathcal{Z}_\varepsilon$, with $d = \dim(\mathcal{Z}_{T_c}) = \dim(\mathcal{Z}_\varepsilon)$. System (8.7) is stabilizable in a finite time if and only if $Re(\lambda(A)) \leq 0$, $\text{rank}(\mathcal{C}(A, Z_\varepsilon)) = n$ and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}_\varepsilon$.

Proof. By construction $\text{Im}(Z_\varepsilon) = \text{span}(\mathcal{Z}_\varepsilon)$, so that $\mathcal{C}(A, Z_\varepsilon)$ is a controllability matrix associated with system (8.7). We made the assumption that $0 \in \text{relint}(\mathcal{Z}_\varepsilon)$, so we can apply Proposition 20 which guarantees that system (8.7) is stabilizable in finite time. □

We have now established sufficient conditions for resilient reachability and resilient trajectory tracking for linear systems in the presence of actuation delay. Thus, resilience theory does not require anymore the simplifying assumption of instantaneous knowledge of the undesirable input by the controller. We will now investigate how to extend resilience theory to nonlinear dynamics.

8.4 Resilience of nonlinear systems

In this dissertation we established resilient reachability conditions for linear systems with energy bounded inputs in Chapter 3 and for linear systems with amplitude bounded inputs in Chapter 7. These two approaches relied respectively on the reachability condition of [24] and on the duality theorem of [23]. However, both of these results were established for linear systems and their proofs actually require linear dynamics. Resilience theory would then need novel proofs of these two results to extend to nonlinear dynamics.

When studying the proof of Hájek's duality theorem [23], we notice that the two implications constituting the equivalence result are very different. Indeed, one of these implications is straightforward, while the other one requires a complex use of linearity. Then, we can extend one of the implications of Hájek's duality theorem [23] to nonlinear dynamics.

Consider the two following systems of state $x \in \mathbb{R}^n$:

$$\dot{x}(t) = f\big(t, x(t)\big) + g\big(t, x(t)\big)\big[Bu(t) + Cw(t)\big], \qquad x(0) = x_0 \in \mathbb{R}^n, \qquad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W}, \tag{8.9}$$

$$\dot{x}(t) = f\big(t, x(t)\big) + g\big(t, x(t)\big)z(t), \qquad x(0) = x_0 \in \mathbb{R}^n \qquad z(t) \in \mathcal{Z} := B\mathcal{U} \ominus (-C\mathcal{W}), \tag{8.10}$$

with f and g piecewise continuous in t and locally Lipschitz in x, constant matrices $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{n \times p}$, and compact sets $\mathcal{U} \subseteq \mathbb{R}^m$ and $\mathcal{W} \subseteq \mathbb{R}^p$.

Theorem 26: If system (8.10) is controllable, then system (8.9) is resilient.

Proof. Let $x_0 \in \mathbb{R}^n$, $x_{goal} \in \mathbb{R}^n$ and $w \in \mathcal{F}(\mathcal{W})$. Since system (8.10) is controllable, there exists $T \geq 0$ and $z \in \mathcal{F}(\mathcal{Z})$ driving the state of system (8.10) from x_0 to x_{goal} in time T. By definition of $\mathcal{Z}$, there exists $u \in \mathcal{F}(\mathcal{U})$ such that $Bu(t) = z(t) - Cw(t)$ for all $t \in [0, T]$. Then, applying input signals u and w to system (8.9) drives its state from x_0 to x_{goal} in time T. Since we found such a u for any w, system (8.9) is resilient. □

Note that the reverse implication of Theorem 26 holds for linear systems [23] but remains an open question for nonlinear dynamics. Note that Theorem 26 might not be extremely useful in practice since verifying the controllability of nonlinear system (8.10) is a difficult problem [111]. However, we can deduce immediately two corollaries of Theorem 26 with the exact same proof that might be much more practical.

Corollary 7: If system (8.10) is stabilizable, then system (8.9) is resiliently stabilizable.

Corollary 8: If target $x_g \in \mathbb{R}^n$ is reachable in time T by system (8.10), then x_g is resiliently reachable in time T by system (8.9).

We then have obtained several sufficient conditions for nonlinear resilient reachability.

8.5 Summary

In this chapter, we first extended resilience theory to cover trajectory tracking objectives. Then, we removed the main simplifying assumption of this theory by taking into account the realistically unavoidable delays affecting the knowledge of the undesirable input by the controller. Finally, we introduced a partial extension of resilience theory to nonlinear systems.

Chapter 9

Resilience of Linear Networks

9.1 Introduction

Resilience to catastrophic events is a crucial infrastructural challenge, recognized across populations and government levels [126], [127]. Natural disasters, terrorist acts, and widespread power failures all have the potential to rapidly deteriorate the infrastructure's capabilities to meet population needs. Inability to adapt to such events in real time also impedes emergency services, evacuation, and supply chain operations. The need for resilience — already amply displayed during natural disasters in past decades [128] — is made stronger with growing reliance on cyber-physical systems for control of infrastructure. For instance, a team of security researchers recently demonstrated the capability to influence traffic signals over the internet in at least ten cities [129], potentially causing system-wide disorder in the cities' traffic flows. Motivated by these examples of networks failures and encouraged by Problem 4, we will extend resilience theory to the study of linear networks based on our work [40].

Following a catastrophic natural event or an adversarial attack, the network under study endures a partial *loss of control authority* [53] over the actuators of one of its subsystems. This malfunction consists in some of the actuators of the malfunctioning subsystem to produce uncontrolled and thus possibly undesirable inputs with their full capabilities. We already have the theory to verify whether this isolated subsystem is resilient to its loss of control authority. However, the network framework raises different and more interesting questions, such as: is the subsystem resilient despite/thanks to its connection with the rest of the network? and how is the rest of the network affected by this malfunctioning subsystem? In this chapter, we will try to answer these questions.

The contributions of this chapter are threefold. First, we establish stabilizability and controllability conditions for networks of linear systems. Second, we build resilience conditions for networks with a subsystem suffering a partial loss of control authority over its actuators. Third, we quantify how unresilient can a subsystem be so as not to destabilize the whole network after a loss of control authority in said subsystem.

The remainder of this chapter is organized as follows. Section 9.2 introduces the network dynamics and states our problems of interest. Section 9.3 builds on Chapter 7 to establish necessary and sufficient conditions for the resilient stabilizability of linear systems. Armed with these results, we are then able to study the resilient stabilizability of networks of N subsystems in Section 9.4.

9.2 Networks preliminaries

Inspired by [130], we consider a network of $N \geq 2$ linear subsystems of dynamics

$$\dot{x}_1(t) = A_1 x_1(t) + \bar{B}_1 \bar{u}_1(t) + \sum_{k \in \mathcal{N}_1} L_{1,k} y_k(t), \qquad y_1(t) = F_1 x_1(t), \qquad x_1(0) = x_1^0 \in \mathbb{R}^{n_1}, \tag{9.1-1}$$

$$\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots$$

$$\dot{x}_N(t) = A_N x_N(t) + \bar{B}_N \bar{u}_N(t) + \sum_{k \in \mathcal{N}_N} L_{N,k} y_k(t), \qquad y_N(t) = F_N x_N(t), \qquad x_N(0) = x_N^0 \in \mathbb{R}^{n_N}, \tag{9.1-N}$$

where $x_i \in \mathbb{R}^{n_i}$, $\bar{u}_i \in \mathbb{R}^{m_i}$ and $y_i \in \mathbb{R}^{q_i}$ are respectively the state, the control input and the output of subsystem $i \in [\![1, N]\!]$. Additionally, $\mathcal{N}_i \subseteq [\![1, N]\!]$ is the set of neighbors of subsystem i with $i \notin \mathcal{N}_i$, while $A_i \in \mathbb{R}^{n_i \times n_i}$, $\bar{B}_i \in \mathbb{R}^{n_i \times m_i}$, $L_{i,k} \in \mathbb{R}^{n_i \times q_i}$ and $F_i \in \mathbb{R}^{q_i \times n_i}$ are constant matrices. The set of admissible control inputs for subsystem i is the unit hypercube of $\mathbb{R}^{m_i}$, i.e., $\bar{u}_i(t) \in \bar{\mathcal{U}}_i := [-1, 1]^{m_i}$. To alleviate the notations, we introduce matrices $D_{i,k} := L_{i,k} F_k$ to represent the connection of subsystems i and k for $i \in [\![1, N]\!]$ and $k \in \mathcal{N}_i$. Then,

$$\dot{x}_1(t) = A_1 x_1(t) + \bar{B}_1 \bar{u}_1(t) + \sum_{k \in \mathcal{N}_1} D_{1,k} x_k(t), \qquad x_1(0) = x_1^0 \in \mathbb{R}^{n_1}, \tag{9.2-1}$$

$$\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots$$

$$\dot{x}_N(t) = A_N x_N(t) + \bar{B}_N \bar{u}_N(t) + \sum_{k \in \mathcal{N}_N} D_{N,k} x_k(t), \qquad x_N(0) = x_N^0 \in \mathbb{R}^{n_N}. \tag{9.2-N}$$

We are interested by the properties of *stabilizability* and *controllability* of network (9.2) and how they derive from the stabilizability and controllability of each of its subsystems.

Definition 22: Tuple $(A_i, \bar{B}_i, \bar{\mathcal{U}}_i)$ is *stabilizable* (resp. *controllable*) if there exists an admissible control signal $\bar{u}_i \in \mathcal{F}(\bar{\mathcal{U}}_i)$ driving the state of system $\dot{x}_i(t) = A_i x_i(t) + \bar{B}_i \bar{u}_i(t)$ from any $x_i(0) \in \mathbb{R}^{n_i}$ to $0 \in \mathbb{R}^{n_i}$ (resp. to any $x_{goal} \in \mathbb{R}^{n_i}$).

Following Definition 22, the stabilizability and controllability of tuple $(A_i, \bar{B}_i, \bar{\mathcal{U}}_i)$ characterize subsystem i isolated from its neighbors. These are local properties from which we will want to derive the associated overall network properties. To do so, we define the network state $X := \big(x_1, x_2, \ldots, x_N\big) \in \mathbb{R}^{n_\Sigma}$ with $n_\Sigma := n_1 + \ldots + n_N$, and the network control input $\bar{u}(t) := \big(\bar{u}_1(t), \ldots, \bar{u}_N(t)\big) \in \bar{\mathcal{U}} := \bar{\mathcal{U}}_1 \times \ldots \times \bar{\mathcal{U}}_N \subseteq \mathbb{R}^{m_\Sigma}$ with $m_\Sigma := m_1 + \ldots + m_N$. Network dynamics (9.2) can then be written more concisely as

$$\dot{X}(t) = (A + D)X(t) + \bar{B}\bar{u}(t), \qquad X(0) = X_0 = \big(x_1^0, \ldots, x_N^0\big) \in \mathbb{R}^{n_\Sigma}, \tag{9.3}$$

with the constant matrices $A := \mathrm{diag}(A_1, \ldots, A_N)$, $B := \mathrm{diag}(\bar{B}_1, \ldots, \bar{B}_N)$ and $D := \big(D_{i,j}\big)_{(i,j) \in [\![1,N]\!]}$ with $D_{i,k} = 0$ if $k \notin \mathcal{N}_i$ and $D_{i,i} = 0$ for all $i \in [\![1, N]\!]$.

Definition 23: Network (9.3) is *stabilizable* (resp. *controllable*) if tuple $\big(A + D, \bar{B}, \bar{\mathcal{U}}\big)$ is stabilizable (resp. controllable).

We can then state our first problem of interest.

Problem 13: Assuming that each tuple $(A_i, \bar{B}_i, \bar{\mathcal{U}}_i)$ is stabilizable (resp. controllable), under what conditions is network (9.3) stabilizable (resp. controllable)?

Addressing Problem 13 allows us to investigate how overall network properties derive from those same properties at the subsystem level. This investigation will be crucial when studying how a malfunctioning subsystem affects the overall network.

Let us now describe said malfunction. Following a software bug or an adversarial attack, assume that subsystem (9.2-N) suffers a loss of control authority over a number $p_N \in [\![1, m_N]\!]$ of its m_N actuators. We then split the nominal input $\bar{u}_N$ between the remaining controlled inputs $u_N \in \mathcal{F}(\mathcal{U}_N)$, $\mathcal{U}_N = [-1,1]^{m_N - p_N}$ and the uncontrolled and possibly undesirable inputs $w_N \in \mathcal{F}(\mathcal{W}_N)$, $\mathcal{W}_N = [-1,1]^{p_N}$. We split accordingly matrix $\bar{B}_N$ into $B_N \in \mathbb{R}^{n_N \times (m_N - p_N)}$ and $C_N \in \mathbb{R}^{n_N \times p_N}$, so that the dynamics of subsystem (9.2-N) become

$$\dot{x}_N(t) = A_N x_N(t) + B_N u_N(t) + C_N w_N(t) + \sum_{k \,\in\, \mathcal{N}_N} D_{N,k} x_k(t), \qquad x_N(0) = x_N^0 \in \mathbb{R}^{n_N}. \tag{9.4}$$

We want to study how the loss of control authority over actuators of subsystem (9.2-N) affects the *stabilizability* and the *controllability* of the whole network. To adapt these properties to malfunctioning system (9.4), we first need the notion of *resilient reachability* introduced in [31].

Definition 24: A target $x_{goal} \in \mathbb{R}^{n_i}$ is *resiliently reachable* from $x_i(0) \in \mathbb{R}^{n_i}$ by malfunctioning system $\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + C_i w_i(t)$ if for all $w_i \in \mathcal{F}(\mathcal{W}_i)$, there exists $T \geq 0$ and $u_i \in \mathcal{F}(\mathcal{U}_i)$ such that $u_i(t)$ only depends on $w_i([0,t])$ and the solution exists, is unique and $x_i(T) = x_{goal}$.

Definition 25: Tuple $(A_i, B_i, C_i, \mathcal{U}_i, \mathcal{W}_i)$ is *resiliently stabilizable* (resp. *resilient*) if $0 \in \mathbb{R}^{n_i}$ (resp. every $x_{goal} \in \mathbb{R}^{n_i}$) is resiliently reachable from any $x_i(0) \in \mathbb{R}^{n_i}$ by malfunctioning system $\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + C_i w_i(t)$.

The network dynamics (9.3) are also impacted by the loss of control authority in subsystem (9.2-N). We define the network control input $u(t) := \big(\bar{u}_1(t), \ldots, \bar{u}_{N-1}(t), u_N(t)\big) \in \mathcal{U} := \bar{\mathcal{U}}_1 \times \ldots \times \bar{\mathcal{U}}_{N-1} \times \mathcal{U}_N \subseteq \mathbb{R}^{m_\Sigma - p_N}$. The network dynamics (9.3) then become

$$\dot{X}(t) = (A + D)X(t) + Bu(t) + C w_N(t), \qquad X(0) = X_0 = \big(x_1^0, \ldots, x_N^0\big) \in \mathbb{R}^{n_\Sigma}, \tag{9.5}$$

with the constant matrices $B = \text{diag}(\bar{B}_1, \ldots, \bar{B}_{N-1}, B_N)$ and $C = \begin{pmatrix} 0_{(n_\Sigma - n_N) \times p_N} \\ C_N \end{pmatrix}$.

Definition 26: Network (9.5) is *resiliently stabilizable* (resp. *resilient*) if tuple $\big(A + D, B, C, \mathcal{U}, \mathcal{W}_N\big)$ is resiliently stabilizable (resp. resilient).

We are now led to the following problems of interest.

Problem 14: Assuming that $(A_N, B_N, C_N, \mathcal{U}_N, \mathcal{W}_N)$ is resiliently stabilizable (resp. resilient) and $(A_i, \bar{B}_i, \bar{\mathcal{U}}_i)$ is stabilizable (resp. controllable) for $i \in [\![1, N-1]\!]$, under what conditions is network (9.5) resiliently stabilizable (resp. resilient)?

Then, we will consider the case where $(A_N, B_N, C_N, \mathcal{U}_N, \mathcal{W}_N)$ is not resiliently stabilizable or not resilient and study whether the other subsystems of the network are stabilizable or controllable despite the perturbations arising from the coupling between subsystems.

Definition 27: Subsystem $i \in [\![1, N-1]\!]$ of network (9.5) is *stabilizable* (resp. *controllable*) if for every $X_0 \in \mathbb{R}^{n_\Sigma}$ (resp. and every $x_{goal} \in \mathbb{R}^{n_i}$) and every $w_N \in \mathcal{F}(\mathcal{W}_N)$ there exists $T \geq 0$ and $u \in \mathcal{F}(\mathcal{U})$ such that the solution to (9.5) exists, is unique and $x_i(T) = 0$ (resp. $x_i(T) = x_{goal}$).

Note the difference between Definitions 22 and 27. The stabilizability of tuple $(A_i, \bar{B}_i, \mathcal{U}_i)$ only states that x_i can be driven to 0 with dynamics $\dot{x}_i(t) = A_i x_i(t) + \bar{B}_i \bar{u}_i(t)$. Whereas the stabilizability of subsystem i concerns dynamics (9.2-i) that possess the extra coupling term $\sum_{k \in \mathcal{N}_i} D_{i,k} x_k(t)$. This essential distinction leads to the following self-evident result.

Lemma 26: If subsystem $i \in [\![1, N-1]\!]$ is stabilizable (resp. resilient), then so is tuple $(A_i, \bar{B}_i, \mathcal{U}_i)$. Additionally, if all the $D_{i,k} = 0$ for $k \in \mathcal{N}_i$ or equivalently if $\mathcal{N}_i = \emptyset$, then the stabilizability (resp. resilience) of tuple $(A_i, \bar{B}_i, \mathcal{U}_i)$ implies that of subsystem i.

We can then state our third problem of interest.

Problem 15: Assuming that $(A_N, B_N, C_N, \mathcal{U}_N, \mathcal{W}_N)$ is *not* resiliently stabilizable (resp. *not* resilient) and $(A_i, \bar{B}_i, \bar{\mathcal{U}_i})$ is stabilizable (resp. controllable) for $i \in [\![1, N-1]\!]$, under what conditions are all the subsystems of network (9.5) stabilizable (resp. controllable)?

To solve our problems of interest, we first need several background results from Chapter 7 concerning the resilience and resilient stabilizability of isolated linear systems.

9.3 Background results

We consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + \bar{B}\bar{u}(t), \quad x(0) = x_0 \in \mathbb{R}^n, \quad \bar{u}(t) \in \bar{\mathcal{U}}, \tag{9.6}$$

with $A \in \mathbb{R}^{n \times n}$ and $\bar{B} \in \mathbb{R}^{n \times m}$ constant matrices and $\bar{\mathcal{U}} = [-1, 1]^m$. The controllability and stabilizability of system (9.6) can be assessed with Corollaries 3.6 and 3.7 of Brammer [91], restated here together as Theorem 27.

Theorem 27 (Brammer's conditions [91])**:** If $\bar{\mathcal{U}} \cap \ker(\bar{B}) \neq \emptyset$ and $\mathrm{int}(\mathrm{co}(\bar{\mathcal{U}})) \neq \emptyset$, then system (9.6) is stabilizable (resp. controllable) if and only if $\mathrm{rank}\big(\mathcal{C}(A, \bar{B})\big) = n$, $Re\big(\lambda(A)\big) \leq 0$ (resp. $= 0$) and there is no real eigenvector v of $A^\top$ satisfying $v^\top \bar{B}\bar{u} \leq 0$ for all $\bar{u} \in \bar{\mathcal{U}}$.

When $0 \in \mathrm{int}(\bar{\mathcal{U}})$, Theorem 27 boils down to Sontag's stabilizability condition [131] as the eigenvector criteria can be removed.

Theorem 28 (Sontag's condition [131])**:** If $0 \in \mathrm{int}(\bar{\mathcal{U}})$, then system (9.6) is stabilizable (resp. controllable) if and only if $\mathrm{rank}\big(\mathcal{C}(A, \bar{B})\big) = n$ and $Re\big(\lambda(A)\big) \leq 0$ (resp. $= 0$).

After a loss of control authority over p of the m actuators of system (9.6), the input signal $\bar{u}$ is split between the remaining controlled inputs $u \in \mathcal{F}(\mathcal{U})$, $\mathcal{U} = [-1, 1]^{m-p}$ and the uncontrolled and possibly undesirable inputs $w \in \mathcal{F}(\mathcal{W})$, $\mathcal{W} = [-1, 1]^p$. Matrix $\bar{B}$ is accordingly split into two constant matrices $B \in \mathbb{R}^{n \times (m-p)}$ and $C \in \mathbb{R}^{n \times p}$ so that the system dynamics become

$$\dot{x}(t) = Ax(t) + Bu(t) + Cw(t), \quad x(0) = x_0 \in \mathbb{R}^n, \quad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W}. \tag{9.7}$$

Resilience conditions established in Chapter 7 use Hájek's approach [23] and hence require the following system associated to dynamics (9.7)

$$\dot{x}(t) = Ax(t) + z(t), \qquad x(0) = x_0 \in \mathbb{R}^n, \qquad z(t) \in \mathcal{Z},$$

where $\mathcal{Z} \subseteq \mathbb{R}^n$ is the Minkowski difference between the set of admissible control inputs $B\mathcal{U} := \big\{Bu : u \in \mathcal{U}\big\}$ and the opposite of the set of undesirable inputs $C\mathcal{W} := \big\{Cw : w \in \mathcal{W}\big\}$, i.e.,

$$\mathcal{Z} := \big[B\mathcal{U} \ominus (-C\mathcal{W})\big] \cap B\mathcal{U} := \big\{z \in B\mathcal{U} : z - Cw \in B\mathcal{U} \text{ for all } w \in \mathcal{W}\big\}.$$

Informally, $\mathcal{Z}$ represents the control available after counteracting any undesirable input. The first resilience condition established in Chapter 7 is as follows.

Proposition 27: If $\text{int}(\mathcal{Z}) \neq \emptyset$, then system (9.7) is resiliently stabilizable (resp. resilient) if and only if $Re\big(\lambda(A)\big) \leq 0$ (resp. $= 0$).

The main issue with Proposition 27 is the requirement that $\text{int}(\mathcal{Z}) \neq \emptyset$ in $\mathbb{R}^n$, i.e., $\mathcal{Z}$ must be of dimension n, which implies that matrices B and $\bar{B}$ must be full rank. To remove this restrictive requirement, Chapter 7 relied on a matrix $Z \in \mathbb{R}^{n \times r}$ with $r := \dim(\mathcal{Z})$ such that $\text{Im}(Z) = \text{span}(\mathcal{Z})$.

Theorem 29 (Necessary and sufficient condition [37])**:** System (9.7) is resiliently stabilizable (resp. resilient) if and only if $Re\big(\lambda(A)\big) \leq 0$ (resp. $= 0$), rank $\big(\mathcal{C}(A, Z)\big) = n$ and there is no real eigenvector v of $A^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$.

Corollary 9: If $\dim(\mathcal{Z}) = \text{rank}(B)$, then system (9.7) is resiliently stabilizable (resp. resilient) if and only if system (9.6) is stabilizable (resp. controllable).

Notice how the resilience conditions only differ from the resilient stabilizability ones by further restricting the eigenvalues of A. Because of the similarity between these two concepts, we will only focus on resilient stabilizability.

9.4 Network stabilizability

Before studying the impact of a partial loss of control authority, let us investigate the stabilizability of the initial network (9.3) to address Problem 13.

9.4.1 Stabilizability of the initial network

Since $0 \in \text{int}(\bar{\mathcal{U}}) = [-1, 1]^{m_\Sigma}$, a direct application of Theorem 28 yields the following result.

Proposition 28: Network (9.3) is stabilizable if and only if rank $\big(\mathcal{C}(A + D, \bar{B})\big) = n_\Sigma$ and $Re(\lambda(A + D)) \leq 0$.

To address Problem 13, we need to establish conditions that rely on the stabilizability of tuples $\big(A_i, \bar{B}_i, \bar{\mathcal{U}}_i\big)$ unlike Proposition 28. We will then establish several sufficient conditions for stabilizability by studying the rank and eigenvalue conditions of Proposition 28.

First, note that having rank $\big(\mathcal{C}(A_i, \bar{B}_i)\big) = n_i$ for all $i \in [\![1, N]\!]$ does not necessarily imply rank $\big(\mathcal{C}(A + D, \bar{B})\big) = n_\Sigma$, even for matrices D with a small norm compared to that of A. As expected, we need a condition on matrix D to ensure that the coupling between subsystems does not alter their stabilizability. We could use the Popov-Belevitch-Hautus (PBH) controllability test [132], i.e., whether rank $\big[A + D - sI, \bar{B}\big] = n_\Sigma$ for all $s \in \mathbb{C}$, which is equivalent to verifying whether $\bar{B}x \neq 0$ for all x eigenvectors of $A + D$. However, relating the eigenvectors of $A + D$ to those of A is very complicated, as detailed in Corollary 7.2.6. of [100].

Instead, we will prefer the *distance to uncontrollability* $\mu(A, \bar{B})$ defined in [132] as

$$\mu(A, \bar{B}) := \min\big\{\|\Delta A, \Delta\bar{B}\| : (A + \Delta A, \bar{B} + \Delta\bar{B}) \text{ is uncontrollable}\big\} = \min\big\{\sigma_n\big(A - sI, \bar{B}\big) : s \in \mathbb{C}\big\}.$$

Since $\bar{B}$ is not affected by the coupling D, we define $\mu_{\bar{B}}(A) := \min\big\{\|\Delta A\| : (A+\Delta A, \bar{B}) \text{ is uncontrollable}\big\} \geq \mu(A, \bar{B})$. We also introduce the *real stability radius* of A,

$$r_{\mathbb{R}}(A) := \inf\big\{\|D\| : D \in \mathbb{R}^{n\times n} \text{ and } A+D \text{ is unstable}\big\} \text{ [133]}.$$

To approximate $r_{\mathbb{R}}(A)$ numerous lower bounds are provided in [133].

Proposition 29: Sufficient stabilizability conditions for network (9.3) can be derived from Proposition 28 coupled with the following statements.

(a) If $\text{rank}(\bar{B}_i) = n_i$ for all $i \in [\![1, N]\!]$, then $\text{rank}\big(\mathcal{C}(A+D, \bar{B})\big) = n_\Sigma$ for all $D \in \mathbb{R}^{n\times n}$.

(b) If there exists a matrix $F \in \mathbb{R}^{m_\Sigma \times n_\Sigma}$ such that $D = \bar{B}F$ and pairs $(A_i, \bar{B}_i)$ are controllable, then $\text{rank}\big(\mathcal{C}(A+D, \bar{B})\big) = n_\Sigma$.

(c) If $\|D\| \leq \mu_{\bar{B}}(A)$, then $\text{rank}\big(\mathcal{C}(A+D, \bar{B})\big) = n_\Sigma$.

(d) If $\|D\| \leq r_{\mathbb{R}}(A)$, then $Re\big(\lambda(A+D)\big) \leq 0$.

Proof. (a) Assume that $\text{rank}(\bar{B}_i) = n_i$. Because $\bar{B} = diag(\bar{B}_1, \ldots, \bar{B}_N)$, we have $\text{rank}(\bar{B}) = n_\Sigma$, which yields $\text{rank}\big(\mathcal{C}(A+D, \bar{B})\big) = \text{rank}\big(\bar{B}, (A+D)\bar{B}, \ldots\big) = n_\Sigma$.

(b) If D can be written as state feedback, $D = \bar{B}F$, then $\text{rank}\big(\mathcal{C}(A+D, \bar{B})\big) = \text{rank}\big(\mathcal{C}(A, \bar{B})\big)$ [134]. Because A and $\bar{B}$ are block diagonal matrices,

$$\text{rank}\big(\mathcal{C}(A, \bar{B})\big) = \sum_{i=1}^{N} \text{rank}\big(\mathcal{C}(A_i, \bar{B}_i)\big) = \sum_{i=1}^{N} n_i = n_\Sigma$$

because each $(A_i, \bar{B}_i)$ is controllable.

(c) By definition of $\mu_{\bar{B}}(A)$, $\|D\| \leq \mu_{\bar{B}}(A)$ leads to the controllability of pair $(A+D, \bar{B})$, i.e., to $\text{rank}\big(\mathcal{C}(A+D, \bar{B})\big) = n_\Sigma$. Note that $\mu_{\bar{B}}(A) \geq \mu(A, \bar{B})$.

(d) By definition of the stability radius $\|D\| \leq r_{\mathbb{R}}(A)$ leads to the stability of $A+D$, i.e., $Re(\lambda(A+D)) \leq 0$.

Combining statements (a), (b) or (c) with statement (d) yield three different sufficient stabilizability conditions for network (9.3) thanks to Proposition 28. □

Note that having $\mu_{\bar{B}}(A) > 0$ and $r_{\mathbb{R}}(A) > 0$ implicitly requires the stabilizability of all the tuples $\big(A_i, \bar{B}_i, \bar{\mathcal{U}}_i\big)$. Indeed, $\mu_{\bar{B}}(A) > 0$ requires $(A, \bar{B})$ to be controllable, i.e., each tuple $(A_i, \bar{B}_i)$ must be controllable because of the diagonal structure of A and $\bar{B}$. Similarly, $r_{\mathbb{R}}(A) > 0$ requires $Re\big(\lambda(A)\big) \leq 0$, but $\lambda(A) = \lambda(A_1) \cup \ldots \cup \lambda(A_N)$ because $A = diag(A_1, \ldots, A_N)$, hence $Re\big(\lambda(A_i)\big) \leq 0$. To sum up, $\mu_{\bar{B}}(A) > 0$ and $r_{\mathbb{R}}(A) > 0$ require $\text{rank}(A_i, \bar{B}_i) = n_i$ and $Re\big(\lambda(A_i)\big) \leq 0$, which are exactly the stabilizability conditions of Sontag for tuple $\big(A_i, \bar{B}_i, \bar{\mathcal{U}}_i\big)$ stated in Theorem 28.

Proposition 29 provides several ways to solve Problem 13. We will now address Problem 14.

9.4.2 Resilient stabilizability of the network

Inspired by the work completed before Proposition 27, we define the input sets of the network (9.5)

$$B\mathcal{U} := \big\{Bu : u \in \mathcal{U}\big\}, \quad C\mathcal{W} := \big\{Cw_N : w_N \in \mathcal{W}_N\big\}, \quad \text{and} \quad \mathcal{Z} := B\mathcal{U} \ominus (-C\mathcal{W}) \subseteq \mathbb{R}^{n_\Sigma}.$$

Similarly, we introduce the input sets of each subsystems

$$\bar{B}_i\bar{\mathcal{U}}_i := \left\{\bar{B}_i\bar{u}_i : \bar{u}_i \in \bar{\mathcal{U}}_i\right\} \qquad \text{for} \quad i \in [\![1, N-1]\!],$$

$$B_N\mathcal{U}_N := \left\{B_N u_N : u_N \in \mathcal{U}_N\right\}, \quad C_N\mathcal{W}_N := \left\{C_N w_N : w_N \in \mathcal{W}_N\right\} \quad \text{and} \quad \mathcal{Z}_N := B_N\mathcal{U}_N \ominus (-C_N\mathcal{W}_N).$$

These sets are all linked together with the following result.

Lemma 27: $\mathcal{Z} = \bar{B}_1\bar{\mathcal{U}}_1 \times \ldots \times \bar{B}_{N-1}\bar{\mathcal{U}}_{N-1} \times \mathcal{Z}_N$.

Proof. Take $z = (z_1, \ldots, z_N) \in \mathcal{Z}$. We want to show that $z_i \in \bar{B}_i\bar{\mathcal{U}}_i$ for $i \in [\![1, N-1]\!]$ and that $z_N \in \mathcal{Z}_N$. Let $w_N \in \mathcal{W}_N$. Since $z \in \mathcal{Z}$, there exists $u = (\bar{u}_1, \ldots, \bar{u}_{N-1}, u_N) \in \mathcal{U} = \bar{\mathcal{U}}_1 \times \ldots \times \bar{\mathcal{U}}_{N-1} \times \mathcal{U}_N$ such that

$$z - Cw_N = Bu = \begin{pmatrix} \bar{B}_1\bar{u}_1 \\ \vdots \\ \bar{B}_{N-1}\bar{u}_{N-1} \\ B_N u_N \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_{N-1} \\ z_N - C_N w_N \end{pmatrix}$$

Then, $z_i \in \bar{B}_i\bar{\mathcal{U}}_i$ for $i \in [\![1, N-1]\!]$. Additionally, for all $w_N \in \mathcal{W}_N$ we have $z_N - C_N w_N \in B_N\mathcal{U}_N$, i.e., $z_N \in \mathcal{Z}_N$.

On the other hand, let $\bar{u}_i \in \bar{\mathcal{U}}_i$ for $i \in [\![1, N-1]\!]$, $z_N \in \mathcal{Z}_N$ and define $z = \big(\bar{B}_1\bar{u}_1, \ldots, \bar{B}_{N-1}\bar{u}_{N-1}, z_N\big)$. We want to show that $z \in \mathcal{Z}$, so we take some $w_N \in \mathcal{W}_N$. Since $z_N \in \mathcal{Z}_N$, there exists $u_N \in \mathcal{U}_N$ such that $z_N - C_N w_N = B_N u_N$. Then,

$$z - Cw_N = \begin{pmatrix} \bar{B}_1\bar{u}_1 \\ \vdots \\ \bar{B}_{N-1}\bar{u}_{N-1} \\ z_N \end{pmatrix} - \begin{pmatrix} 0 \\ \vdots \\ 0 \\ C_N \end{pmatrix} w_N = \begin{pmatrix} \bar{B}_1\bar{u}_1 \\ \vdots \\ \bar{B}_{N-1}\bar{u}_{N-1} \\ B_N u_N \end{pmatrix} \in B\mathcal{U}, \quad \text{so } z \in \mathcal{Z}.$$

□

Let us now address Problem 14 by considering the case where $(A_N, B_N, C_N, \mathcal{U}_N, \mathcal{W}_N)$ is resiliently stabilizable. Using Proposition 27 we derive a sufficient condition for resilient stabilizability.

Proposition 30: If $\text{rank}(\bar{B}_i) = n_i$ for all $i \in [\![1, N-1]\!]$, $\text{int}(\mathcal{Z}_N) \neq \emptyset$ and $\|D\| \leq r_{\mathbb{R}}(A)$, then network (9.3) is resiliently stabilizable.

Proof. Since $\text{rank}(\bar{B}_i) = n_i$, $\text{int}(\bar{B}_i\bar{\mathcal{U}}_i) \neq \emptyset$, so that according to Lemma 27 we have $\text{int}(\mathcal{Z}) \neq \emptyset$. By assumption, we have $\|D\| \leq r_{\mathbb{R}}(A)$, i.e., $Re(\lambda(A + D)) \leq 0$. Then, Proposition 27 states that network (9.3) is resiliently stabilizable. □

Proposition 30 provides a straightforward resilient stabilizability condition for the network in a case that is similar to Proposition 29 (a). As mentioned after Proposition 27, the condition $\text{int}(\mathcal{Z}_N) \neq \emptyset$ requires $\text{rank}(B_N) = \text{rank}(\bar{B}_N) = n_N$. Then, Proposition 30 requires all $\bar{B}_i$ to be full rank, which is very restrictive and not necessary for stabilizability. Instead, we want to use Theorem 29 to derive a less restrictive resilient stabilizability condition for the network. To use this theorem, we must first build a matrix $Z \in \mathbb{R}^{n_\Sigma \times r_\Sigma}$ with $r_\Sigma = \dim(\mathcal{Z})$ and satisfying $\text{Im}(Z) = \text{span}(\mathcal{Z})$. In practice, matrix Z is built by collating r_Σ linearly independent vectors from set $\mathcal{Z}$.

Proposition 31: If $\|D\| \leq \min\big\{r_{\mathbb{R}}(A), \mu_Z(A)\big\}$ and there is no real eigenvector v of $(A+D)^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$, then network (9.5) is resiliently stabilizable.

Proof. We apply Theorem 29 to network (9.5) and obtain that it is resiliently stabilizable if and only if $Re\big(\lambda(A+D)\big) \leq 0$, rank $\big(\mathcal{C}(A+D, Z)\big) = n_\Sigma$ and there is no real eigenvector v of $(A+D)^\top$ satisfying $v^\top z \leq 0$ for all $z \in \mathcal{Z}$. The eigenvalue and rank conditions are satisfied thanks to $\|D\| \leq \min\{r_{\mathbb{R}}(A), \mu_Z(A)\}$, while the eigenvector condition is verified by assumption. □

As before, the fact that $(A_i, \bar{B}_i, \bar{\mathcal{U}}_i)$ are stabilizable and that $(A_N, B_N, C_N, \mathcal{U}_N, \mathcal{W}_N)$ is resiliently stabilizable, are implied by the conditions of Proposition 31.

When $\mathcal{Z}$ is not of full dimension, the eigenvector condition of Proposition 31 is difficult to verify. Indeed, the space $\mathcal{Z}^\perp$ is non-trivial and thus might encompass a real eigenvector of $A+D$ even if none of the eigenvectors of A are part of $\mathcal{Z}^\perp$. Intuitively, when D is small, the eigenvectors of $A+D$ should be 'close' to those of A. This intuition is formalized in Corollary 7.2.6 of [100], but the complexity of its statement prevents the derivation of a simple condition to be verified by A and D. Then, Propositions 30 and 31 are our solutions to Problem 14

9.4.3 Loss of control authority affecting a non-resilient subsystem

We now study the eventuality where $(A_N, B_N, C_N, \mathcal{U}_N, \mathcal{W}_N)$ is not resiliently stabilizable. More specifically, we consider the case where $-C_N\mathcal{W}_N \nsubseteq B_N\mathcal{U}_N$, i.e., subsystem (9.4) lost an actuator to which it is not resilient. Then, there are some undesirable inputs w_N that no control input u_N can overcome. Such undesirable inputs w_N can prevent stabilizability of subsystem N as demonstrated in Lemma 6 of [37].

To evaluate the resilient stabilizability of the network, we need to study the worst-case scenario where w_N is the most destabilizing undesirable input for subsystem (9.4). We will focus on the case where A_N is Hurwitz, so that the state x_N cannot be forced too far from the origin by w_N. Then, the terms $D_{i,N}x_N$ impacting subsystems (9.2-i) are bounded and might be counteracted if the controls $\bar{B}_i\bar{u}_i$ are strong enough for all $i \in [\![1, N-1]\!]$. On the other hand, if A_N is not Hurwitz, since $-C_N\mathcal{W}_N \nsubseteq B_N\mathcal{U}_N$, the state x_N can be driven to infinity by some w_N.

Compared to Theorem 29, we need stronger condition on A_N (its Hurwitzness) in order to employ Lyapunov theory. We will quantify the maximal degree of non-resilience of subsystem (9.4) despite which all other subsystems (9.2-i) remain stabilizable. Since we will isolate subsystem (9.4) to study its effect on the other subsystems, we need to split matrix D accordingly:

$$D = \left(\begin{array}{cccc|c} 0 & D_{1,2} & \dots & D_{1,N-1} & D_{1,N} \\ & \ddots & & & \vdots \\ D_{N-1,1} & \dots & D_{N-1,N-2} & 0 & D_{N-1,N} \\ \hline D_{N,1} & \dots & \dots & D_{N,N-1} & 0 \end{array}\right) := \left(\begin{array}{c|c} \hat{D} & D_{-,N} \\ \hline D_{N,-} & 0 \end{array}\right). \tag{9.8}$$

Then, the last row of D without the last diagonal element is $D_{N,-}$, while the last columns of D without the last diagonal element is $D_{-,N}$. Additionally, let $X(t)$ be the combined state of the first $N-1$ subsystems, $X(t) := \big(x_1(t), \dots, x_{N-1}(t)\big)$. We will now calculate how far from the origin can w_N push the state of subsystem (9.4) despite the best u_N and the Hurwitzness of A_N.

Proposition 32: If A_N is Hurwitz and $-C_N\mathcal{W}_N \nsubseteq B_N\mathcal{U}_N$, then for all $t \geq 0$

$$\|x_N(t)\|_{P_N} \leq e^{\alpha_N t}\left(\|x_N(0)\|_{P_N} + \int_0^t e^{-\alpha_N \tau}\beta_N(\tau)\, d\tau\right), \tag{9.9}$$

for all $P_N \succ 0$ and $Q_N \succ 0$ such that $A_N^\top P_N + P_N A_N = -Q_N$ and with

$$\alpha_N := \frac{-\lambda_{min}^{Q_N}}{2\lambda_{max}^{P_N}},\ \beta_N(t) := z_{max}^{P_N} + \|D_{N,-}X(t)\|_{P_N},\ z_{max}^{P_N} := \max_{w_N\,\in\mathcal{W}_N}\Big\{\min_{u_N\,\in\mathcal{U}_N}\|C_N w_N + B_N u_N\|_{P_N}\Big\}.$$

Proof. Since A_N is Hurwitz, there exists $P_N \succ 0$ and $Q_N \succ 0$ such that $A_N^\top P_N + P_N A_N = -Q_N$ according to Lyapunov theory [121]. Let us consider any such pair (P_N, Q_N). Then, inspired by Example 15 of [121], we study the P_N-norm of x_N, i.e., $x_N^\top P_N x_N = \|x_N\|_{P_N}^2$ when state x_N is following dynamics (9.4)

$$\begin{aligned}\frac{d}{dt}\|x_N(t)\|_{P_N}^2 &= \dot{x}_N^\top P_N x_N + x_N^\top P_N \dot{x}_N \\ &= x_N^\top\big(A_N^\top P_N + P_N A_N\big)x_N + 2x_N^\top P_N(B_N u_N + C_N w_N) + 2x_N^\top P_N\sum_{i=1}^{N-1} D_{N,i}x_i.\end{aligned}$$

With the notation of (9.8), we have $\sum_{i=1}^{N-1} D_{N,i}x_i = D_{N,-}X$. Since $P_N \succ 0$, the Cauchy-Schwarz inequality [100] as stated in Lemma 28 yields

$$x_N^\top P_N D_{N,-}X \leq \|x_N\|_{P_N}\|D_{N,-}X\|_{P_N} \quad \text{and} \quad x_N^\top P_N(B_N u_N + C_N w_N) \leq \|x_N\|_{P_N}\|B_N u_N + C_N w_N\|_{P_N}.$$

To stabilize x_N, we take the control u_N minimizing $\|B_N u_N + C_N w_N\|_{P_N}$ when w_N is chosen to maximize this norm, which yields $\|B_N u_N + C_N w_N\|_{P_N} \leq z_{max}^{P_N}$ by definition. Then,

$$\frac{d}{dt}\|x_N(t)\|_{P_N}^2 \leq -x_N^\top Q_N x_N + 2\|x_N\|_{P_N}\left(z_{max}^{P_N} + \|D_{N,-}X\|_{P_N}\right) \leq -\frac{\lambda_{min}^{Q_N}}{\lambda_{max}^{P_N}}\|x_N\|_{P_N}^2 + 2\|x_N\|_{P_N}\beta_N.$$

Indeed, $Q_N \succ 0$ yields $-x_N^\top Q_N x_N \leq -\lambda_{min}^{Q_N} x_N^\top x_N$ [122] and $\|x_N\|_{P_N}^2 \leq \lambda_{max}^{P_N} x_N^\top x_N$ leads to $-x_N^\top x_N \leq \frac{-1}{\lambda_{max}^{P_N}}\|x_N\|_{P_N}^2$. Hence, we obtain

$$\frac{d}{dt}\|x_N(t)\|_{P_N}^2 \leq 2\alpha_N\|x_N(t)\|_{P_N}^2 + 2\beta_N(t)\|x_N(t)\|_{P_N}.$$

We define $y_N(t) := \|x_N(t)\|_{P_N}$, so that we have $\frac{d}{dt}y_N^2(t) = 2y_N(t)\dot{y}_N(t) \leq 2\alpha_N y_N(t)^2 + 2\beta_N(t)y_N(t)$. For $y_N(t) > 0$, we then have $\dot{y}_N(t) \leq \alpha_N y_N(t) + \beta_N(t)$. Define also the function $f_N\big(t, s(t)\big) := \alpha_N s(t) + \beta_N(t)$.

The solution of the differential equation $\dot{z}(t) = f_N\big(t, z(t)\big)$ with initial condition $z(0) = \|x_N(0)\|_{P_N}$ is $z(t) = e^{\alpha_N t}\Big(\|x_N(0)\|_{P_N} + \int_0^t e^{-\alpha_N\tau}\beta_N(\tau)\, d\tau\Big)$. Since $f_N(t,z)$ is Lipschitz in z and continuous in t, $\dot{y}_N(t) \leq f_N\big(t, y_N(t)\big)$ and $y_N(0) = z(0)$, the Comparison Lemma of [122] states that $y_N(t) \leq z(t)$ for all $t \geq 0$, hence (9.9) holds. □

The Hurwitzness of A_N allows to bound the steady-state value of the state x_N despite undesirable inputs that cannot be counteracted. We will now study the impact of x_N on the rest of the network, whose dynamics follow

$$\dot{X}(t) = \hat{A}X(t) + \hat{B}\hat{u}(t) + \hat{D}X(t) + D_{-,N}x_N(t), \tag{9.10}$$

with $\hat{A} := diag\big(A_1, \dots, A_{N-1}\big)$, $\hat{B} := diag\big(\bar{B}_1, \dots, \bar{B}_{N-1}\big)$, $\hat{\mathcal{U}} := \bar{\mathcal{U}}_1 \times \dots \times \bar{\mathcal{U}}_{N-1} = [-1,1]^{n_\Sigma - n_N}$, and $\hat{u} := \big(\bar{u}_1, \dots, \bar{u}_{N-1}\big)$. Recall that $X = \big(x_1, \dots, x_{N-1}\big)$ and $\hat{D}$ was defined in (9.8). Dynamics (9.10) are perturbed by the term $D_{-,N}x_N(t)$ bounded in Proposition 32. Recall that $D_{-,N}$ designates the last column of matrix D without its last element as defined in (9.8). We can then evaluate how term $D_{-,N}x_N(t)$ impacts $X(t)$ with the following result.

Proposition 33: If $\hat{A}+\hat{D}$ and A_N are Hurwitz, $\hat{B}$ is full rank, and $C_N\mathcal{W}_N \not\subseteq B_N\mathcal{U}_N$, then for any $\hat{P} \succ 0$ and $\hat{Q} \succ 0$ such that $(\hat{A}+\hat{D})^\top \hat{P} + \hat{P}(\hat{A}+\hat{D}) = -\hat{Q}$ we define the constants $b_{min}^{\hat{P}} := \min\limits_{\hat{u}\,\in\,\partial\hat{\mathcal{U}}} \big\{\|\hat{B}\hat{u}\|_{\hat{P}}\big\}$, $\alpha := \frac{-\lambda_{min}^{\hat{Q}}}{2\lambda_{max}^{\hat{P}}}$, $\gamma := \|D_{-,N}\|_{\hat{P}} \sqrt{\frac{\lambda_{max}^{\hat{P}}}{\lambda_{min}^{P_N}}}$ and $\gamma_N := \|D_{N,-}\|_{\hat{P}} \sqrt{\frac{\lambda_{max}^{P_N}}{\lambda_{min}^{\hat{P}}}}$. If $\alpha\alpha_N \neq \gamma\gamma_N$, then there exists $h_\pm \in \mathbb{R}$ and $r_\pm \in \mathbb{R}$ such that

$$\|X(t)\|_{\hat{P}} \leq \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{\alpha\alpha_N - \gamma\gamma_N} + h_+ e^{(r_+ + \alpha_N)t} + h_- e^{(r_- + \alpha_N)t} \qquad \text{as long as} \quad \|X(t)\|_{\hat{P}} > 0. \tag{9.11}$$

If $\alpha\alpha_N = \gamma\gamma_N$, there are constants $h_\pm \in \mathbb{R}$ such that

$$\|X(t)\|_{\hat{P}} \leq \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{-\alpha - \alpha_N} t + h_+ e^{(\alpha + \alpha_N)t} + h_- \qquad \text{as long as} \quad \|X(t)\|_{\hat{P}} > 0. \tag{9.12}$$

Proof. Since $\hat{A}+\hat{D}$ is Hurwitz, there exists $\hat{P} \succ 0$ and $\hat{Q} \succ 0$ such that $(\hat{A}+\hat{D})^\top \hat{P} + \hat{P}(\hat{A}+\hat{D}) = -\hat{Q}$ according to Lyapunov theory [121]. Following the same steps as in the proof of Proposition 32 with $X^\top \hat{P} X = \|X\|_{\hat{P}}^2$ and X following the dynamics (9.10), we first obtain

$$\frac{d}{dt}\|X(t)\|_{\hat{P}}^2 = X^\top \big((\hat{A}+\hat{D})^\top \hat{P} + \hat{P}(\hat{A}+\hat{D})\big)X + 2X^\top \hat{P}\hat{B}\hat{u} + 2X^\top \hat{P} D_{-,N} x_N.$$

We apply the control law $\hat{B}\hat{u}(t) = -\frac{X(t)}{\|X(t)\|_{\hat{P}}} b_{min}^{\hat{P}}$ when $X(t) \neq 0$, which yields

$$X(t)^\top \hat{P}\hat{B}\hat{u}(t) = \frac{-X(t)^\top \hat{P} X(t)}{\|X(t)\|_{\hat{P}}} b_{min}^{\hat{P}} = -\|X(t)\|_{\hat{P}} b_{min}^{\hat{P}}.$$

Note that $\hat{B}$ being full rank guarantees that the control is always admissible as shown in the proof of Proposition 3 of [33]. Since $\|\cdot\|_{\hat{P}}$ is a norm, it verifies the Cauchy-Schwarz inequality [100] $X^\top \hat{P} D_{-,N} x_N \leq \|X\|_{\hat{P}} \|D_{-,N} x_N\|_{\hat{P}}$. Then,

$$\frac{d}{dt}\|X(t)\|_{\hat{P}}^2 \leq -X^\top \hat{Q} X - 2\|X\|_{\hat{P}} b_{min}^{\hat{P}} + 2\|X\|_{\hat{P}} \|D_{-,N} x_N\|_{\hat{P}}.$$

Because $\hat{P} \succ 0$ and $\hat{Q} \succ 0$, we obtain $-X^\top \hat{Q} X \leq -\frac{\lambda_{min}^{\hat{Q}}}{\lambda_{max}^{\hat{P}}} \|X\|_{\hat{P}}^2$. The associated $\hat{P}$-norm for matrices yields $\|D_{-,N} x_N\|_{\hat{P}} \leq \|D_{-,N}\|_{\hat{P}} \|x_N\|_{\hat{P}}$. Using the positive definiteness of $\hat{P}$ and P_N leads to

$$\|x_N\|_{\hat{P}} = \sqrt{x_N^\top \hat{P} x_N} \leq \sqrt{\lambda_{max}^{\hat{P}}\, x_N^\top x_N} \leq \sqrt{\frac{\lambda_{max}^{\hat{P}}}{\lambda_{min}^{P_N}} x_N^\top P_N x_N} = \|x_N\|_{P_N} \sqrt{\frac{\lambda_{max}^{\hat{P}}}{\lambda_{min}^{P_N}}}. \tag{9.13}$$

Following Proposition 32, we combine (9.9) with the preceding inequalities which yield

$$\frac{d}{dt}\|X(t)\|^2_{\hat{P}} \leq -\frac{\lambda^{\hat{Q}}_{min}}{\lambda^{\hat{P}}_{max}}\|X(t)\|^2_{\hat{P}} + 2\|X(t)\|_{\hat{P}}\left(\|D_{-,N}\|_{\hat{P}}\sqrt{\frac{\lambda^{\hat{P}}_{max}}{\lambda^{P_N}_{min}}}e^{\alpha_N t}\left(\|x_N(0)\|_{P_N}+\int_0^t e^{-\alpha_N\tau}\beta_N(\tau)d\tau\right) - b^{\hat{P}}_{min}\right).$$

Define $y(t) := \|X(t)\|_{\hat{P}}$ and $y_N(t) := \|x_N(t)\|_{P_N}$. Using a similar process as in (9.13), we also obtain $\|D_{N,-}X\|_{P_N} \leq \sqrt{\frac{\lambda^{P_N}_{max}}{\lambda^{\hat{P}}_{min}}}\|D_{N,-}X\|_{\hat{P}}$, which can be used in β_N defined in Proposition 32

$$\beta_N(\tau) = z^{P_N}_{max} + \|D_{N,-}X(\tau)\|_{P_N} \leq z^{P_N}_{max} + \sqrt{\frac{\lambda^{P_N}_{max}}{\lambda^{\hat{P}}_{min}}}\|D_{N,-}X(\tau)\|_{\hat{P}} \leq z^{P_N}_{max} + \gamma_N y(\tau). \tag{9.14}$$

If $y(t) \leq 0$, then $X(t) = 0$, so that (9.11) and (9.12) both hold. Otherwise, for $y(t) > 0$ we notice that $\frac{d}{dt}\|X(t)\|^2_{\hat{P}} = 2y(t)\dot{y}(t)$, and we can divide both sides of the inequality on $\frac{d}{dt}\|X(t)\|^2_{\hat{P}}$ by $2y(t)$ to obtain

$$\dot{y}(t) \leq \alpha y(t) + \left(\gamma e^{\alpha_N t}\left(y_N(0) + \int_0^t e^{-\alpha_N\tau}\left(z^{P_N}_{max} + \gamma_N y(\tau)\right)d\tau\right) - b^{\hat{P}}_{min}\right).$$

We calculate the following integral

$$e^{\alpha_N t}\int_0^t e^{-\alpha_N\tau}\,d\tau = e^{\alpha_N t}\frac{e^{-\alpha_N t}-1}{-\alpha_N} = \frac{1-e^{\alpha_N t}}{-\alpha_N},$$

so that the differential inequality becomes

$$\begin{aligned}\dot{y}(t) &\leq \alpha y(t) - b^{\hat{P}}_{min} + \gamma y_N(0)e^{\alpha_N t} + \gamma z^{P_N}_{max}\frac{1-e^{\alpha_N t}}{-\alpha_N} + \gamma\gamma_N e^{\alpha_N t}\int_0^t e^{-\alpha_N\tau}y(\tau)\,d\tau,\\ &\leq \alpha y(t) - \left(\frac{\gamma}{\alpha_N}z^{P_N}_{max} + b^{\hat{P}}_{min}\right) + \gamma\left(y_N(0) + \frac{z^{P_N}_{max}}{\alpha_N}\right)e^{\alpha_N t} + \gamma\gamma_N e^{\alpha_N t}\int_0^t e^{-\alpha_N\tau}y(\tau)\,d\tau.\end{aligned}$$

Now multiply both sides by $e^{-\alpha_N t} > 0$ and define $v(t) = e^{-\alpha_N t}y(t)$. Then, $\dot{v}(t) = -\alpha_N v(t) + e^{-\alpha_N t}\dot{y}(t)$, which leads to

$$\dot{v}(t) + \alpha_N v(t) \leq \alpha v(t) - \left(\frac{\gamma}{\alpha_N}z^{P_N}_{max} + b^{\hat{P}}_{min}\right)e^{-\alpha_N t} + \gamma\left(y_N(0) + \frac{z^{P_N}_{max}}{\alpha_N}\right) + \gamma\gamma_N\int_0^t v(\tau)\,d\tau.$$

With the function $f\big(t, v(t)\big) := (\alpha - \alpha_N)v(t) - \left(\frac{\gamma}{\alpha_N}z^{P_N}_{max} + b^{\hat{P}}_{min}\right)e^{-\alpha_N t} + \gamma\left(y_N(0) + \frac{z^{P_N}_{max}}{\alpha_N}\right) + \gamma\gamma_N\int_0^t v(\tau)\,d\tau$, we have $\dot{v}(t) \leq f\big(t, v(t)\big)$. Now we search for a solution to the differential equation $\dot{s}(t) = f\big(t, s(t)\big)$. Differentiating this equation yields

$$\begin{aligned}\ddot{s}(t) &= (\alpha - \alpha_N)\dot{s}(t) + \alpha_N\left(\frac{\gamma}{\alpha_N}z^{P_N}_{max} + b^{\hat{P}}_{min}\right)e^{-\alpha_N t} + \gamma\gamma_N s(t), \quad \text{i.e.,}\\ \ddot{s}(t) &+ (\alpha_N - \alpha)\dot{s}(t) - \gamma\gamma_N s(t) - \left(\gamma z^{P_N}_{max} + \alpha_N b^{\hat{P}}_{min}\right)e^{-\alpha_N t} = 0.\end{aligned} \tag{9.15}$$

We first study the linear homogeneous differential equation associated with (9.15), i.e.,

$$\ddot{s}(t) + (\alpha_N - \alpha)\dot{s}(t) - \gamma\gamma_N s(t) = 0. \tag{9.16}$$

Solutions of (9.16) can be written as $s_h(t) = e^{rt}$ with $r \in \mathbb{C}$. Plugging s_h in (9.16) leads to the quadratic equation $r^2 + (\alpha_N - \alpha)r - \gamma\gamma_N = 0$ after diving by e^{rt}. The solutions of this quadratic equation are $r_\pm = \frac{1}{2}\big(\alpha - \alpha_N \pm \sqrt{(\alpha_N - \alpha)^2 + 4\gamma\gamma_N}\big)$. Notice that the discriminant is nonnegative, since $\gamma \geq 0$ and $\gamma_N \geq 0$, so both $r_\pm \in \mathbb{R}$. We also need a particular solution of the non-homogeneous equation (9.15). We take $p \in \mathbb{R}$ such that $s_p(t) = pe^{-\alpha_N t}$ and plug it in (9.15) to obtain

$$\big(p\alpha_N^2 - p\alpha_N(\alpha_N - \alpha) - \gamma\gamma_N p - \gamma z_{max}^{P_N} - \alpha_N b_{min}^{\hat{P}}\big)e^{-\alpha_N t} = 0,$$
$$\text{i.e.,} \quad p = \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{\alpha_N^2 - \alpha_N(\alpha_N - \alpha) - \gamma\gamma_N} = \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{\alpha\alpha_N - \gamma\gamma_N}.$$

Let us first treat the case where $\alpha\alpha_N \neq \gamma\gamma_N$, so that p is well-defined. In this case, the general solution of (9.15) is $s(t) = pe^{-\alpha_N t} + h_+ e^{r_+ t} + h_- e^{r_- t}$ with $h_\pm \in \mathbb{R}$ two constants to choose. Since we obtained our solution by solving $\ddot{s}(t) = \frac{\partial f}{\partial t}\big(t, s(t)\big)$ instead of $\dot{s}(t) = f\big(t, s(t)\big)$, we have an additional initial condition to satisfy: $\dot{s}(0) = f\big(0, s(0)\big)$.

Now we can apply the Comparison Lemma of [122] stating that if $\dot{s}(t) = f\big(t, s(t)\big)$, f is continuous in t and locally Lipschitz in s and $s(0) = v(0)$, then $\dot{v}(t) \leq f\big(t, v(t)\big)$ implies $v(t) \leq s(t)$ for all $t \geq 0$. Using $\|X(t)\|_{\hat{P}} = y(t) = e^{\alpha_N t}v(t) \leq e^{\alpha_N t}s(t)$, we finally obtain (9.11). To determine the value of the constants $h_\pm$, we use the initial conditions $s(0) = v(0) = y(0)$ and $\dot{s}(0) = f\big(0, s(0)\big)$, i.e.,

$$p + h_+ + h_- = \|X(0)\|_{\hat{P}} \quad \text{and} \quad -\alpha_N p + h_+ r_+ + h_- r_- = (\alpha_N - \alpha)\|X(0)\|_{\hat{P}} - b_{min}^{\hat{P}} + \gamma\|x_N(0)\|_{P_N}.$$

We can solve these equations as

$$h_\pm = \frac{(\alpha_N - \alpha - r_\mp)\|X(0)\|_{\hat{P}} + \gamma\|x_N(0)\|_{P_N} - b_{min}^{\hat{P}} + (r_\mp + \alpha_N)p}{\pm\sqrt{(\alpha_N - \alpha)^2 + 4\gamma\gamma_N}}.$$

In the case $\alpha\alpha_N = \gamma\gamma_N$, the discriminant of the quadratic equation arising from the homogeneous differential equation is $(\alpha_N - \alpha)^2 + 4\alpha\alpha_N = (\alpha + \alpha_N)^2$, which yields $r_+ = \alpha$ and $r_- = -\alpha_N$. Hence $e^{-\alpha_N t}$ is an homogeneous solution and cannot be a particular solution of the non-homogeneous differential equation (9.15). Instead, we try $s_p(t) = pte^{-\alpha_N t}$ as a particular solution. We calculate its derivatives $\dot{s}_p(t) = p(1 - \alpha_N t)e^{-\alpha_N t}$, $\ddot{s}_p(t) = p(-2\alpha_N + \alpha_N^2 t)e^{-\alpha_N t}$ and plug it in (9.15). After dividing by $e^{-\alpha_N t}$ we obtain

$$\begin{aligned} 0 &= p(-2\alpha_N + \alpha_N^2 t) + (\alpha_N - \alpha)p(1 - \alpha_N t) - \alpha\alpha_N pt - \gamma z_{max}^{P_N} - \alpha_N b_{min}^{\hat{P}} \\ &= p(-2\alpha_N + \alpha_N - \alpha) + pt\big(\alpha_N^2 - \alpha_N(\alpha_N - \alpha) - \alpha\alpha_N\big) - \gamma z_{max}^{P_N} - \alpha_N b_{min}^{\hat{P}}, \end{aligned}$$

i.e., $p = \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{-\alpha - \alpha_N}$. In this case p is well-defined since $\alpha < 0$ and $\alpha_N < 0$. The general solution is then $s(t) = pte^{-\alpha_N t} + h_+ e^{\alpha t} + h_- e^{-\alpha_N t}$ with $h_\pm \in \mathbb{R}$ two constants. Applying the Comparison Lemma of [122] as above, we obtain $\|X(t)\|_{\hat{P}} = y(t) = e^{\alpha_N t}v(t) \leq e^{\alpha_N t}s(t)$, which yields (9.12). The initial conditions $s(0) = y(0)$ and $\dot{s}(0) = f\big(0, s(0)\big)$ lead to

$$h_+ + h_- = \|X(0)\|_{\hat{P}} \quad \text{and} \quad p + h_+\alpha - h_-\alpha_N = (\alpha_N - \alpha)\|X(0)\|_{\hat{P}} - b_{min}^{\hat{P}} + \gamma\|x_N(0)\|_{P_N}.$$

We can solve these equations as

$$h_\pm = \frac{\frac{1}{2}\big(-\alpha_N - \alpha \pm 3(\alpha - \alpha_N)\big)\|X(0)\|_{\hat{P}} \mp \gamma \|x_N(0)\|_{P_N} \pm b_{min}^{\hat{P}} \pm p}{-\alpha_N - \alpha}.$$

□

We can now derive conditions for subsystem (9.10) to be stabilizable despite the perturbations created by x_N.

Theorem 30: If $\hat{A} + \hat{D}$ and A_N are Hurwitz, $\hat{B}$ is full rank, and $C_N \mathcal{W}_N \nsubseteq B_N \mathcal{U}_N$, $\gamma\gamma_N \leq \alpha\alpha_N$ and $\gamma z_{max}^{P_N} < (-\alpha_N) b_{min}^{\hat{P}}$, then subsystem (9.10) is stabilizable in finite time.

Proof. Let us first consider the case $\gamma\gamma_N = \alpha\alpha_N$. By definition $\alpha + \alpha_N < 0$, so the exponential term in (9.12) goes to zero asymptotically. By assumption $\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}} < 0$ and $-\alpha - \alpha_N > 0$, so the ratio of these factors is negative. Because this ratio is multiplied by t in (9.12), there exists some time $T \geq 0$ such that for all $t \geq T$

$$\frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{-\alpha - \alpha_N} t + h_- \leq 0. \tag{9.17}$$

Since the terms in (9.17) are linearly decreasing with time, while $h_+ e^{(\alpha+\alpha_N)t}$ is asymptotically converging to zero their sum is reaching zero in finite time. Since their sum constitutes, the right-hand side of (9.12), subsystem (9.10) is stabilizable in finite time. Note that after this upper bound reaches 0 it loses its validity since we assumed $\|X(t)\|_{\hat{P}} > 0$ in the derivation of this bound, just after (9.14). Therefore, the right-hand side of (9.12) becoming negative does not create an issue.

Now consider the case $\gamma\gamma_N < \alpha\alpha_N$. Note that this inequality is equivalent to $r_+ + \alpha_N < 0$ as shown below

$$\begin{aligned} r_+ + \alpha_N < 0 \iff \frac{1}{2}(\alpha + \alpha_N) + \frac{1}{2}\sqrt{(\alpha_N - \alpha)^2 + 4\gamma\gamma_N} < 0 \iff (\alpha_N - \alpha)^2 + 4\gamma\gamma_N < (\alpha + \alpha_N)^2 \\ \iff -2\alpha\alpha_N + 4\gamma\gamma_N < 2\alpha\alpha_N \iff \gamma\gamma_N < \alpha\alpha_N. \end{aligned}$$

Since $r_- \leq r_+$, we also have $r_- + \alpha_N < 0$, so both exponential terms in (9.11) converge to zero. Additionally, the fraction term in (9.11) is negative, so the right-hand side of (9.11) reaches zero in finite time. Therefore, subsystem (9.10) is stabilizable in finite time. As in the previous case, (9.11) is only valid while $\|X(t)\|_{\hat{P}} > 0$, hence its upper bound is allowed to become negative. □

Let us now give some intuition concerning Theorem 30. Since γ is proportional to the norm of the matrix $D_{-,N}$ which multiplies $x_N(t)$ in (9.10), γ quantifies the impact of the $x_N(t)$ on $X(t)$. To put it in words, γ quantifies the impact of nonresilient subsystem (9.4) on the rest of the network (9.10). Reciprocally, γ_N quantifies the impact of $X(t)$ on $x_N(t)$. On the other hand, $\alpha = \frac{-\lambda_{min}^{\hat{Q}}}{2\lambda_{max}^{\hat{P}}}$ relates to the Hurwitzness of the first $N-1$ subsystems of network (9.10), while α_N relates to the Hurwitzness of malfunctioning subsystem (9.4). Therefore, condition $\gamma\gamma_N \leq \alpha\alpha_N$ follows the intuition that the magnitude of the perturbations arising from the coupling between subsystems (9.10) and (9.4) must be weaker than the Hurwitzness and hence stabilizability of each of these subsystems.

Let us now give some intuition concerning condition $\gamma z_{max}^{P_N} < (-\alpha_N) b_{min}^{\hat{P}}$ of Theorem 30. Since $z_{max}^{P_N}$ describes the magnitude of the destabilizing inputs in subsystem (9.4), term $\gamma z_{max}^{P_N}$ quantifies the destabilizing influence of w_N on the state of the rest of the network X. On the other hand, $b_{min}^{\hat{P}}$ relates to the magnitude of the stabilizing inputs in subsystem (9.10) and α_N relates to the Hurwitzness of malfunctioning subsystem (9.4).

Therefore, condition $\gamma z_{max}^{P_N} < (-\alpha_N) b_{min}^{\hat{P}}$ carries the intuition that the stabilizing terms of the network must overcome the destabilizing ones.

Since we have bounded the state X of the first $N-1$ subsystems, we can use this knowledge to bound the state of the malfunctioning subsystem N. Indeed, the bound derived in Proposition 32 depends on $X(t)$ through the term $\beta_N(t)$.

Proposition 34: If $\hat{A} + \hat{D}$ and A_N are Hurwitz, $\hat{B}$ is full rank, and $C_N \mathcal{W}_N \not\subseteq B_N \mathcal{U}_N$, we can bound the state of subsystem (9.4). If $\alpha \alpha_N \neq \gamma \gamma_N$, then as long as $\|x_N(t)\|_{P_N} > 0$,

$$\|x_N(t)\|_{P_N} \leq \frac{\alpha z_{max}^{P_N} + \gamma_N b_{min}^{\hat{P}}}{\gamma \gamma_N - \alpha \alpha_N} \big(1 - e^{\alpha_N t}\big) + e^{\alpha_N t} \left(\|x_N(0)\|_{P_N} + \frac{\gamma_N h_+}{r_+} \big(e^{r_+ t} - 1\big) + \frac{\gamma_N h_-}{r_-} \big(e^{r_- t} - 1\big) \right). \tag{9.18}$$

If $\alpha \alpha_N = \gamma \gamma_N$, then as long as $\|x_N(t)\|_{P_N} > 0$, we have

$$\begin{aligned} \|x_N(t)\|_{P_N} \quad & \leq e^{\alpha_N t} \|x_N(0)\|_{P_N} + \tfrac{1 - e^{\alpha_N t}}{-\alpha_N} \Big(\gamma_N h_- + \tfrac{\gamma_N b_{min}^{\hat{P}} - \alpha_N z_{max}^{P_N}}{-\alpha - \alpha_N} \Big) \\ & + \tfrac{\alpha z_{max}^{P_N} + \gamma_N b_{min}^{\hat{P}}}{\alpha + \alpha_N} t + \tfrac{\gamma_N h_+}{\alpha} \big(e^{\alpha t} - 1\big) e^{\alpha_N t}. \end{aligned} \tag{9.19}$$

Proof. We recall from Proposition 32 that $\|x_N(t)\|_{P_N} \leq e^{\alpha_N t} \Big(\|x_N(0)\|_{P_N} + \int_0^t e^{-\alpha_N \tau} \beta_N(\tau)\, d\tau \Big)$ (9.9). Following (9.14), we have $\beta_N(t) \leq z_{max}^{P_N} + \gamma_N \|X(t)\|_{\hat{P}}$. We can bound $\|X(t)\|_{\hat{P}}$ with (9.11) or (9.12) from Proposition 33 depending on the values of $\alpha \alpha_N$ and $\gamma \gamma_N$.

We start with the generic case where $\alpha \alpha_N \neq \gamma \gamma_N$. Then, bound (9.11) yields

$$\begin{aligned} \int_0^t e^{-\alpha_N \tau} \beta_N(\tau)\, d\tau & \leq \int_0^t e^{-\alpha_N \tau} \big(z_{max}^{P_N} + \gamma_N p + \gamma_N h_+ e^{(r_+ + \alpha_N)\tau} + \gamma_N h_- e^{(r_- + \alpha_N)\tau} \big) d\tau \\ & = \int_0^t e^{-\alpha_N \tau} \big(z_{max}^{P_N} + \gamma_N p \big) d\tau + \gamma_N \int_0^t h_+ e^{r_+ \tau} + h_- e^{r_- \tau} d\tau \\ & = \frac{e^{-\alpha_N t} - 1}{-\alpha_N} \big(z_{max}^{P_N} + \gamma_N p \big) + \frac{\gamma_N h_+}{r_+} \big(e^{r_+ t} - 1 \big) + \frac{\gamma_N h_-}{r_-} \big(e^{r_- t} - 1 \big). \end{aligned}$$

We replace p in

$$\frac{z_{max}^{P_N} + \gamma_N p}{-\alpha_N} = \frac{\alpha z_{max}^{P_N} + \gamma_N b_{min}^{\hat{P}}}{\gamma \gamma_N - \alpha \alpha_N} \qquad \text{with} \qquad p = \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{\alpha \alpha_N - \gamma \gamma_N}.$$

Then, plugging the integral calculated above in (9.9), we obtain

$$\|x_N(t)\|_{P_N} \leq e^{\alpha_N t} \left(\|x_N(0)\|_{P_N} + \frac{\alpha z_{max}^{P_N} + \gamma_N b_{min}^{\hat{P}}}{\gamma \gamma_N - \alpha \alpha_N} \big(e^{-\alpha_N t} - 1 \big) + \frac{\gamma_N h_+}{r_+} \big(e^{r_+ t} - 1 \big) + \frac{\gamma_N h_-}{r_-} \big(e^{r_- t} - 1 \big) \right),$$

which yields (9.18).

We can now address the other case where $\alpha\alpha_N = \gamma\gamma_N$ and $\|X(t)\|_{\hat{P}}$ is bounded by (9.12), which yields

$$\begin{aligned}
\int_0^t e^{-\alpha_N\tau}\beta_N(\tau)\,d\tau &\leq \int_0^t e^{-\alpha_N\tau}\big(z_{max}^{P_N} + \gamma_N p\tau + \gamma_N h_+ e^{(\alpha+\alpha_N)\tau} + \gamma_N h_-\big)d\tau \\
&= \big(z_{max}^{P_N} + \gamma_N h_-\big)\int_0^t e^{-\alpha_N\tau}\,d\tau + \gamma_N p\int_0^t \tau e^{-\alpha_N\tau}\,d\tau + \gamma_N h_+\int_0^t e^{\alpha\tau}\,d\tau \\
&= \big(z_{max}^{P_N} + \gamma_N h_-\big)\frac{e^{-\alpha_N t}-1}{-\alpha_N} + \frac{\gamma_N p}{\alpha_N^2}\big(1 - e^{-\alpha_N t} - \alpha_N t e^{-\alpha_N t}\big) + \frac{\gamma_N h_+}{\alpha}\big(e^{\alpha t}-1\big) \\
&= \frac{e^{-\alpha_N t}-1}{-\alpha_N}\Big(z_{max}^{P_N} + \gamma_N h_- + \frac{\gamma_N p}{\alpha_N}\Big) - \frac{\gamma_N p}{\alpha_N}te^{-\alpha_N t} + \frac{\gamma_N h_+}{\alpha}\big(e^{\alpha t}-1\big).
\end{aligned}$$

Then, we replace p with its definition:

$$\frac{\gamma_N p}{-\alpha_N} = \frac{\gamma\gamma_N z_{max}^{P_N} + \gamma_N\alpha_N b_{min}^{\hat{P}}}{\alpha_N(\alpha+\alpha_N)} = \frac{\alpha z_{max}^{P_N} + \gamma_N b_{min}^{\hat{P}}}{\alpha+\alpha_N} \qquad \text{thanks to} \qquad p = \frac{\gamma z_{max}^{P_N} + \alpha_N b_{min}^{\hat{P}}}{-\alpha-\alpha_N}.$$

Multiplying the integral calculated previously by $e^{\alpha_N t}$ yields

$$\int_0^t e^{\alpha_N(t-\tau)}\beta_N(\tau)d\tau \leq \frac{1-e^{\alpha_N t}}{-\alpha_N}\Big(\gamma_N h_- + \frac{\gamma_N b_{min}^{\hat{P}} - \alpha_N z_{max}^{P_N}}{-\alpha-\alpha_N}\Big) + \frac{\alpha z_{max}^{P_N} + \gamma_N b_{min}^{\hat{P}}}{\alpha+\alpha_N}t + \frac{\gamma_N h_+}{\alpha}\big(e^{\alpha t}-1\big)e^{\alpha_N t}.$$

We finally obtain (9.19) thanks to (9.9). □

Propositions 33 and 34 describe the state of the network after a partial loss of control authority to which the network dynamics were not resilient. These two results relied on the full rank assumption of $\hat{B}$, the control matrix of the unaffected part of the network. Because this assumption might be too restrictive, we will now employ a different approach to bound the states of the network.

9.4.4 Underactuated network

Let us now assume that $\hat{B}$ is not full rank, preventing the use of Proposition 33. Instead of the stabilizing control input of constant magnitude $\hat{B}\hat{u}(t) = -\frac{X(t)}{\|X(t)\|_{\hat{P}}}b_{min}^{\hat{P}}$ used in Proposition 33, we will employ a linear control $\hat{u}(t) = -KX(t)$ to stabilize X.

Let K be a matrix such that $\hat{A} + \hat{D} - \hat{B}K$ is Hurwitz. Then, for any $\hat{P} \succ 0$ and $\hat{Q} \succ 0$ such that $(\hat{A} + \hat{D} - \hat{B}K)^\top\hat{P} + \hat{P}(\hat{A} + \hat{D} - \hat{B}K) = -\hat{Q}$, we can define the same constants as in Proposition 33, namely $\alpha = \frac{-\lambda_{min}^{\hat{Q}}}{2\lambda_{max}^{\hat{P}}}$, $\gamma = \|D_{-,N}\|_{\hat{P}}\sqrt{\frac{\lambda_{max}^{\hat{P}}}{\lambda_{min}^{P_N}}}$ and $\gamma_N = \|D_{N,-}\|_{\hat{P}}\sqrt{\frac{\lambda_{max}^{P_N}}{\lambda_{min}^{\hat{P}}}}$.

Proposition 35: If pair $\big(\hat{A} + \hat{D}, \hat{B}\big)$ is controllable, A_N is Hurwitz, $C_N\mathcal{W}_N \nsubseteq B_N\mathcal{U}_N$, $\gamma\gamma_N < \alpha\alpha_N$ and $m \leq \frac{\sqrt{\lambda_{min}^{\hat{P}}}}{\|K\|}$, then $\|X(t)\|_{\hat{P}} \leq m$ for all $t \geq 0$, with $m := \max\big\{p + h_+ + h_-,\ p + h_+,\ p + h_-,\ p\big\}$,

$$h_\pm = \frac{(\alpha_N - \alpha - r_\mp)\|X(0)\|_{\hat{P}} + \gamma\|x_N(0)\|_{P_N} + (r_\mp + \alpha_N)p}{\pm\sqrt{(\alpha_N - \alpha)^2 + 4\gamma\gamma_N}} \quad \text{and} \quad p = \frac{\gamma z_{max}^{P_N}}{\alpha\alpha_N - \gamma\gamma_N}.$$

Proof. Since pair $\big(\hat{A} + \hat{D}, \hat{B}\big)$ is controllable, there exists a matrix K such that $\hat{A} + \hat{D} - \hat{B}K$ is Hurwitz [122]. Then, there exists $\hat{P} \succ 0$ and $\hat{Q} \succ 0$ such that $(\hat{A} + \hat{D} - \hat{B}K)^\top\hat{P} + \hat{P}(\hat{A} + \hat{D} - \hat{B}K) = -\hat{Q}$ according to Lyapunov theory [121]. We will follow the same steps as in the proof of Proposition 33 with $X^\top\hat{P}X = \|X\|_{\hat{P}}^2$ and X following the dynamics (9.10) with $\hat{u}(t) = -KX(t)$. Once we obtain bounds on $X(t)$ we will verify under which conditions is $\hat{u}$ admissible. We first obtain

$$\frac{d}{dt}\|X(t)\|_{\hat{P}}^2 = X(t)^\top \big((\hat{A}+\hat{D}-\hat{B}K)^\top \hat{P} + \hat{P}(\hat{A}+\hat{D}-\hat{B}K)\big)X(t) + 2X(t)^\top \hat{P} D_{-,N} x_N(t).$$

We then proceed as in Proposition 33, but without the term $b_{min}^{\hat{P}}$. If $\alpha\alpha_N \neq \gamma\gamma_N$, then $r_\pm = \frac{1}{2}\big(\alpha - \alpha_N \pm \sqrt{(\alpha_N-\alpha)^2 + 4\gamma\gamma_N}\big)$, $p = \frac{\gamma z_{max}^{P_N}}{\alpha\alpha_N - \gamma\gamma_N} > 0$ and there are constants $h_\pm \in \mathbb{R}$ such that

$$\|X(t)\|_{\hat{P}} \leq p + h_+ e^{(r_+ + \alpha_N)t} + h_- e^{(r_- + \alpha_N)t} \qquad \text{for all } t \geq 0, \tag{9.20}$$

$$p + h_+ + h_- = \|X(0)\|_{\hat{P}} \quad \text{and} \quad -\alpha_N p + h_+ r_+ + h_- r_- = (\alpha_N - \alpha)\|X(0)\|_{\hat{P}} + \gamma\|x_N(0)\|_{P_N}.$$

Similarly, in the case $\alpha\alpha_N = \gamma\gamma_N$, there are constants $h_\pm \in \mathbb{R}$ such that

$$\|X(t)\|_{\hat{P}} \leq \frac{\gamma z_{max}^{P_N}}{-\alpha - \alpha_N} t + h_+ e^{(\alpha+\alpha_N)t} + h_- \qquad \text{for all} \quad t \geq 0. \tag{9.21}$$

Bounds (9.20) and (9.21) are only valid when $\hat{u}(t) = -KX(t) \in \mathcal{U} = [-1,1]^m$. For this control law to be admissible, we then need

$$\|X(t)\|_{\hat{P}} \leq \frac{\sqrt{\lambda_{min}^{\hat{P}}}}{\|K\|} \quad \text{at all times } t \geq 0, \text{ since} \quad \|\hat{u}(t)\| \leq \|K\|\|X(t)\| \leq \|K\| \frac{\|X(t)\|_{\hat{P}}}{\sqrt{\lambda_{min}^{\hat{P}}}}.$$

If $\alpha\alpha_N = \gamma\gamma_N$, the term linear in t of (9.21) grows unbounded since $\gamma z_{max}^{P_N} > 0$ and $-\alpha - \alpha_N > 0$. In this case, we cannot guarantee to keep $X(t)$ bounded with linear control $\hat{u}(t) = -KX(t)$.

If $\gamma\gamma_N < \alpha\alpha_N$, then according to the equivalence derived in the proof of Theorem 30, both $r_\pm + \alpha_N < 0$. Then, both exponential terms of (9.20) converge to 0, leaving only the constant term $p > 0$. In this case,

$$\|X(t)\|_{\hat{P}} \leq \max\big\{p + h_+ + h_-,\ p + h_+,\ p + h_-,\ p\big\} = m.$$

Thus, $m \leq \frac{\sqrt{\lambda_{min}^{\hat{P}}}}{\|K\|}$ guarantees the admissibility of $\hat{u}$. □

Note that the perturbation from subsystem N in norm bounds (9.20) and (9.21) is modeled by term $z_{max}^{P_N} > 0$ of constant magnitude. Hence, this perturbation cannot be overcome when X is near 0 by the linear control $\hat{u}(t) = -KX(t)$ used in Proposition 35. That is why Proposition 35 only guarantees the boundedness of X and not its stabilizability.

Maximum m depends on the sign of $h_\pm$, which depend on the initial conditions $\|X(0)\|_{\hat{P}}$ and $\|x_N(0)\|_{P_N}$.

Remark: If the network is initially at rest when the loss of control authority occurs, i.e., if $X(0) = 0$ and $x_N(0) = 0$, then $h_+ < 0$ and $h_- > 0$, so that $m = p + h_- = -h_+ = \frac{-(r_- + \alpha_N)\gamma z_{max}^{P_N}}{(\alpha\alpha_N - \gamma\gamma_N)\sqrt{(\alpha_N-\alpha)^2 + 4\gamma\gamma_N}}$.

9.5 Supporting lemmata

Since $(x,y) \mapsto x^\top P y$ defines a scalar product for any $P \succ 0$, it verifies the Cauchy-Schwarz inequality [100]. We provide here a more constructive proof of this result for the reader.

Lemma 28 (Cauchy-Schwarz inequality for the P-norm)**:** Let $P \in \mathbb{R}^{n\times n}$, $P \succ 0$ and $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$. Then, $x^\top P y \leq \|x\|_P \|y\|_P$.

Proof. Since $P \succ 0$, there exists a matrix $M \in \mathbb{R}^{n \times n}$ such that $P = M^\top M$ [100]. Then,

$$x^\top P y = x^\top M^\top M y = (Mx)^\top M y \leq \|Mx\| \, \|My\|,$$

by the Cauchy-Schwarz inequality applied to the Euclidean norm on $\mathbb{R}^n$ [100]. Note that

$$\|Mx\| = \sqrt{(Mx)^\top Mx} = \sqrt{x^\top M^\top M x} = \sqrt{x^\top P x} = \|x\|_P.$$

Similarly, $\|My\| = \|y\|_P$. Thus, $x^\top P y \leq \|x\|_P \|y\|_P$. □

We now show how the non-resilience of subsytem (9.4) translates to the value of $z_{max}^{P_N}$.

Lemma 29: With $P_N \succ 0$ and $z_{max}^{P_N} = \max\limits_{w_N \in \mathcal{W}_N} \big\{ \min\limits_{u_N \in \mathcal{U}_N} \|C_N w_N + B_N u_N\|_{P_N} \big\}$, we have $-C_N \mathcal{W}_N \nsubseteq B_N \mathcal{U}_N \iff z_{max}^{P_N} > 0$.

Proof. If $-C_N \mathcal{W}_N \subseteq B_N \mathcal{U}_N$, then for all $w_N \in \mathcal{W}_N$, there exists $u_N \in \mathcal{U}_N$ such that $C_N w_N + B_N u_N = 0$. Hence, $\min\limits_{u_N \in \mathcal{U}_N} \big\{ \|C_N w_N + B_N u_N\|_{P_N} \big\} = 0$ for all $w_N \in \mathcal{W}_N$, i.e., $z_{max}^{P_N} = 0$.

On the other hand, if $-C_N \mathcal{W}_N \nsubseteq B_N \mathcal{U}_N$, there exists $w_N \in \mathcal{W}_N$ such that $C_N w_N + B_N u_N \neq 0$ for all $u_N \in \mathcal{U}_N$. The function $u_N \mapsto \|C_N w_N + B_N u_N\|_{P_N}$ is continuous, nonnegative and $\mathcal{U}_N$ is compact, hence it reaches a minimum which cannot be null on $\mathcal{U}_N$, i.e., $\min\limits_{u_N \in \mathcal{U}_N} \big\{ \|C_N w_N + B_N u_N\|_{P_N} \big\} > 0$. Then, $z_{max}^{P_N} > 0$. □

9.6 Summary

In this chapter we studied the resilience of linear networks and established network specific stabilizability conditions. Building on these results, we investigated how a partial loss of control authority over actuators of a subsystem would affect the stabilizability of the whole network, depending on the resilience of the malfunctioning subsystem. We were able to quantify how unresilient can a subsystem be so as not to destabilize its network.

Chapter 10

Resilience of an Orbital Inspection Mission

10.1 Introduction

With an increase in the number of active satellites, there is a growing demand for on-orbit satellite inspection, e.g., to assess damage on satellites, prevent unnecessary spacewalks of astronauts, or enforce the ban of space weapons [135]–[137]. The importance of satellite inspection is also reflected by the creation of spacecraft entirely dedicated to on-orbit inspections, like the robot Laura from the Rogue Space Systems Corporation[1].

Partly inspired by the on-orbit servicing Restore-L mission [12], our scenario of interest consists of an *orbital inspection* of a satellite by a spacecraft that completes a full revolution around the target satellite. Following an on-board computer error, the inspecting spacecraft endures a *loss of control authority* over one of its thrusters, similarly to what happened to the Nauka module when docked to the International Space Station [13]. This malfunction consists in one of the thrusters producing uncontrolled and thus possibly undesirable thrust with the same capabilities as before the malfunction.

Because of the reaction times of the sensors and thrusters [26], [27], the controller is likely not able counteract the undesirable thrust in real-time. Our objective is then to develop a control strategy for safely carrying out the inspection mission despite the malfunctioning thruster and the actuation delay. More specifically, we want the damaged spacecraft to accurately follow a safe reference trajectory. In our simulation, we choose a minimal-fuel reference trajectory generated by the convex optimization method of [138].

Relying on our works [34], [39], this chapter studies the resilience of the orbital inspection mission aforementioned and has three main contributions. Firstly, we establish the resilience of a spacecraft with nonlinear dynamics. Secondly, we build a resilient trajectory tracking controller with guaranteed performance for the nonlinear spacecraft dynamics. Finally, we demonstrate that on-orbit inspection can be performed safely despite actuation delay and a loss of control authority over a thruster.

The remainder of this chapter is structured as follows. Section 10.2 introduces our problem of interest and the relative dynamics of the satellites. In Section 10.3, we ignore actuation delay to apply existing resilience theory to the malfunctioning spacecraft to demonstrate its remaining capabilities in terms of resilient reachability and trajectory tracking. Section 10.4 builds on the theory of Section 8.3 to produce a resilient

[1] https://rogue.space/orbots/

trajectory tracking controller with guaranteed performance despite actuation delay. Finally, Section 10.5 implements this controller in a numerical simulation of the inspection mission.

10.2 Motivation and background

We consider two spacecraft on circular orbit around Earth. The mission of the chaser spacecraft is to inspect the target spacecraft. As we are interested in proximity operations, we employ the Clohessy-Wiltshire equations in a local-vertical, local-horizontal frame [138]. The state vector $X = \big(x \; y \; z \; \dot{x} \; \dot{y} \; \dot{z}\big) \in \mathbb{R}^6$ represents the difference in position and velocity between the two spacecraft and initially follows the dynamics

$$\dot{X}(t) = AX(t) + r\bar{B}\bar{u}(t), \qquad X(0) = X_0 \in \mathbb{R}^6,$$

with a *thrust-to-mass ratio* $r = 1.5 \times 10^{-4}\, m/s^2$, as we consider a chaser spacecraft of mass $600\, kg$ and five PPS-1350 thrusters of maximal thrust $90\, mN$ [139] controlled by the inputs $\bar{u} = \big(\bar{u}_1 \; \bar{u}_2 \; \bar{u}_3 \; \bar{u}_4 \; \bar{u}_5\big) \in [0,1]^5$. Because the z-dynamics of the Clohessy-Witshire equations are decoupled from the other two axis, we focus on the two-dimensional dynamics in the (x,y)-plane, with matrices:

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 3\Omega^2 & 0 & 0 & 2\Omega \\ 0 & 0 & -2\Omega & 0 \end{bmatrix} \quad \text{and} \quad \bar{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -\sqrt{2} & -1 \\ 1 & -1 & -1 & 0 & 1 \end{bmatrix},$$

where $\Omega = 0.00106\, s^{-1}$ is the mean orbital rate of the target's orbit. The thrusters do not create any torque [51] since they are rigidly fixed on the spacecraft and are aligned with its center of mass, as illustrated on Fig. 10.1. To perform its inspection mission, the chaser spacecraft relies on a fixed camera constantly pointing at the target thanks to the reaction wheels controlling the attitude of the chaser, as shown on Fig. 10.1. Because of these attitude changes, the relative dynamics lose their linearity to become

$$\dot{X}(t) = AX(t) + rR_\theta(t)\bar{B}\bar{u}(t), \qquad \text{with} \qquad R_\theta(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\big(\theta(t)\big) & -\sin\big(\theta(t)\big) \\ 0 & 0 & \sin\big(\theta(t)\big) & \cos\big(\theta(t)\big) \end{bmatrix}, \tag{10.1}$$

where θ is illustrated on Fig. 10.1 and is defined as the 2-argument arctangent $\theta(t) := \text{atan2}\big(y(t),\, x(t)\big)$.

Following the Restore-L protocol [12], we assume that the chaser must come within $80\, m$ of the target for a precise optical inspection. Hence, we want the chaser to occupy successively the 5 following holding points $(0,80)$, $(-80,0)$, $(0,-80)$, $(80,0)$ and $(0,80)$. Using the convex optimization method [138] we compute on Fig. 10.2 the minimal fuel trajectory linking these waypoints with 90 minutes transfers. For safety considerations, we consider a keep-out sphere (KOS) of radius $R_{KOS} = 50\, m$ around the target as in the Restore-L mission [12].

Because the chaser is constantly pointing its camera towards the target, its orientation angle θ (see Fig. 10.1) varies throughout the trajectory as shown on Fig. 10.3(a) and starts at $\theta(0) = 90°$ since the initial position of the spacecraft is on the y-axis, as illustrated in Fig. 10.2.

When the chaser follows the minimal-fuel reference trajectory of Fig. 10.2, the resulting thrust profile is

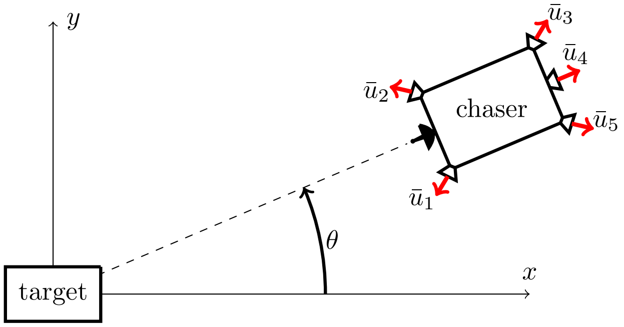


Figure 10.1: Relative positions and attitudes of the two satellites, with the camera of the chaser always pointing at the target thanks to an independent attitude control system.

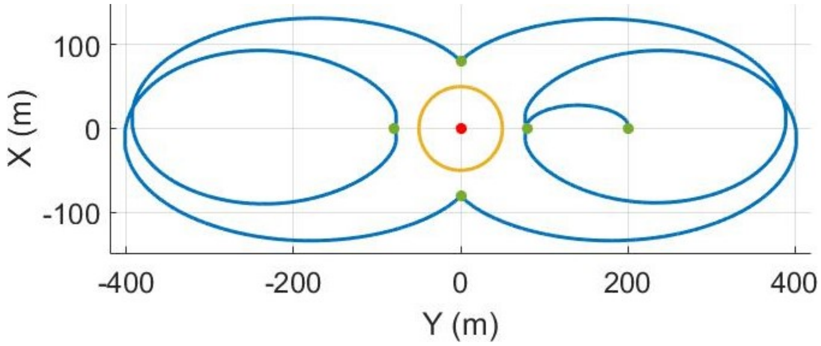
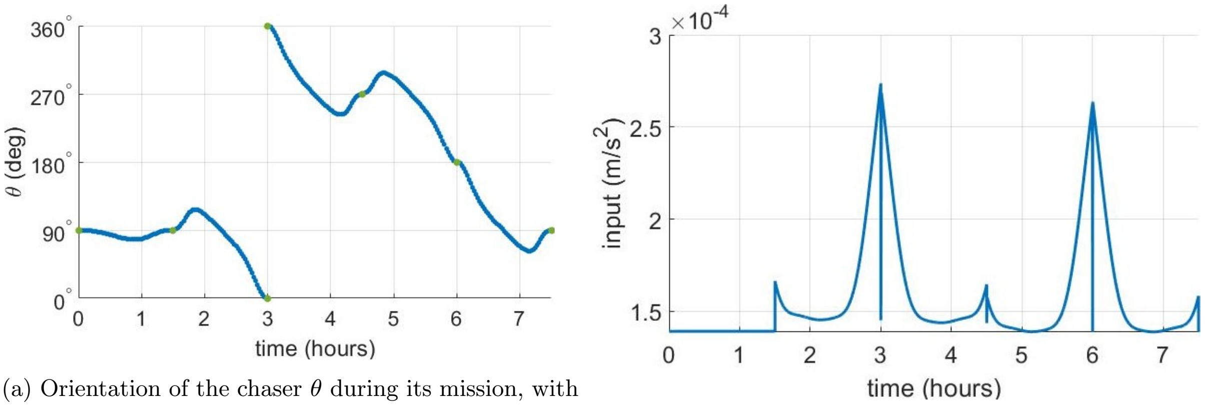


Figure 10.2: Reference minimal-fuel trajectory (blue) linking the four waypoints (green) to inspect the target satellite (red) without breaching the KOS (yellow).



(a) Orientation of the chaser θ during its mission, with the waypoints in green.

(b) Combined magnitude of the reference thrust signal $\|u_{\text{ref}}\|$.

Figure 10.3: Chaser orientation and thrust profile for the reference trajectory.

represented on Fig. 10.3(b) and shows several impulses. Their symmetry comes from the symmetry of the reference trajectory of Fig. 10.2. This thrust profile is obtained from the convex optimization method [138] by propagating dynamics (10.1) along the fuel-optimal trajectory.

Having described the nominal dynamics and the reference trajectory, we now study the malfunction impacting the chaser. Similarly to what happened to the Nauka module docked to the ISS [13], we assume that an error in the on-board computer of the chaser satellite causes the controller to lose authority over one of the thrusters. The input signal $\bar{u} \in \mathcal{F}(\bar{\mathcal{U}})$ of (10.1) is then split between the undesirable signal $w \in \mathcal{F}(\mathcal{W})$, $\mathcal{W} = [0, 1]$ and the controlled signal $u \in \mathcal{F}(\mathcal{U})$, $\mathcal{U} = [0, 1]^4$. Matrix $\bar{B}$ is accordingly split into two constant matrices $B \in \mathbb{R}^{4\times 4}$ and $C \in \mathbb{R}^4$ so that the dynamics of the malfunctioning satellite become

$$\dot{X}(t) = AX(t) + rR_\theta(t)Bu(t) + rR_\theta(t)Cw(t), \qquad X(0) = X_0 \in \mathbb{R}^4. \tag{10.2}$$

In order to account for the unavoidable sensors and thrusters delays on spacecraft [26], we should assume that the controller operates with a constant input delay $\tau > 0$ so that the dynamics of the spacecraft are in fact

$$\dot{X}(t) = AX(t) + rR_\theta(t)Bu\big(t, X(t-\tau), w(t-\tau)\big) + rR_\theta(t)Cw(t), \quad X(0) = X_0 \in \mathbb{R}^4. \tag{10.3}$$

The controller cannot react immediately to a change of the undesirable input $w(t)$ and cancel it instantaneously. Only starting at $t + \tau$ can the controller try to counteract $w(t)$. We can then formulate our problem of interest.

Problem 16: With what accuracy can the chaser satellite track the reference trajectory even after enduring a loss of control authority over any one of its thrusters?

To address Problem 16, we start by determining over which thrusters the spacecraft can resiliently lose control. This investigation is carried out in Section 10.3 on dynamics (10.2) based on the 'snap decision rule' of [23] where the controller $u(t)$ has instantaneous knowledge of the state $X(t)$ and of the uncontrolled input $w(t)$. This assumption is lifted in subsequent sections where we study how can system (10.3) perform resilient trajectory tracking despite actuation delay.

10.3 Spacecraft resilience with instantaneous control

We rely on the resilience theory established in Chapter 7 to determine over which thrusters the chaser spacecraft can lose control, while still remaining capable of accomplishing its mission.

10.3.1 Resilient reachability

Let us first recall the notion of *resilience* adapted to system (10.1).

Definition 28: System (10.1) is *resilient* to the loss of control authority over one of its thrusters if for any target $X_{goal} \in \mathbb{R}^4$ and any undesirable signal $w \in \mathcal{F}(\mathcal{W})$ there exists a control signal $u \in \mathcal{F}(\mathcal{U})$ such that the resulting malfunctioning system (10.2) can reach X_{goal} in finite time.

Resilience is not sufficient to complete the mission of Problem 16 since Definition 28 only concerns target reachability and not trajectory tracking. However, resilience is necessary for mission completion. Indeed, if the spacecraft is not resilient to the loss of one of its thrusters, then it cannot track a trajectory despite any undesirable thrust. We will then study the resilience of the nonlinear dynamics (10.1) thanks to the work accomplished in Section 8.4. System (10.2) fits within the scope of Theorem 26, hence resilience of system (10.1) is implied by *controllability* of system

$$\dot{X}(t) = AX(t) + rR_\theta(t)p(t), \qquad X(0) = X_0, \qquad p(t) \in \mathcal{P} = B\mathcal{U} \ominus (-C\mathcal{W}). \tag{10.4}$$

Note that $\mathcal{P}$ represents the amount of control authority remaining after counteracting the worst undesirable input. The controllability and stabilizability definitions from [91] can be adapted to system (10.4) as follows.

Definition 29: System (10.4) is *controllable* (resp. *stabilizable*) if for all $X_0 \in \mathbb{R}^4$ and all $X_{goal} \in \mathbb{R}^4$, there exists a time T and a control signal $p \in \mathcal{F}(\mathcal{P})$ driving the state of system (10.4) from $X(0) = X_0$ to $X(T) = X_{goal}$ (resp. to $X(T) = 0$).

We start by investigating whether system (10.1) is resilient to a loss of control authority over thruster no. 4. Indeed, this thruster plays a special role in the actuation of the chaser spacecraft due to its location shown on Fig. 10.1 and yields

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \qquad \text{and} \qquad C = \begin{bmatrix} 0 \\ 0 \\ -\sqrt{2} \\ 0 \end{bmatrix}. \tag{10.5}$$

The polytopes $B\mathcal{U}$ and $-C\mathcal{W}$ are both in $\mathbb{R}^4$, but since they are of dimension 2, we only represent these last two dimensions in Fig. 10.4. Similarly, the Minkowski difference $\mathcal{P}$ is of dimension 2 and is also illustrated in Fig. 10.4.

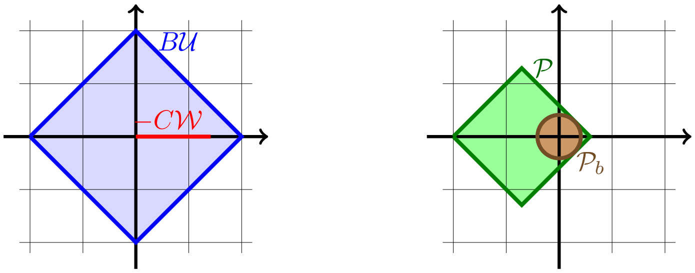


Figure 10.4: Illustration of dimensions 3 and 4 of $B\mathcal{U}$ (blue), $-C\mathcal{W}$ (red), their Minkowski difference $\mathcal{P}$ (green) and the largest ball $\mathcal{P}_b$ (brown) centered at 0 fitting inside $\mathcal{P}$ for the loss of control authority over thruster no. 4.

To prove the resilience of system (10.1) to a loss of control authority over thruster no. 4, we need to verify the controllability of nonlinear system (10.4). However, this is generally a difficult problem [111]. We will then construct a related linear time-invariant system, whose controllability implies that of system (10.4).

Proposition 36: System (10.1) is resilient to a loss of control authority over thruster no. 4.

Proof. Following Theorem 26, we will prove controllability of system (10.4) to obtain resilience of system (10.1). Because $-C\mathcal{W} \subseteq \text{int}\big(B\mathcal{U}\big)$, we have $0 \in \text{int}\big(\mathcal{P}\big)$, as seen on Fig. 10.4. Then, we can define ρ_{max} as the radius of the largest ball of dimension 2 centered at 0 and fitting inside $\mathcal{P}$, i.e., $\rho_{max} := \max\big\{\rho \geq 0 : \mathbb{B}^2(0,\rho) \subseteq \mathcal{P}\big\}$. In our case $\rho_{max} = \sqrt{2} - 1 = 0.414$. Then, the ball $\mathcal{P}_b := \mathbb{B}^2(0, \rho_{max})$ is a subset of $\mathcal{P}$ as illustrated on Fig. 10.4. Because $\mathcal{P}_b$ is a ball, there is a one-to-one correspondence between inputs $p \in \mathcal{P}_b$ and $R_\theta p \in \mathcal{P}_b$, so the dynamics (10.4) with inputs constrained to $\mathcal{P}_b$ are in fact

$$\dot{X}(t) = AX(t) + r\hat{B}p(t), \qquad p \in \mathcal{P}_b \subset \mathbb{R}^2, \qquad \hat{B} = \begin{bmatrix} 0_{2\times 2} \\ I_2 \end{bmatrix}. \tag{10.6}$$

Because the first two rows of B and C defined in Eq. (10.5) are null, the geometrical work we completed above only concerns the last two coordinates of the inputs, which explains the structure of matrix $\hat{B}$ in (10.6). To prove the controllability of system (10.6), we verify the conditions of Corollary 3.7 of [91]:

- $0 \in \mathcal{P}_b$, so taking $p = 0$ makes $\hat{B}p = 0$;
- the convex hull of $\mathcal{P}_b$ has a non-empty interior in $\mathbb{R}^2$;
- $\big[\hat{B}, A\hat{B}\big] = \begin{bmatrix} 0_{2\times 2} & I_2 \\ I_2 & * \end{bmatrix}$, so $\operatorname{rank}\big(\big[\hat{B}, A\hat{B}\big]\big) = 4$, i.e., the controllability matrix has full rank;
- the real eigenvectors of $A^\top$ are all scalar multiples of $v = (2\Omega, 0, 0, 1)$, which makes $v^\top \hat{B}p = p_2$ for all $p = (p_1, p_2) \in \mathcal{P}_b$ and p_2 can be chosen positive or negative since $\rho_{max} > 0$;
- the eigenvalues of A are $\big\{0, 0, \pm j\Omega\big\}$, so they all have a zero real part.

Hence, system (10.6) is controllable. Because system (10.4) follows the same dynamics as (10.6), and has a larger input set encompassing that of system (10.6), it is also controllable. Then, according to Theorem 26 system (10.1) is resilient to the loss of control over thruster no. 4. □

We can now proceed to the case of the other four thrusters. Because of their symmetric placement as shown on Fig. 10.1, we only need to study one thruster and similar conclusions will hold for the others. For the loss of control authority over thruster no. 1, we represent dimensions 3 and 4 of polytopes $B\mathcal{U}$, $-C\mathcal{W}$, and $\mathcal{P}$ on Fig. 10.5.

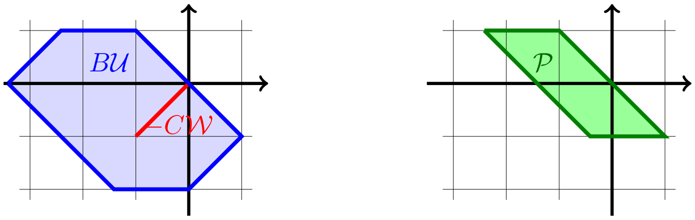


Figure 10.5: Illustration of dimensions 3 and 4 of $B\mathcal{U}$ (blue), $-C\mathcal{W}$ (red) and their Minkowski difference $\mathcal{P}$ (green) for the loss of control authority over thruster no. 1.

Note that $(0,0)$ is on the boundary of $\mathcal{P}$, so that $\rho_{max} = 0$, no ball of positive radius centered at 0 can fit inside $\mathcal{P}$. This issue is much more problematic than just preventing us from reusing the proof of Proposition 36. Indeed, let $\mathcal{T}_{\text{ref}} := \big\{X_{\text{ref}}(t) : \dot{X}_{\text{ref}}(T) = AX_{\text{ref}} + rR_\theta(t)p_{\text{ref}}(t),\ t \geq 0\big\}$ be the reference trajectory of Fig. 10.2, where control law $p_{\text{ref}} \in \mathcal{F}(\mathcal{P}_{\text{ref}})$ is obtained with the trajectory propagation algorithm of [138]. To produce this trajectory, we need $0 \in \text{int}(\mathcal{P}_{\text{ref}})$ as shown on Fig. 10.6. Since $0 \notin \text{int}(\mathcal{P})$, we have $\mathcal{P}_{\text{ref}} \nsubseteq \mathcal{P}$. Therefore, the spacecraft cannot track $\mathcal{T}_{\text{ref}}$ after the loss of control authority over a thruster other than no. 4. Note that adding a sixth thruster instead of the camera (see Fig. 10.1) would guarantee $0 \in \text{int}(\mathcal{P})$ after the loss of control over any single thruster.

To simplify further discussions, let us assume that $\mathcal{P}_{\text{ref}}$ is the smallest ball centered at 0 encompassing all $p_{\text{ref}}(t)$, i.e., $\mathcal{P}_{\text{ref}} = \mathbb{B}^2(0, \rho_{\text{ref}})$ with $\rho_{\text{ref}} := \min\big\{\rho > 0 : p_{\text{ref}}(t) \in \mathbb{B}(0,\rho) \text{ for all } t \geq 0\big\}$. For the reference trajectory of Fig. 10.2, radius $\rho_{\text{ref}} = 4.85 \times 10^{-4}$ and is illustrated on Fig. 10.6.

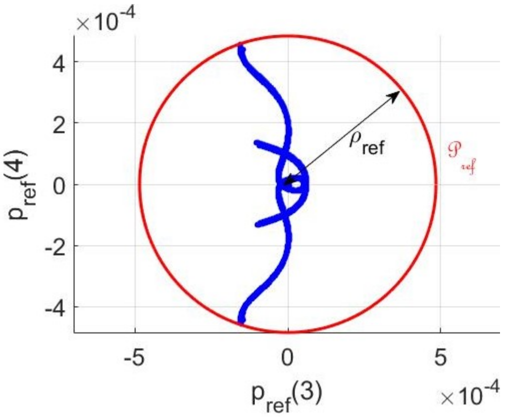


Figure 10.6: Dimensions 3 and 4 of the reference thrust inputs p_{ref} (blue) included in the ball $\mathcal{P}_{\text{ref}}$ (red) of radius ρ_{ref} (black) for the reference trajectory $\mathcal{T}_{\text{ref}}$.

10.3.2 Resilient trajectory tracking and robustness to initial state difference

Following the discussion above, we will only investigate resilient trajectory tracking for the loss of control authority over thruster no. 4. In this scenario, $\rho_{max} = 0.414 >> \rho_{\text{ref}} = 4.85 \times 10^{-4}$. Then, the malfunctioning spacecraft has a large amount of control authority left even after counteracting the worst undesirable thrust and producing the reference thrust input. Let us detail why this remaining thrust capability will be sorely needed.

The initial state of the malfunctioning spacecraft X_0 is most likely not exactly equal to $X_{\text{ref}}(0)$, the initial of reference trajectory $\mathcal{T}_{\text{ref}}$, which was designed before the spaceflight. We then need to design a tracking controller with robustness to uncertainty on the initial state. Moreover, if the difference $X_0 - X_{\text{ref}}(0)$ is not actively reduced, it can grow exponentially with time [122]. Thus, we need the extra thrust capability mentioned earlier to counteract $X(t) - X_{\text{ref}}(t)$. Formally, we pick $\varepsilon > 0$ and define input set $\mathcal{P}_\varepsilon := \mathbb{B}^2(0, \varepsilon)$ to overcome $X(t) - X_{\text{ref}}(t)$. For the robust tracking of $\mathcal{T}_{\text{ref}}$ to be admissible, we then need $\mathcal{P}_\varepsilon \oplus \mathcal{P}_{\text{ref}} \subseteq \mathcal{P}$, where we recall $\mathcal{P}$ as the set of control inputs remaining after counteracting the worst undesirable thrust from malfunctioning thruster no. 4. We now introduce the dynamics tasked with counteracting the initial state error

$$\dot{Y}(t) = AY(t) + rR_\theta(t)p_\varepsilon(t), \quad Y(0) = X_0 - X_{\text{ref}}(0), \quad p_\varepsilon(t) \in \mathcal{P}_\varepsilon, \tag{10.7}$$

where $R_\theta(t)$ is the rotation matrix tracking position $X(t)$ of system (10.2).

Proposition 37: If $\varepsilon + \rho_{\text{ref}} \leq \rho_{max}$, then system (10.7) is stabilizable in a finite time t_f and the reference trajectory $\mathcal{T}_{\text{ref}}$ can be tracked exactly by system (10.2) after time t_f, i.e., $X(T) = X_{\text{ref}}(T)$ for all $t \geq t_f$.

Proof. We start with the same trick as in the proof of Proposition 36 by noticing that $\mathcal{P}_\varepsilon = \mathbb{B}^2(0, \varepsilon)$ is left unchanged by the rotation matrix R_θ. Then, system (10.7) has a one-to-one correspondence with the following linear system

$$\dot{Y}(t) = AY(t) + r\hat{B}p_\varepsilon(t), \quad Y(0) = X_0 - X_{\text{ref}}(0), \quad p_\varepsilon(t) \in \mathcal{P}_\varepsilon, \quad \hat{B} = \begin{bmatrix} 0_{2\times 2} \\ I_2 \end{bmatrix}. \tag{10.8}$$

Since $0 \in \text{int}(\mathcal{P}_\varepsilon)$, $Re(\lambda(A)) \leq 0$, and $\text{rank}\left(\left[\hat{B}\ A\hat{B}\right]\right) = 4$, Corollary 3.6 of [91] states that system (10.8) is stabilizable in a finite time t_f and so is system (10.7) by construction.

Therefore, there exists a signal $p_\varepsilon \in \mathcal{F}(\mathcal{P}_\varepsilon)$ on $[0, t_f]$ yielding $Y(t_f) = 0$ in system (10.7). Since $0 \in \mathcal{P}_\varepsilon$, we extend the control signal to $p_\varepsilon(t) = 0$ for all $t > t_f$. We now define the control law $p_{track}(t) := p_\varepsilon(t) + p_{\text{ref}}(t)$. Note that $\mathcal{P}_\varepsilon \oplus \mathcal{P}_{\text{ref}} = \mathbb{B}(0, \varepsilon) \oplus \mathbb{B}(0, \rho_{\text{ref}}) = \mathbb{B}(0, \varepsilon + \rho_{\text{ref}}) \subseteq \mathbb{B}(0, \rho_{max})$ since $\varepsilon + \rho_{\text{ref}} \leq \rho_{max}$. By definition of ρ_{max}, $\mathbb{B}(0, \rho_{max}) \subseteq \mathcal{P}$. Thus, $\mathcal{P}_\varepsilon \oplus \mathcal{P}_{\text{ref}} \subseteq \mathcal{P}$, i.e., $p_{track}(t) \in \mathcal{P}$ for all $t \geq 0$.

Let $w \in \mathcal{F}(\mathcal{W})$ be any undesirable input signal. Then, by definition of $\mathcal{P}$, there exists $u \in \mathcal{F}(\mathcal{U})$ such that $Bu(t) = p_{track}(t) - Cw(t)$ for all $t \geq 0$. We now implement this controller for $T \geq t_f$ in system (10.2):

$$\begin{aligned} X(T) &= e^{AT}\left(X_0 + \int_0^T e^{-At} r R_\theta(t)\big(Bu(t) + Cw(t)\big)\, dt\right) = e^{AT}\left(X_0 + \int_0^T e^{-At} r R_\theta(t)\big(p_\varepsilon(t) + p_{\text{ref}}(t)\big)\, dt\right) \\ &= e^{AT}\left(X_0 + \int_0^T e^{-At} r R_\theta(t) p_\varepsilon(t)\, dt + e^{-AT} X_{\text{ref}}(T) - X_{\text{ref}}(0)\right), \end{aligned}$$

because $X_{\text{ref}}(T) = e^{AT}\left(X_{\text{ref}}(0) + \int_0^T e^{-At} r R_\theta(t) p_{\text{ref}}(t) dt\right)$. Then,

$$X(T) - X_{\text{ref}}(T) = e^{AT}\left(X_0 - X_{\text{ref}}(0) + \int_0^T e^{-At} r R_\theta(t) p_\varepsilon(t)\, dt\right) = e^{AT}\left(X_0 - X_{\text{ref}}(0) + \int_0^{t_f} e^{-At} r R_\theta(t) p_\varepsilon(t)\, dt\right),$$

since $p_\varepsilon(t) = 0$ for $t > t_f$. By definition of p_ε,

$$Y(t_f) = 0 = e^{At_f}\left(Y(0) + \int_0^{t_f} e^{-At} r R_\theta(t) p_\varepsilon(t)\, dt\right), \quad \text{i.e.,} \quad X_0 - X_{\text{ref}}(0) + \int_0^{t_f} e^{-At} r R_\theta(t) p_\varepsilon(t)\, dt = 0.$$

Therefore, $X(T) = X_{\text{ref}}(T)$ for all $T \geq t_f$. □

Proposition 37 states that as long as $\varepsilon + \rho_{\text{ref}} \leq \rho_{max}$, there exists a finite time t_f after which any trajectory $\mathcal{T}_{\text{ref}}$ can be tracked perfectly despite the loss of control authority over a thruster. Since ε describes the maximal input magnitude of system (10.7), ε is inversely correlated with its stabilization time t_f. Then, the constraint $\varepsilon + \rho_{\text{ref}} \leq \rho_{max}$ yields that the smaller ρ_{ref}, the larger ε and so the smaller t_f is. In other words, the smaller the inputs required to track the reference trajectory, the faster the spacecraft can resume perfect tracking after a loss of control authority. Let us now investigate how the spacecraft would perform if the controller could not react instantly to undesirable thrust inputs.

10.4 Spacecraft resilience in the presence of actuation delay

In this section we extend the resilience theory of linear systems with actuation delays of Section 8.3 to the rotating dynamics (10.3) to build an answer to Problem 16. We start by verifying whether the open-loop controller of Section 8.3.3 can be applied to the spacecraft.

10.4.1 Open-loop controller

To apply the open-loop controller of Section 8.3.3, we need to extend the minimal correction time T_c of Definition 21 to the nonlinear spacecraft dynamics (10.3) as $T_c = \inf\big\{T \geq \tau : -R_\theta^{-1}(t+T) e^{AT} R_\theta(t) C\mathcal{W} \subseteq B\mathcal{U}$ for all $t \geq 0\big\}$. Indeed, to extend the proof of Theorem 25 we need to cancel the following terms appearing in the calculation of $X(T)$:

$$\int_0^{T-T_c} e^{-At} R_\theta(t) Cw(t) dt + \int_{T_c}^T e^{-At} R_\theta(t) Bu(t) dt = \int_0^{T-T_c} e^{-At}\big[R_\theta(t) Cw(t) + e^{-AT_c} R_\theta(t+T_c) Bu(t+T_c)\big] dt,$$

where the bracketed term can only be set to 0 for all $w(t) \in \mathcal{W}$ if the aforementioned extension of T_c holds. However, this definition of T_c creates a circular dependency of T_c on $R_\theta(t)$, which depends on state $X(t)$, which is in turn modified by controller u relying on T_c. Then, T_c can only be properly defined if it is not impacted by $\theta(t)$. Let us investigate if such a minimal correction time T_c can be defined.

Thanks to the two recursions: $A^{2n+2} = (-1)^n \Omega^{2n} A^2$ and $A^{2n+3} = (-1)^n \Omega^{2n} A^3$ for all $n \in \mathbb{N}$, we can calculate the exponential series

$$e^{At} = \begin{bmatrix} 4 - 3\cos(\Omega t) & 0 & \frac{1}{\Omega}\sin(\Omega t) & \frac{2}{\Omega}\big(1 - \cos(\Omega t)\big) \\ 6\big(\sin(\Omega t) - \Omega t\big) & 1 & \frac{2}{\Omega}\big(\cos(\Omega t) - 1\big) & -3t + \frac{4}{\Omega}\sin(\Omega t) \\ 3\Omega\sin(\Omega t) & 0 & \cos(\Omega t) & 2\sin(\Omega t) \\ 6\Omega\big(\cos(\Omega t) - 1\big) & 0 & -2\sin(\Omega t) & 4\cos(\Omega t) - 3 \end{bmatrix}.$$

After the loss of control authority over thruster no. 4, matrices B and C are defined in (10.5). Since the first two rows of B are null, the first two components of $-R_\theta^{-1}(t + T_c)e^{AT_c}R_\theta(t)C$ should also be zero for all $t \geq 0$, i.e.,

$$0 = \cos\big(\theta(t)\big)\sin(\Omega T_c) + 2\sin\big(\theta(t)\big)\big(1 - \cos(\Omega T_c)\big), \tag{10.9}$$

$$0 = 2\cos\big(\theta(t)\big)\big(\cos(\Omega T_c) - 1\big) + \sin\big(\theta(t)\big)\big(4\sin(\Omega T_c) - 3\Omega T_c\big). \tag{10.10}$$

For (10.9) to hold independently of $\theta(t)$, we need $T_c = \frac{2\pi}{\Omega}n$, $n \in \mathbb{N}$. However, (10.10) would yield $\sin\big(\theta(t)\big) = 0$, which prevents to track trajectory $\mathcal{T}_{\text{ref}}$. Therefore, we cannot define a minimal correction time T_c for the nonlinear spacecraft dynamics (10.3). Then, we cannot cancel exactly $Cw(t)$ after some actuation delay as we did in Section 8.3.3. Without this perfect cancellation an open-loop controller like in Theorem 25 would not be able to track a trajectory. We will then transform this controller into a feedback controller.

10.4.2 Closed-loop controller

Motivated by the ISS malfunction [13] where the undesirable thrust was constant, we will assume in this section that w is Lipschitz. As in Theorem 25 we partition $\mathcal{P} = B\mathcal{U} \ominus -C\mathcal{W}$ into two parts: $\mathcal{P}_\varepsilon$ for the feedback correction and $\mathcal{P}_{\text{ref}}$ for the trajectory tracking, but we want to replace p_ε by a linear feedback controller. However, $u(t)$ has only access to $X(t - \tau)$ and not $X(t)$, hence a straightforward linear feedback is not possible. We will replace $X(t)$ by a state predictor $X_p(t)$, designed to predict $X(t)$ based on the information available at time $t - \tau$. We will use a predictor adapted from [94] which takes advantage of the system's dynamics:

$$X_p(t) = e^{A\tau}X(t - \tau) + \int_{t-\tau}^{t} e^{A(t-s)} r R_\theta(s)\big(Bu(s) + Cw(s - \tau)\big)ds. \tag{10.11}$$

Before stating our main theorem for resilient trajectory tracking, we recall the definitions of ρ_{max} from Proposition 36, $\rho_{max} = \max\big\{\rho \geq 0 : \mathbb{B}^2(0, \rho) \subseteq \mathcal{P}\big\}$, the log-norm $\mu(A) = \max\big\{\lambda((A + A^\top)/2)\big\}$, and the reference trajectory $\mathcal{T}_{\text{ref}} = \big\{X_{\text{ref}}(t) : \dot{X}_{\text{ref}}(t) = AX_{\text{ref}}(t) + rp_{\text{ref}}(t) \text{ for all } t \geq 0\big\}$, $p_{\text{ref}} \in \mathcal{F}(\mathcal{P}_{\text{ref}})$ with $\mathcal{P}_{\text{ref}} = \mathbb{B}^2(0, \rho_{\text{ref}})$.

Theorem 31: Let $K \in \mathbb{R}^{4\times 4}$ such that $\tilde{A} := A - rBK$ is Hurwitz, and let $P \in \mathbb{R}^{4\times 4}$ and $Q \in \mathbb{R}^{4\times 4}$ such that $P \succ 0$, $Q \succ 0$ and $\tilde{A}^\top P + P\tilde{A} = -Q$. Define $\alpha := \frac{\lambda^Q_{min}}{2\lambda^P_{max}}$, $\beta := r\sqrt{\lambda^P_{max}}\|C\|L\tau$, and $\gamma := r\|BK\|\frac{e^{\mu(A)\tau} - 1}{\mu(A)}$. For $L > 0$, let $\varepsilon := \frac{\|BK\|}{\sqrt{\lambda^P_{min}}}\max\Big(\|X_{\text{ref}}(0) - X(0)\|_P, \frac{\beta}{\alpha}(1 + \gamma)\Big) + \gamma\|C\|L\tau$.

If $\varepsilon+\rho_{\text{ref}} \leq \rho_{max}$, then, for all $w \in \mathcal{F}(\mathcal{W})$ with a Lipschitz constant L, malfunctioning spacecraft (10.3) can track reference trajectory $\mathcal{T}_{\text{ref}}$ with a tolerance $\big\|X_{\text{ref}}(t)-X(t)\big\| \leq \frac{1}{\sqrt{\lambda^P_{min}}} \max\Big(\|X_{\text{ref}}(0)-X(0)\|_P,\ \frac{\beta}{\alpha}(1+\gamma)\Big)$ for all $t \geq \tau$.

Proof. The existence of matrix K is justified by the controllability of the pair (A, B) [122]. Since the resulting $\tilde{A}$ is Hurwitz, matrices $P \succ 0$ and $Q \succ 0$ exist according to Lyapunov theory [121]. We consider any $w \in \mathcal{F}(\mathcal{W})$ with a Lipschitz constant L and assume that $\varepsilon + \rho_{\text{ref}} \leq \rho_{max}$. We define $\mathcal{P}_\varepsilon := \mathbb{B}^2(0,\varepsilon)$ and recall $\mathcal{P}_b = \mathbb{B}^2(0,\rho_{max})$ as introduced in Proposition 36. Then,

$$\mathcal{P}_\varepsilon \oplus \mathcal{P}_{\text{ref}} = \mathbb{B}^2(0,\varepsilon) \oplus \mathbb{B}^2(0,\rho_{\text{ref}}) = \mathbb{B}^2(0,\varepsilon+\rho_{\text{ref}}) \subseteq \mathbb{B}^2(0,\rho_{max}) = \mathcal{P}_b \subseteq \mathcal{P}.$$

For $t \geq \tau$ we introduce control signals u_w, u_{ref} and u_ε such that $Bu_w(t) := -Cw(t-\tau)$, $Bu_{\text{ref}}(t) := R_\theta^{-1}(t)p_{\text{ref}}(t)$ and $Bu_\varepsilon(t) := R_\theta^{-1}(t)BK\big(X_{\text{ref}}(t)-X_p(t)\big)$ with the predictor X_p from (10.11). We consequently define the feedback control law u by

$$Bu(t) := Bu_w(t) + Bu_{\text{ref}}(t) + Bu_\varepsilon(t) = -Cw(t-\tau) + R_\theta^{-1}(t)p_{\text{ref}}(t) + R_\theta^{-1}(t)BK\big(X_{\text{ref}}(t) - X_p(t)\big). \quad (10.12)$$

To prove that controller (10.12) is admissible we need to show that $Bu(t) \in B\mathcal{U}$ for all $t \geq \tau$. Firstly, $Bu_w(t) = -Cw(t-\tau) \in -C\mathcal{W}$. Because $\mathcal{P}_{\text{ref}}$ is a ball centered on 0, it is invariant by rotation R_θ. Then, $R_\theta^{-1}(t)p_{\text{ref}}(t) \in \mathcal{P}_{\text{ref}}$, i.e., $Bu_{\text{ref}}(t) \in \mathcal{P}_{\text{ref}}$. Since $-C\mathcal{W} \oplus \mathcal{P}_{\text{ref}} \oplus \mathcal{P}_\varepsilon \subseteq B\mathcal{U}$, it now suffices to show that $Bu_\varepsilon(t) \in \mathcal{P}_\varepsilon = \mathbb{B}^2(0,\varepsilon)$. To do so, we first apply (10.12) to dynamics (10.3). By definition of $\mathcal{T}_{\text{ref}}$, we have $rp_{\text{ref}}(t) = \dot{X}_{\text{ref}}(t) - AX_{\text{ref}}(t)$, and thus

$$\begin{aligned}\dot{X}(t) &= AX(t) + rR_\theta(t)Bu(t) + rR_\theta(t)Cw(t)\\ &= AX(t) - rR_\theta(t)Cw(t-\tau) + \dot{X}_{\text{ref}}(t) - AX_{\text{ref}}(t) + rBK\big(X_{\text{ref}}(t) - X_p(t)\big) + rR_\theta(t)Cw(t),\end{aligned}$$

$$\text{i.e.,}\quad \dot{X}(t) - \dot{X}_{\text{ref}}(t) = (A - rBK)\big(X(t) - X_{\text{ref}}(t)\big) + rR_\theta(t)\big(Cw(t) - Cw(t-\tau)\big) + rBK\big(X(t) - X_p(t)\big).$$

We define $Y(t) := X(t) - X_{\text{ref}}(t)$, $\Delta C(t) := rR_\theta(t)\big(Cw(t) - Cw(t-\tau)\big)$ and $\Delta X(t) := rBK\big(X(t) - X_p(t)\big)$ so that $\dot{Y}(t) = \tilde{A}Y(t) + \Delta C(t) + \Delta X(t)$. Inspired by the method described in Section 9.3 of [122], we will now show that $Y(t)$ is bounded, which in turn will prove that $Bu_\varepsilon(t) \in \mathcal{P}_\varepsilon$ and hence that control law (10.12) is admissible. We consider the derivative of the norm $Y(t)^\top PY(t) = \|Y(t)\|_P^2$ and obtain the following:

$$\frac{d}{dt}\|Y(t)\|_P^2 = \dot{Y}(t)^\top PY(t) + Y(t)^\top P\dot{Y}(t) = Y(t)^\top\big(\tilde{A}^\top P + P\tilde{A}\big)Y(t) + 2Y(t)^\top P\big(\Delta C(t) + \Delta X(t)\big).$$

Since $\|\cdot\|_P$ is a norm, the Cauchy-Schwarz inequality [100] yields $Y(t)^\top P\Delta C(t) \leq \|Y(t)\|_P\|\Delta C(t)\|_P$. Then, using $\|R_\theta(t)\| = 1$ and the Lipschitz constant L of w, we have

$$\|\Delta C(t)\|_P \leq \sqrt{\lambda^P_{max}}r\|R_\theta(t)\|\|Cw(t) - Cw(t-\tau)\| \leq r\sqrt{\lambda^P_{max}}\|C\|\big|w(t) - w(t-\tau)\big| \leq r\sqrt{\lambda^P_{max}}\|C\|L\tau = \beta.$$

Similarly,

$$\|\Delta X(t)\|_P \leq r\sqrt{\lambda^P_{max}}\big\|BK\big(X(t) - X_p(t)\big)\big\| \leq r\sqrt{\lambda^P_{max}}\|BK\|\|X(t) - X_p(t)\|.$$

We write the state of the system $X(t)$ in a form similar to (10.11) to compare it with X_p:

$$X(t) = e^{A\tau}X(t-\tau) + \int_{t-\tau}^{t} e^{A(t-s)} r R_\theta(s)\big(Bu(s) + Cw(s)\big)ds.$$

Then, reusing the log-norm $\mu(A)$ [105] as in Theorem 24, we obtain

$$\begin{aligned}\big\|X(t) - X_p(t)\big\| &\leq \int_{t-\tau}^{t} \big\|e^{A(t-s)}\big\| r \|R_\theta(s)\| \big\|Cw(s) - Cw(s-\tau)\big\| ds \\ &\leq r\int_{t-\tau}^{t} e^{\mu(A)(t-s)}\|C\|L\tau\, ds = r\|C\|L\tau\frac{e^{\mu(A)\tau}-1}{\mu(A)}.\end{aligned}$$

Therefore, $\|\Delta X(t)\|_P \leq r\sqrt{\lambda^P_{max}}\|BK\|r\|C\|L\tau\frac{e^{\mu(A)\tau}-1}{\mu(A)} = \beta\gamma$, so that

$$\frac{d}{dt}\|Y(t)\|_P^2 \leq -Y^\top(t)QY(t) + 2\|Y(t)\|_P\big(\|\Delta C(t)\|_P + \|\Delta X(t)\|_P\big) \leq -\frac{\lambda^Q_{min}}{\lambda^P_{max}}\|Y(t)\|_P^2 + 2\beta(1+\gamma)\|Y(t)\|_P.$$

Indeed, $Q \succ 0$ yields $-Y^\top QY \leq -\lambda^Q_{min}Y^\top Y$ [122] and $\|Y\|_P^2 \leq \lambda^P_{max}Y^\top Y$ leads to $-Y^\top Y \leq \frac{-1}{\lambda^P_{max}}\|Y\|_P^2$. Hence, we obtain

$$\frac{d}{dt}\|Y(t)\|_P^2 \leq -2\alpha\|Y(t)\|_P^2 + 2\beta(1+\gamma)\|Y(t)\|_P.$$

Since $\frac{d}{dt}\|Y(t)\|_P^2 = 2\|Y(t)\|_P\frac{d}{dt}\|Y(t)\|_P$, we have $\frac{d}{dt}\|Y(t)\|_P \leq -\alpha\|Y(t)\|_P + \beta(1+\gamma)$ for $Y(t) \neq 0$. Let us define the function $f(v) := -\alpha v + \beta(1+\gamma)$. The solution of the differential equation $\dot{v}(t) = f\big(v(t)\big)$ with initial condition $v(0) = \|Y(0)\|_P$ is $v(t) = e^{-\alpha t}\left(\|Y(0)\|_P - \frac{\beta}{\alpha}(1+\gamma)\right) + \frac{\beta}{\alpha}(1+\gamma)$. Since $f(v)$ is Lipschitz in v and $\frac{d}{dt}\|Y(t)\|_P \leq f\big(\|Y(t)\|_P\big)$, the Comparison Lemma of [122] states that $\|Y(t)\|_P \leq v(t)$ for all $t \geq 0$. Then,

$$\|Y(t)\|_P \leq e^{-\alpha t}\left(\|Y(0)\|_P - \frac{\beta}{\alpha}(1+\gamma)\right) + \frac{\beta}{\alpha}(1+\gamma) \xrightarrow[t\to\infty]{} \frac{\beta}{\alpha}(1+\gamma).$$

Since this bound on $\|Y(t)\|_P$ is monotonic, we have $\|Y(t)\|_P \leq \max\left(\|Y(0)\|_P,\ \frac{\beta}{\alpha}(1+\gamma)\right)$. Then,

$$\begin{aligned}\|Bu_\varepsilon(t)\| = \big\|R_\theta^{-1}(t)BK\big(X_{\text{ref}}(t) - X_p(t)\big)\big\| &\leq \big\|R_\theta^{-1}(t)\big\|\,\|BK\|\,\big(\|X_{\text{ref}}(t) - X(t)\| + \|X(t) - X_p(t)\|\big) \\ &\leq \|BK\|\left(\|Y(t)\| + r\|C\|L\tau\frac{e^{\mu(A)\tau}-1}{\mu(A)}\right) \leq \|BK\|\frac{\|Y(t)\|_P}{\sqrt{\lambda^P_{min}}} + \gamma\|C\|L\tau \leq \varepsilon,\end{aligned}$$

by definition of ε and using $\big\|R_\theta^{-1}(t)\big\| = 1$. Therefore, $Bu_\varepsilon(t) \in \mathbb{B}^2(0,\varepsilon) = \mathcal{P}_\varepsilon$. To sum up,

$$Bu(t) = Bu_w(t) + Bu_{\text{ref}}(t) + Bu_\varepsilon(t) \in -C\mathcal{W} \oplus \mathcal{P}_{\text{ref}} \oplus \mathcal{P}_\varepsilon \subseteq -C\mathcal{W} \oplus \mathcal{P} \subseteq B\mathcal{U}$$

for all $t \geq 0$. Therefore, control law (10.12) is admissible and the announced tracking tolerance is verified:

$$\|X_{\text{ref}}(t) - X(t)\| = \|Y(t)\| \leq \frac{\|Y(t)\|_P}{\sqrt{\lambda^P_{min}}} \leq \max\left(\frac{\|Y(0)\|_P}{\sqrt{\lambda^P_{min}}},\ \frac{\beta+\beta\gamma}{\alpha\sqrt{\lambda^P_{min}}}\right).$$

□

Theorem 31 provides a controller with trajectory tracking guarantees for the malfunctioning spacecraft (10.3). The tracking error is dictated by two main terms β and $\beta\gamma$, which respectively bound the prediction errors on the undesirable thrust $\|\Delta C(t)\|_P$ and the state $\|\Delta X(t)\|_P$. The term $\alpha\sqrt{\lambda_{min}^P}$ is just a conversion factor between the P-norm and the Euclidean norm. The term $\|X_{\text{ref}}(0) - X(0)\|_P$ in the definition of ε ensures that the controller is robust to initial state uncertainty as discussed in Section 10.3.2. The tracking tolerance of Theorem 31 can also be interpreted as a convergence radius for the controller. Indeed, in the proof of Theorem 31 we showed that $\|X_{\text{ref}}(t) - X(t)\|$ must be small enough for $u_\varepsilon(t)$ to be admissible. If $\|X_{\text{ref}}(0) - X(0)\|$ is too large, control law (10.12) might not be admissible and the convergence of $X(t)$ to $\mathcal{T}_{\text{ref}}$ cannot be guaranteed. The choice of matrices K, P and Q must then be optimized for the controller to be sufficiently robust to initial state uncertainty. A similar but simplified optimization is discussed in Exercise 9.1 of [122].

We will now implement controller (10.12) embedded with predictor (10.11) on the malfunctioning spacecraft dynamics (10.3) to study its performance over the course of the inspection mission and respond to Problem 16.

10.5 Numerical simulation

In this section we study whether controller (10.12) can fulfill the mission scenario of Section 10.2. Recall that the statement of Problem 16 specifies neither the malfunctioning thruster, nor the regularity of the undesirable thrust signal w, nor the value of the actuation delay τ. As discussed above Fig. 10.6, tracking the reference trajectory of Fig. 10.2 appears to only be possible if the malfunctioning thruster is no. 4. Therefore, we will only investigate scenarios featuring the loss of control authority over thruster no. 4. In such a case $\rho_{max} > 0$, which enables us to apply Theorem 31 and use controller (10.12). Then, to address Problem 16 we will simulate a variety of scenarios with different undesirable thrust signals and different actuation delays. We perform all the simulations in MATLAB and all the codes are accessible on github[2].

10.5.1 Nominal scenario

In this first scenario, we choose an actuation delay $\tau = 0.2\,s$ following [26] and a Lipschitz constant $L = 0.1$ for w so that the malfunctioning thrust cannot vary by more than a tenth of its capability every second since $\mathcal{W} = [0, 1]$. We choose matrices K, P and Q to maximize ε subject to $\varepsilon \leq \rho_{max} - \rho_{\text{ref}}$, where $\rho_{\text{ref}} = 4.85 \times 10^{-4}$ is the maximal input norm on the reference trajectory, as seen on Fig. 10.6. Ample numerical testing on MATLAB led us to believe that the optimal matrices are $Q = I$ and K such that $A - rBK$ has 4 identical eigenvalues. Then,

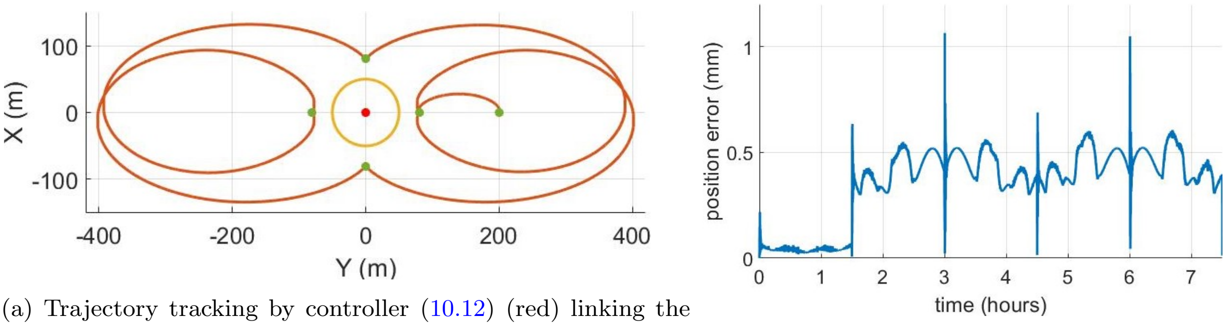
$$P = \begin{bmatrix} 2.77 & 0 & 1.77 & 0.01 \\ 0 & 2.77 & -0.01 & 1.77 \\ 1.77 & -0.01 & 8 & 0 \\ 0.01 & 1.77 & 0 & 8 \end{bmatrix} \quad \text{and} \quad K = 472\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$$

so that $\varepsilon = 0.4133 < \rho_{max} - \rho_{\text{ref}} = 0.4137$ and the tracking tolerance is $\frac{\beta(1+\gamma)}{\alpha\sqrt{\lambda_{min}^P}} = 1.5 \times 10^{-4}$ for $X(0) = X_{\text{ref}}(0)$.

Then, controller (10.12) ensures excellent tracking of the reference trajectory $\mathcal{T}_{\text{ref}}$, as shown on Fig. 10.7(a). We compute the position error between the reference state and the tracking state on Fig. 10.7(b). We observe that the position error is never larger than $1.07\,mm$ and averages only $0.36\,mm$. We acknowledge that these

[2] https://github.com/Jean-BaptisteBouvier/Spacecraft-Resilience

extremely small errors are only possible because all dynamics, states and thrusts are known exactly in our simple simulation.



(a) Trajectory tracking by controller (10.12) (red) linking the waypoints (green) to inspect the target satellite (red) without breaching the KOS (yellow).

(b) Position error between the reference and tracking trajectories.

Figure 10.7: Analysis of the trajectory tracking performance for a stochastic Lipschitz undesirable input w and actuation delay $\tau = 0.2\,s$.

To compare with the tracking tolerance of 1.5×10^{-4}, we also compute the norm difference between the reference and tracking states: $\|X(t) - X_{\text{ref}}(t)\|$. The average norm difference is 3.6×10^{-4}, while the maximal norm difference is 10.5×10^{-4}. Let us investigate why these values are slightly larger than the tracking tolerance. First, note that $\|X(t)\|^2 = x(t)^2 + y(t)^2 + \dot{x}(t)^2 + \dot{y}(t)^2$ where position $(x(t), y(t))$ is of the order of $10^2\,m$ as shown on Fig. 10.7(a), while velocity $(\dot{x}(t), \dot{y}(t))$ is of the order $10^{-1}\,m \cdot s^{-1}$. Because of these orders of magnitudes, the norm difference between reference and tracking states reflects mostly the position error. Based on Fig. 10.7(b), the maximal norm difference occurs at 3 hours and 6 hours, i.e., at the waypoints $x = \pm 80\,m$ and $y = 0\,m$ as shown on Fig. 10.7(a). At every other waypoint Fig. 10.7(b) also shows error spikes albeit of smaller magnitude. Because the sudden stop and start occurring at each waypoint are not well captured by the discrete dynamics of our simulation, the actual norm difference is larger than the threshold value of Theorem 31.

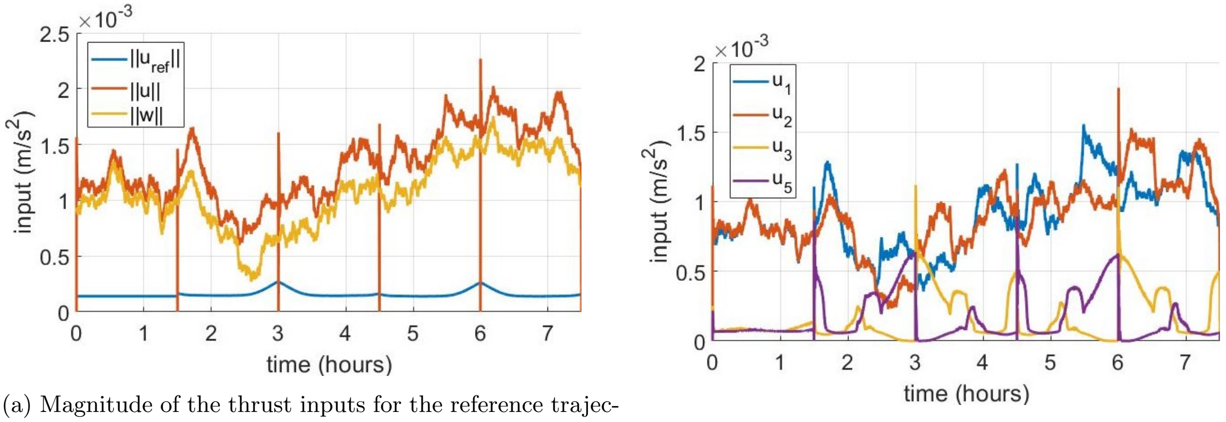
The undesirable thrust input w is generated as a stochastic signal, whose magnitude is represented in yellow in Fig. 10.8(a). To counteract w while following the reference trajectory, the controlled input u verifies approximately the intuitive relation $\|u\| \approx \|u_{\text{ref}}\| + \|w\|$. More specifically, Fig. 10.8(b) shows that thrust inputs u_3 and u_5 replicate the reference thrust profile of Fig. 10.3b, while u_1 and u_2 counteract malfunctioning thruster no. 4 as expected from their opposite placement on Fig. 10.1.

The fuel consumption on the reference and tracking trajectories is displayed on Fig. 10.9(a). The yellow curve represents the mass of fuel $m_w = 1.06\,kg$ used to produce the undesirable thrust, while the red one shows the mass of fuel $m_u = 1.31\,kg$ used by the controlled thrusters. The reference trajectory without malfunctions requires $m_{\text{ref}} = 0.16\,kg$ of fuel. As expected, $m_u \approx m_{\text{ref}} + m_w$. We have the intuition that the gap between m_u and $m_{\text{ref}} + m_w$ will grow with τ and with the unpredictability of w.

As can be expected from the tracking accuracy displayed on Fig. 10.7, the velocity tracking of the reference is also extremely accurate with velocities remaining within $0.35\,mm/s$ of each others, as illustrated on Fig. 10.9(b). As on Fig. 10.7(b), the error spikes at each waypoint and displays also the same symmetry as the orbit.

Based on Fig. 10.7, 10.8 and 10.9, controller (10.12) performs excellently and enables the system to complete its mission when w is Lipschitz and $\tau = 0.2\,s$.

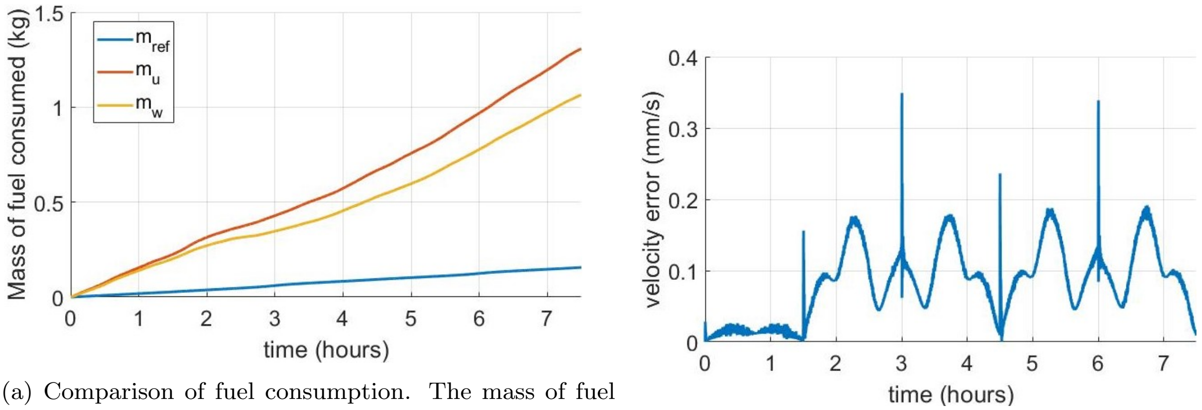
To address Problem 16 we will now study the tracking performance controller (10.12) on a variety of



(a) Magnitude of the thrust inputs for the reference trajectory $\|u_{\mathrm{ref}}\|$ (blue), for the tracking trajectory the controlled input is $\|u\|$ (red) and the undesirable input is $\|w\|$ (yellow).

(b) Thrust profiles for the four controlled thrusters of the chaser satellite on the tracking trajectory.

Figure 10.8: Analysis of the thrust profiles of the malfunctioning satellite for a Lipschitz undesirable input w and actuation delay $\tau = 0.2\,s$.



(a) Comparison of fuel consumption. The mass of fuel used to complete the reference trajectory is m_{ref} (blue). After the loss of control over thruster no. 4, it consumes a mass of fuel m_w (yellow), while the controlled thrusters use a mass m_u (red).

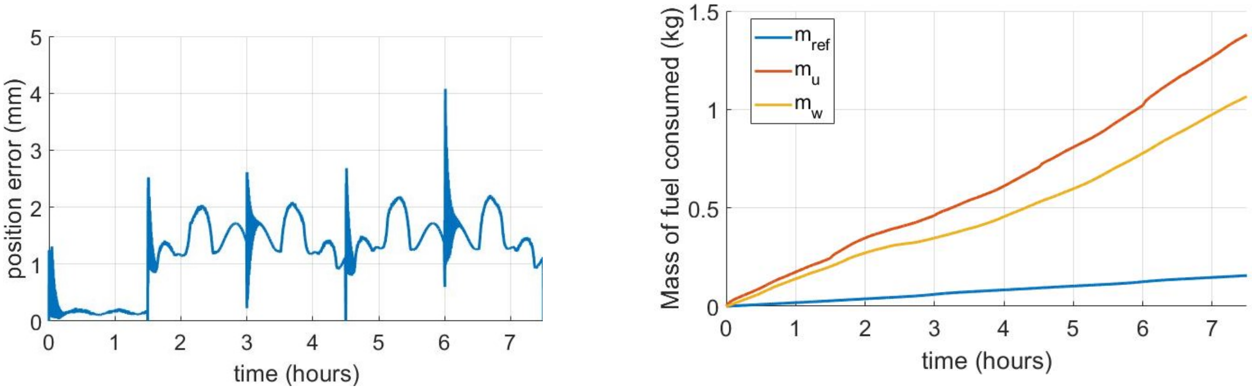
(b) Velocity error between the reference and tracking trajectories.

Figure 10.9: Comparison of the fuel consumption and velocities for a stochastic Lipschitz undesirable input w and actuation delay $\tau = 0.2\,s$.

scenarios. The first parameter that we modify is the regularity of the undesirable thrust signal w. Instead of w being Lipschitz continuous as assumed in Theorem 31, w could be bang-bang. We will also increase the actuation delay τ until controller (10.12) is unable to track the reference trajectory. The third parameter modified in these scenarios is the saturation value of w. Indeed, Fig. 10.8(a) shows that w had an average magnitude of $10^{-3}\, m/s^2$, while w is constrained in $[0,1]$ and hence could be significantly larger.

10.5.2 Lipschitz undesirable thrust and actuation delay of 8 seconds

We first increase the actuation delay τ from $0.2\, s$ to $8\, s$ and keep the same Lipschitz and stochastic undesirable thrust signal w. The previous guarantees of Theorem 31 are not valid anymore, but controller (10.12) still performs sufficiently well to not be distinguishable from the reference as in Fig. 10.7(a). Instead, we analyze the position error shown on Fig. 10.10(a). The trajectory tracks the reference with an average position error of $1.2\, mm$ and a maximal error of $4.1\, mm$. These values are extremely low but still represent a fourfold increase compared to the scenario with $\tau = 0.2\, s$.



(a) Position error between the reference and tracking trajectories.

(b) Comparisons of fuel used to complete the reference trajectory m_{ref} (blue), the tracking trajectory m_u (red), and the fuel used by malfunctioning thruster no. 4 m_w (yellow).

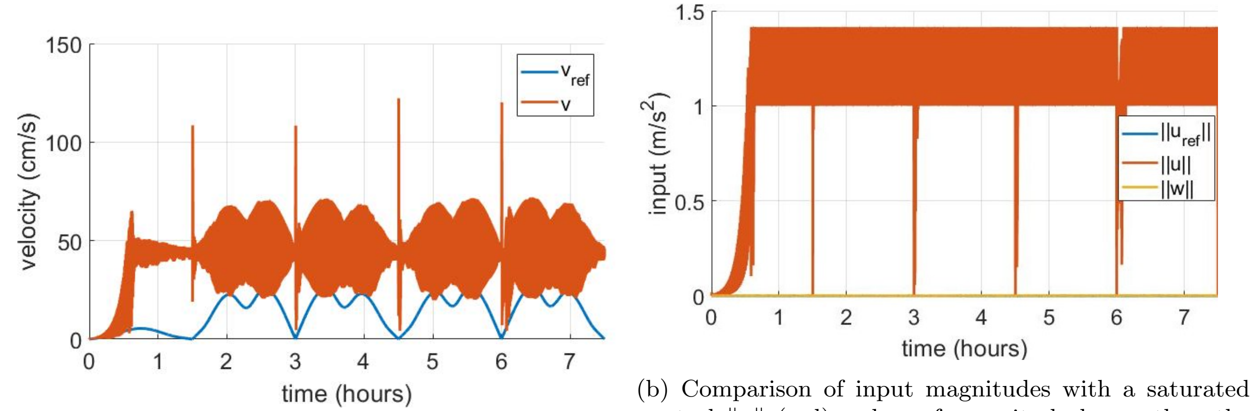
Figure 10.10: Analysis of the trajectory tracking performance for a stochastic Lipschitz undesirable input w and actuation delay $\tau = 8\, s$.

Concerning the fuel efficiency, the pseudo-equality $m \approx m_{\text{ref}} + m_w$ derived from Fig. 10.9(a) still holds approximately since $m_{\text{ref}} = 0.16\, kg$, $m_w = 1.06\, kg$ and $m_u = 1.38\, kg$ in this scenario. The controlled thrusters have only slightly increased their consumption compared to $m_u = 1.31\, kg$ for $\tau = 0.2\, s$. Thus, the actuation delay does not play as crucial a role for the fuel consumption as for the position error.

10.5.3 Lipschitz undesirable thrust and actuation delay of 10 seconds

If we increase further the actuation delay, e.g. $\tau = 10\, s$, controller (10.12) becomes incapable of tracking the reference trajectory as depicted on Fig. 10.11(a). The velocity on the tracking trajectory is on average four times larger than the reference.

The position error has also steeply increased compared to the scenario where $\tau = 8\, s$ since here the average position error is $0.48\, m$ and the maximal error is $3\, m$. These values are still small enough to keep the tracking trajectory indistinguishable from the reference on a figure like Fig. 10.7(a). However, to maintain



(a) Velocity comparison for the reference trajectory v_{ref} (blue) and the tracking trajectory v (red).

(b) Comparison of input magnitudes with a saturated control $\|u\|$ (red) orders of magnitude larger than the reference $\|u_{\text{ref}}\|$ (blue) and the undesirable input $\|w\|$ (yellow).

Figure 10.11: Analysis of the trajectory tracking performance for a stochastic Lipschitz undesirable input w and actuation delay $\tau = 10\,s$.

this accuracy, controller (10.12) had to saturate its thrust inputs as shown on Fig. 10.11(b). This input saturation results in a prohibitive fuel consumption of $503\,kg$ compared to $m_u = 1.38\,kg$ for $\tau = 8\,s$. Now that we have probed the limits of controller (10.12) in terms of actuation delay, let us investigate the impact of the regularity of w on the tracking performance.

10.5.4 Bang-bang undesirable thrust and actuation delay of 1 second

In this scenario we keep the actuation delay $\tau = 1\,s$, but the undesirable thrust signal w is now bang-bang, as illustrated on Fig. 10.13(a). This violates the Lipschitz assumption of Theorem 31 and hence invalidates its performance guarantees.

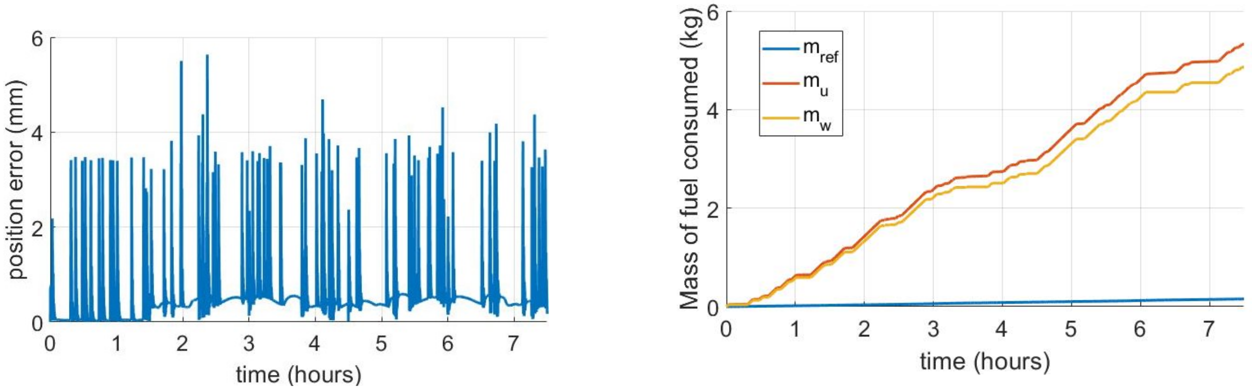
Controller (10.12) generates a trajectory with an average position error of $0.54\,mm$ and a maximal error of $5.6\,mm$ as shown on Fig. 10.12(a). These values are comparable to the precision achieved in the scenario where w was Lipschitz and $\tau = 8\,s$. As expected, increasing the unpredictability of w from Lipschitz to bang-bang led to a degradation of the tracking performance. Concerning the fuel usage in this scenario, Fig. 10.12(b) shows that the bang-bang thrust signal yields a significant consumption increase to $m_w = 4.86\,kg$ compared to $1.06\,kg$ in the Lipschitz scenarios. This increase is reflected on the controller's fuel usage $m_u = 5.33\,kg$, which remains close to $m_{\text{ref}} + m_w = 5.02\,kg$.

Every time the undesirable thrust climbs to its maximum value, the controller reacts after a delay τ and with a 50% higher spike to makeup for this delay, as illustrated on Fig. 10.13(a). This overshoot explains the increased mass of fuel consumption by the controlled thrusters. Note also the similarity between Fig. 10.12(a) and 10.13(a), each position error spike is associated with a spike of w.

Since controller (10.12) is still able to track the reference trajectory, we will consider a more challenging scenario with an increased actuation delay.

10.5.5 Bang-bang undesirable thrust and actuation delay of 8 seconds

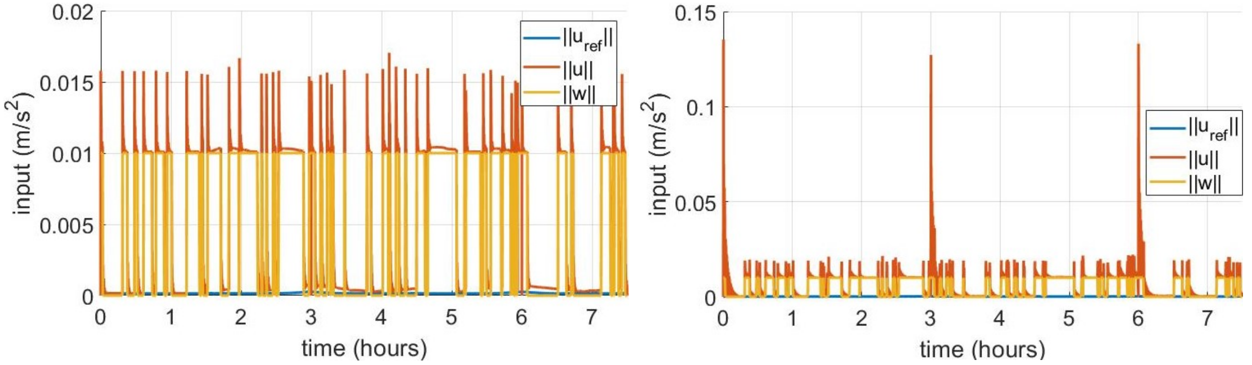
We increase the actuation delay to $\tau = 8\,s$ while keeping the same bang-bang undesirable thrust signal as in the previous scenario. The overshoots of the controller have become much larger at three waypoints as shown



(a) Position error between the reference and tracking trajectories.

(b) Comparisons of fuel used to complete the reference trajectory m_{ref} (blue), the tracking trajectory m_u (red), and the fuel used by malfunctioning thruster no. 4 m_w (yellow).

Figure 10.12: Analysis of the trajectory tracking performance for a bang-bang undesirable thrust signal w and actuation delay $\tau = 1\,s$.



(a) Actuation delay $\tau = 1\,s$.

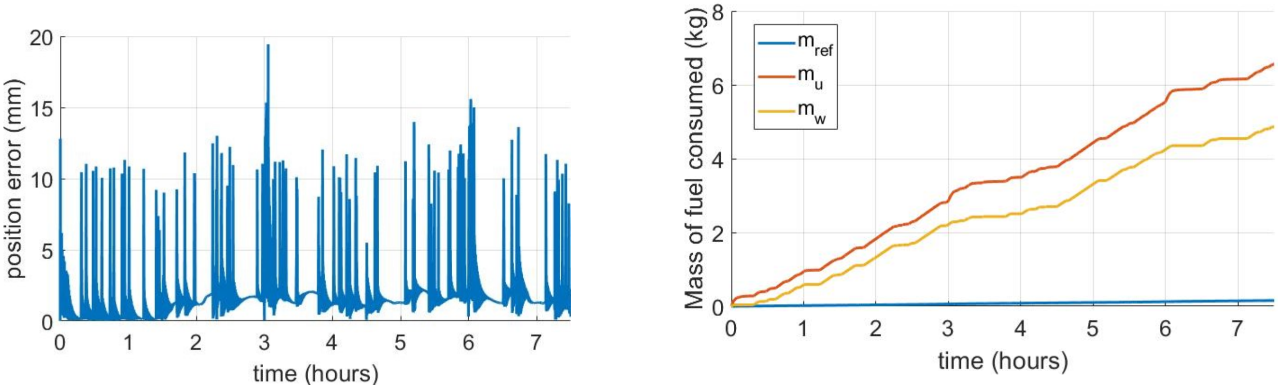
(b) Actuation delay $\tau = 8\,s$.

Figure 10.13: Magnitude of the thrust inputs for the reference trajectory $\|u_{\text{ref}}\|$ (blue), the tracking trajectory $\|u\|$ (red), and the bang-bang undesirable input $\|w\|$ (yellow) for different actuation delays.

on Fig. 10.13(b), while the overshoots at other locations have an amplitude similar to that of w. These large spikes are still an order of magnitude smaller than those of Fig. 10.11(b), so the controller is not saturating yet.

The average position error with respect to the reference trajectory is $1.76\,mm$ and the maximal error is $19.4\,mm$. These values represent approximately a fourfold increase compared to the scenario of Section 10.5.4. As in the Lipschitz cases where we also witnessed a fourfold increase between $\tau = 0.2\,s$ and $\tau = 8\,s$, the increased actuation delay has significant impact on the tracking accuracy.

The undesirable thrust still consumes $m_w = 4.86\,kg$ of fuel, but the controller now needs $m_u = 6.56\,kg$ according to Fig. 10.14(b) instead of $5.33\,kg$ for $\tau = 1\,s$. This consumption increase is most likely caused by the large thrust spikes of Fig. 10.13(b). As in the Lipschitz case, the increased actuation delay does not have a significant impact on the fuel consumption. However, if we increase τ to $10\,s$, then the situation is similar as that of Section 10.5.3 with a prohibitive increase in fuel consumption to keep the malfunctioning spacecraft close to the reference orbit.



(a) Position error between the reference and tracking trajectories.

(b) Comparisons of fuel used to complete the reference trajectory m_{ref} (blue), the tracking trajectory m_u (red), and the fuel used by malfunctioning thruster no. 4 m_w (yellow).

Figure 10.14: Analysis of the trajectory tracking performance for a bang-bang undesirable thrust signal w and actuation delay $\tau = 8\,s$.

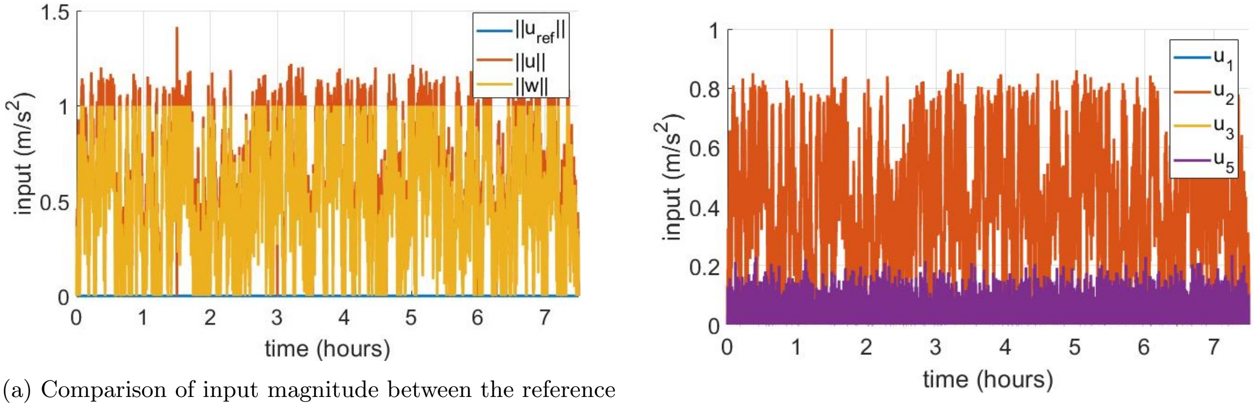
In all scenarios tested so far, the undesirable thrust signal was saturated at 1% of its capability to have $\|w\|$ of the same order of magnitude as $\|u_{\text{ref}}\|$ as depicted on Fig. 10.8(a). In the first scenario $\|u\|$ was also of the same order of magnitude. However, we see in this scenario that u sometimes needs to significantly overshoot w. Therefore, we must also investigate the scenario where w has access to its whole thrust capability, i.e., $w(t) \in [0, 1]$, to assess whether it can be counteracted by u despite its saturation limit.

10.5.6 Saturated Lipschitz undesirable thrust and actuation delay of 2 second

We will now investigate the case of a Lipschitz undesirable thrust input w where $L = 0.1$, $\max\{w(t),\, t \geq 0\} = 1$ and $\tau = 2\,s$. Since w makes use of its full range of thrust actuation, the controlled thrusters might reach their own saturation limit. The simulation results show the undesirable thrust signal meeting both its saturation limits, $w(t) \in [0, 1]$ as seen on Fig. 10.15(a). The controlled thrusters however, do not reach their own saturation since the individual magnitude of each thruster never reaches 1 on Fig. 10.15(b), except at 1 hour 30 minutes. This saturation can also be seen on Fig. 10.15(a) where $\|u\| = \sqrt{2}$.

Based on Fig. 10.15(b), we can see that thruster no. 2 is producing the thrust necessary to counteract w. Thruster no. 1 is actually matching u_2, just as in Fig. 10.8(b), except that we cannot see it on Fig. 10.15(b) because u_2 covers u_1. The average position error is contained to $48\,mm$, while the maximal position error is $0.29\,m$ as shown on Fig. 10.16(a). Then, the tracking trajectory stays sufficiently close to the reference to not be distinguishable on a figure like Fig. 10.7(a). The tracking velocity presents large fluctuations above the reference velocity as shown on Fig. 10.16b(b) while staying much closer than in the scenario of Section 10.5.3 where v was entirely above v_{ref} as seen on Fig. 10.11(a).

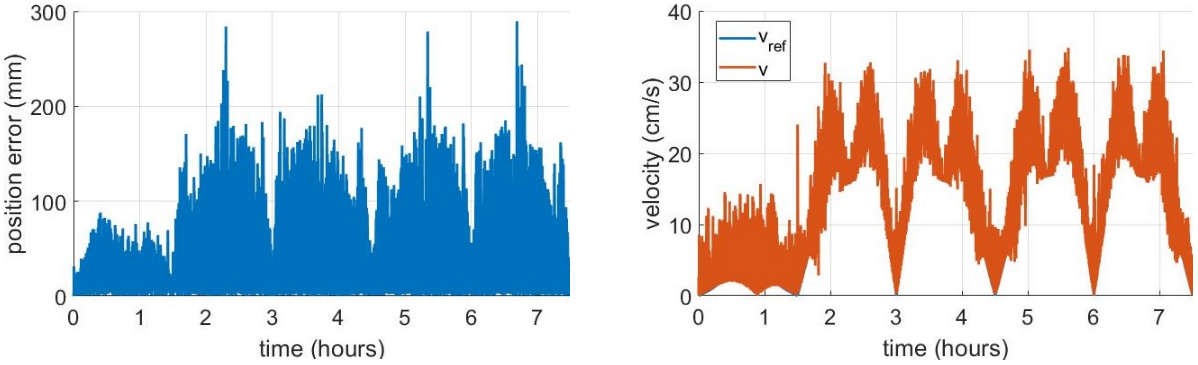
Because of the large thrusts employed in this scenario, the masses of fuel consumed have also significantly increased. The controlled thrusters would need $m_u = 360\,kg$ of fuel, while the malfunctioning thruster is guzzling $m_w = 342\,kg$ of fuel over the 7.5 hours of the mission. These masses are relatively close, within 5% of each other, which tells us that the controller is not wasting too much extra fuel in overshoots, it uses only what is needed to counteract w. However, recall that our spacecraft mass was set at $600\,kg$. Thus, if such a malfunction were to happen, the thrusters would run out of fuel before completing the mission. Nevertheless,



(a) Comparison of input magnitude between the reference $\|u_{\text{ref}}\|$ (blue), tracking control $\|u\|$ (red) and the saturated undesirable input $\|w\|$ (yellow).

(b) Thrust profiles for the four controlled thrusters of the chaser satellite on the tracking trajectory.

Figure 10.15: Analysis of the thrust inputs for a Lipschitz undesirable thrust signal w and actuation delay $\tau = 2\,s$.



(a) Position error between the reference and tracking trajectories.

(b) Velocity comparison for the reference trajectory v_{ref} (blue) and the tracking trajectory v (red).

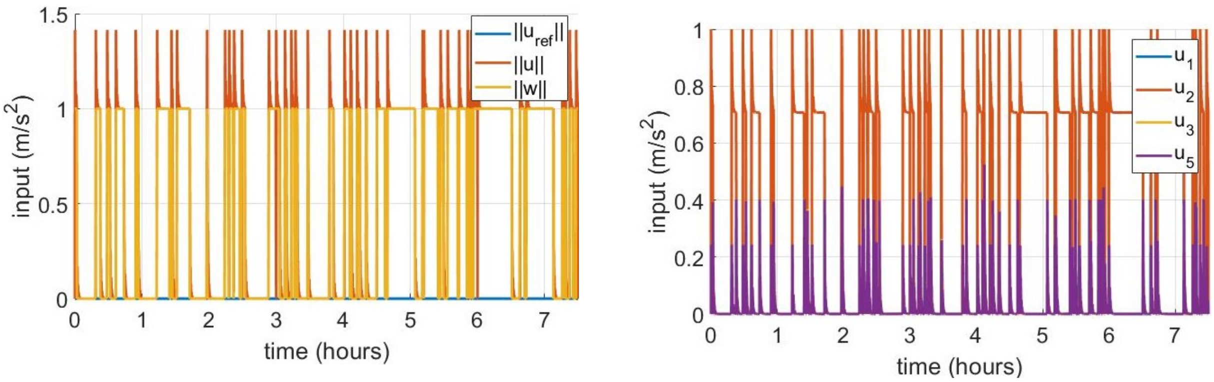
Figure 10.16: Analysis of the trajectory tracking performance for a Lipschitz undesirable thrust signal w and actuation delay $\tau = 2\,s$.

while fuel is available, we now know that controller (10.12) can compensate time-varying undesirable thrust of maximal amplitude.

With the same Lipschitz undesirable thrust signal, but an actuation delay $\tau = 3\,s$ instead of $2\,s$, the trajectory quickly diverge from the reference. This was somewhat predictable from the saturation of u_2 in Fig. 10.15(b). Let us now study how controller (10.12) would fare against a bang-bang undesirable input of similar magnitude.

10.5.7 Saturated bang-bang undesirable thrust and actuation delay of 1 second

In this scenario w is bang-bang in $[0, 1]$ and the actuation delay is $\tau = 1\,s$. The simulation shows clearly the bang-bang behavior of the undesirable thrust signal on Fig. 10.17(a). The controlled thrusters are also reaching their own saturation limit of 1 on Fig. 10.17(b), except at 1 hour 30 minutes. This saturation can also be seen on Fig. 10.15(a) where $\|u\| = \sqrt{2}$.



(a) Comparison of input magnitude between the reference $\|u_{\text{ref}}\|$ (blue), tracking control $\|u\|$ (red) and the bang-bang undesirable input $\|w\|$ (yellow).

(b) Thrust profiles for the four controlled thrusters of the chaser satellite on the tracking trajectory.

Figure 10.17: Analysis of the trajectory tracking performance for a bang-bang undesirable thrust signal w and actuation delay $\tau = 1\,s$.

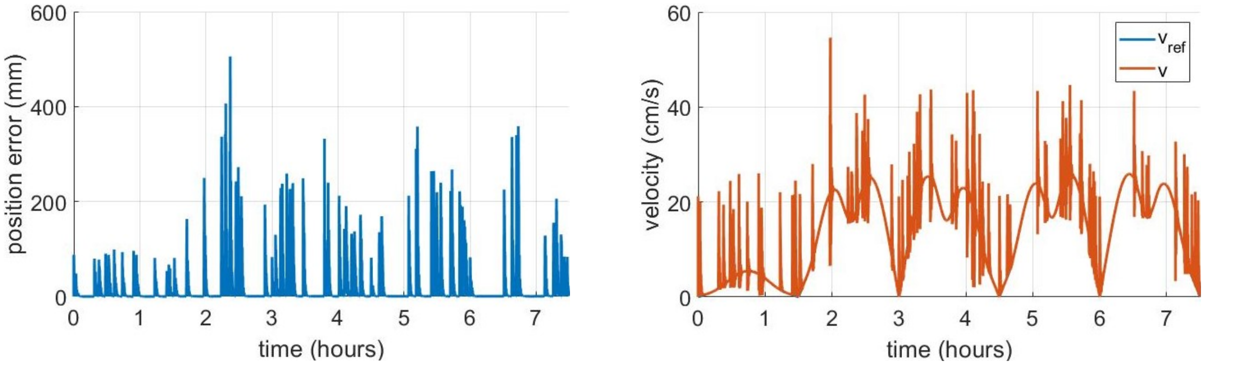
As shown on Fig. 10.18(a), the average position error is $17.1\,mm$ and the maximal position error is $0.5\,m$, so both trajectories are still indistinguishable on a figure like Fig. 10.7(a). We note the presence of a velocity spike on Fig. 10.18(b) for each spike of w on Fig. 10.17(a).

As in the previous scenario, the fuel consumption is too large for the mission to be completed with such a malfunctioning thruster, but while it is active it can be actively counteracted by u_1 and u_2 as shown on Fig. 10.17(b). The masses of fuel consumed by u and w are also relatively close, within 3% of each other, with $m_u = 345\,kg$ and $m_w = 336\,kg$, which relates to the efficiency of the controller.

If we further increase the actuation delay to $\tau = 2\,s$ for the same undesirable thrust w, the trajectory quickly diverge from the reference. Since the controlled inputs were already saturated for $\tau = 1\,s$ as seen on Fig. 10.17(b), the controller was not able to overcome a more unpredictable w and this divergence is not surprising.

10.5.8 Summary of the simulation scenarios

Let us now summarize and compare the scenarios studied. For each of these scenarios, we compute the average and maximal position errors, the mass of fuel used by the controlled thrusters m_u and by malfunctioning



(a) Position error between the reference and tracking trajectories.

(b) Velocity comparison for the reference trajectory v_{ref} (blue) and the tracking trajectory v (red).

Figure 10.18: Analysis of the trajectory tracking performance for a bang-bang undesirable thrust signal w and actuation delay $\tau = 1\,s$.

thruster no. 4 m_w. We also calculate the relative difference of fuel used $\frac{m_u - m_w - m_{\text{ref}}}{m_w + m_{\text{ref}}}$, which is a good metric for the efficiency of controller (10.12) in overcoming w without excessive thrust. We summarize these metrics in Table 10.1.

Scenario	Regularity of w	Actuation delay τ	Saturation of w	Average position error	Maximal position error	Controlled fuel used m_u	Undesirable fuel used m_w	Relative difference of fuel used
1	Lipschitz	$0.2\,s$	0.01	$0.36\,mm$	$1.05\,mm$	$1.31\,kg$	$1.06\,kg$	7.4%
2	Lipschitz	$8\,s$	0.01	$1.2\,mm$	$4.1\,mm$	$1.38\,kg$	$1.06\,kg$	13.1%
3	Lipschitz	$10\,s$	0.01	$484\,mm$	$3 \times 10^3\,mm$	$503\,kg$	$1.06\,kg$	41×10^3%
4	bang-bang	$1\,s$	0.01	$0.54\,mm$	$5.6\,mm$	$5.33\,kg$	$4.86\,kg$	6.2%
5	bang-bang	$8\,s$	0.01	$1.76\,mm$	$19.4\,mm$	$6.56\,kg$	$4.86\,kg$	31%
6	Lipschitz	$2\,s$	1	$48\,mm$	$292\,mm$	$360\,kg$	$342\,kg$	5.2%
7	bang-bang	$1\,s$	1	$17.1\,mm$	$509\,mm$	$345\,kg$	$336\,kg$	2.6%

Table 10.1: Summary table of the simulation scenarios.

As expected, when actuation delay τ increases, so do the position error and the relative difference of fuel used. Indeed, the controller has more difficulty counteracting w and needs to use more corrective thrust when τ is larger.

The hundredfold increase in position error and m_u between scenarios 2 and 3 when τ only increased from $8\,s$ to $10\,s$ shows that controller (10.12) has reached its limit. Indeed, the trajectory diverges when $\tau > 10\,s$.

In scenarios 4, 5 and 7, the undesirable thrust signal w is bang-bang, as illustrated on Tables 10.2 and 10.3. A bang-bang undesirable input violates the Lipschitz assumption of Theorem 31 and hence invalidates its performance guarantees. Nonetheless, controller (10.12) generates a tracking trajectory with position errors of the same order of magnitude as for Lipschitz continuous w.

When comparing scenarios 2 and 5, we see that increasing the unpredictability of w from Lipschitz to bang-bang led to a degradation of the tracking performance.

In scenarios 1 to 5, the undesirable thrust signal was saturated at 1% of its capability to have $\|w\|$ of the same order of magnitude as $\|u_{\text{ref}}\|$ as depicted on Fig. 10.8(a). In the nominal scenario $\|u\|$ was also of the same order of magnitude as shown on Fig. 10.8(a). In scenarios 6 and 7, w has access to its whole thrust capability, i.e., $w(t) \in [0,1]$, to assess whether it can be counteracted by u despite its saturation limit.

Because of the large thrusts employed in scenarios 6 and 7, the masses of fuel consumed have also significantly increased. The controlled thrusters would need $m_u = 360\,kg$ of fuel, while the malfunctioning thruster is guzzling $m_w = 342\,kg$ of fuel over the 7.5 hours of the mission. These masses are relatively close, within 5% of each other, which tells us that the controller is not wasting too much extra fuel in overshoots, it uses only what is needed to counteract w. However, recall that our spacecraft mass was set at $600\,kg$. Thus, if such a malfunction were to happen, the thrusters would run out of fuel before completing the mission. Nevertheless, while fuel is available, we now know that controller (10.12) can compensate time-varying undesirable thrust of maximal amplitude.

If we increase the actuation delay of scenarios 6 and 7 by $1\,s$ each, their trajectory quickly diverge from the reference. This was somewhat predictable from the saturation of the controlled input magnitude $\|u\|$ in Table 10.3.

The average position errors listed in Table 10.1 are small enough to keep the tracking trajectory indistinguishable from the reference on a figure like Fig. 10.7(a). Then, we focus on the more informative graphs of the position error associated with the corresponding thrust profiles for each scenarios grouped together in Tables 10.2 and 10.3.

Let us now summarize the findings of the various scenarios studied. Despite the narrow range of application of Theorem 31, controller (10.12) provides tracking accuracy to the millimeter scale on a much wider range of scenarios than expected with fast-varying undesirable inputs and longer actuation delays. In all the scenarios, the magnitude of the controlled thrusters $\|u\|$ had to be larger, if not significantly larger than $\|w\|$ to counteract its nefarious influence. This is problematic when w reaches its maximal amplitude as u is more likely to saturate. However, scenarios 6 and 7 showed that for small actuation delays, w can still be counteracted. We can visually summarize the performance of controller (10.12) with the Pareto front of Fig. 10.19 on the saturation limit of w and the actuation delay τ. Based on the scenarios investigated in this section, we decided to consider the tracking successful when the position error is smaller than $0.8\,m$, which is 1% of the minimal target distance on the reference trajectory.

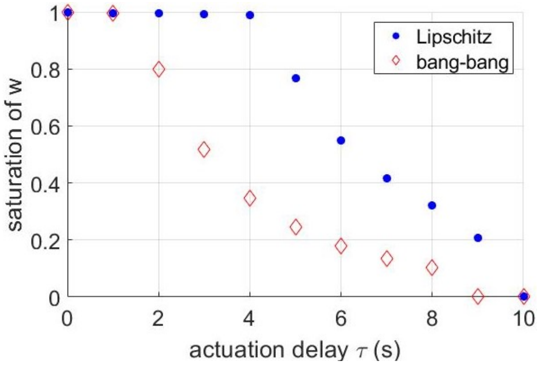


Figure 10.19: Pareto front of the maximal saturation limit of w for which controller (10.12) maintains a position error under $0.8\,m$ despite actuation delay τ.

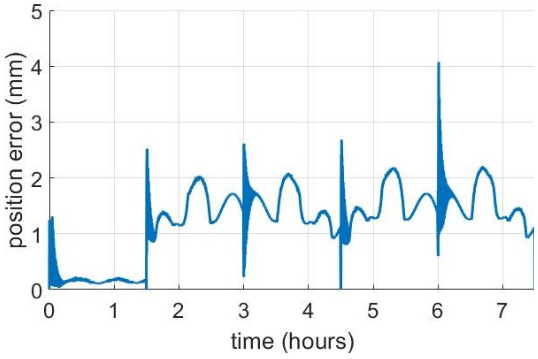
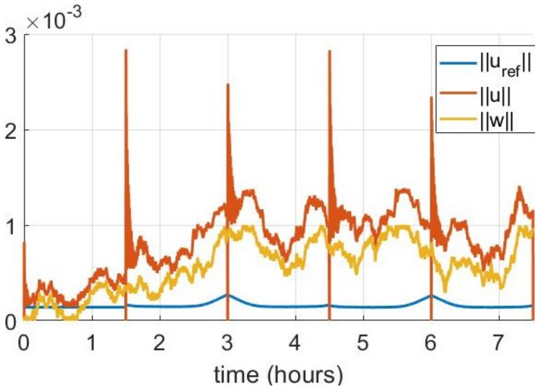
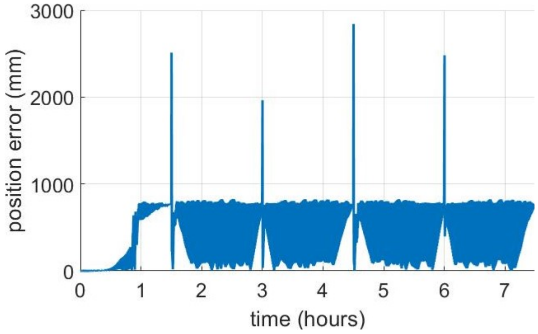
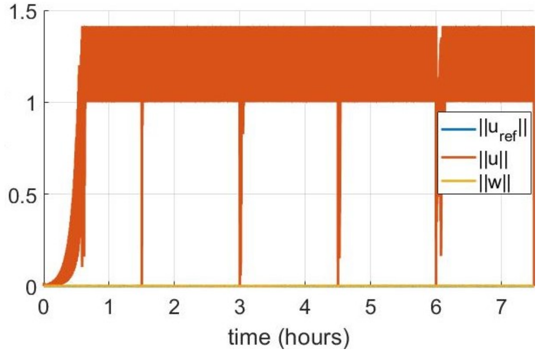
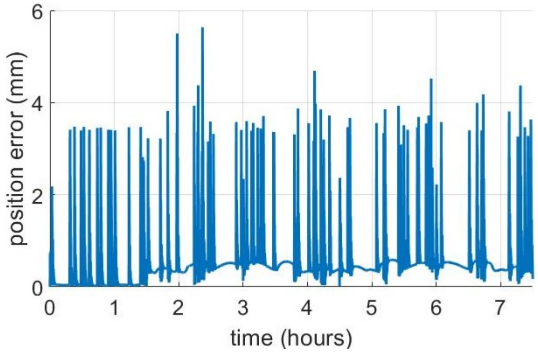
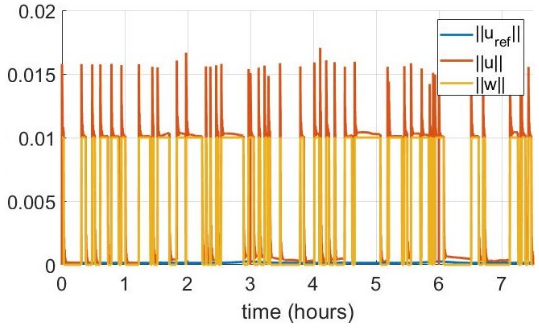
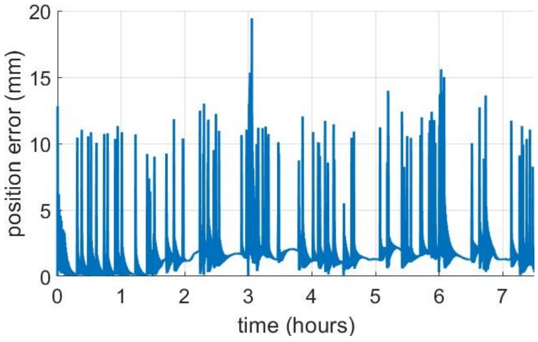
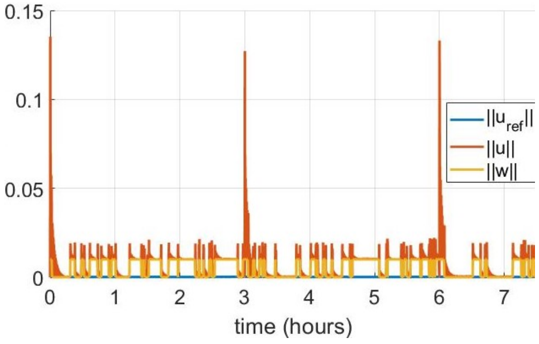
Scenario	Position error	Input magnitudes
2		
3		
4		
5		

Table 10.2: Position error and input magnitude profiles across scenarios 2 to 5.

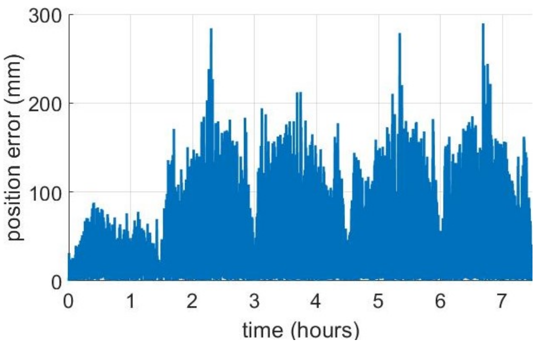
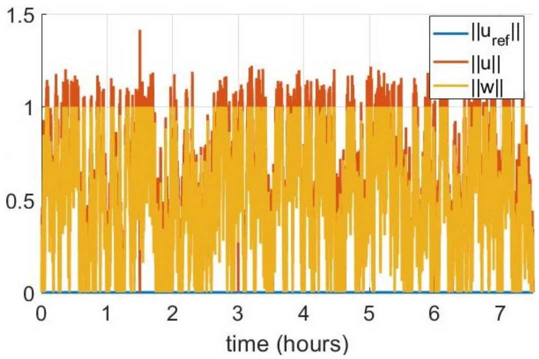
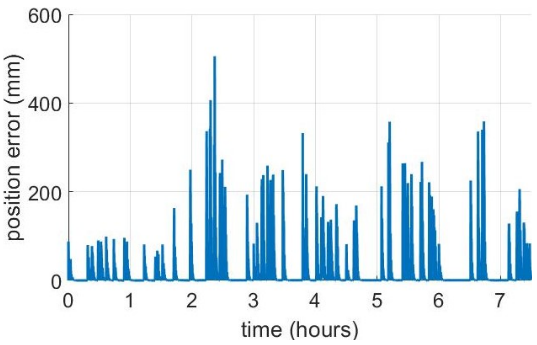
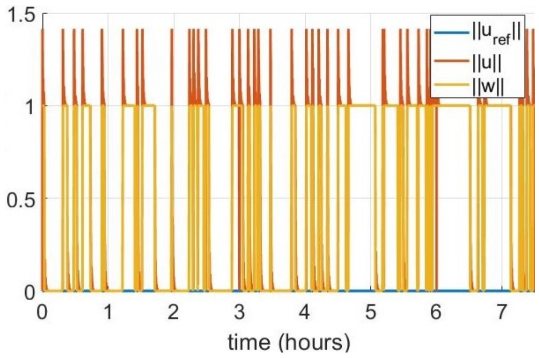
Scenario	Position error	Input magnitudes
6		
7		

Table 10.3: Position error and input magnitude profiles in scenarios 6 and 7.

10.6 Summary

In this chapter we presented a new methodology to safely perform a satellite inspection mission despite actuation delay and the loss of control authority over a thruster. We established theoretical trajectory tracking guarantees on a resilient controller embedded with a state predictor to compensate for the actuation delay. We tested this controller on a variety of scenarios increasingly adversarial to determine the capabilities of our controller. We concluded that it enables a resilient tracking of the reference trajectory and a safe completion of the inspection mission.

Chapter 11

Conclusions and Future Work

We have now reached the final chapter of this dissertation. In Section 11.1 we will summarize the work accomplished so far, before concluding on the success of our approach to address our problems of interest in Section 11.2. Finally, we will highlight several interesting avenues for future work in Section 11.3.

11.1 Summary

In chapter 2 we reviewed the literature related to resilience theory. We studied how resilience fits within the wider approaches of robust, adaptive and fault-tolerant control. We also compared our theory with other notions of resilience found in the literature. Finally, we introduced previous works studying reachability, controllability, differential games theory, and time optimal linear control upon which resilience theory is built.

In chapter 3 we took our first shot at Problem 1 relying on our earliest work on resilience [31]. We established the foundational resilience theory for linear systems with bounded energy. We built on the highly abstract perturbed reachability condition of [24] to derive simple analytical conditions for resilient reachability of driftless linear systems and to understand how reachability evolves with time.

In chapter 4 we relied on our work [35] to investigate the design of resilient driftless linear systems with bounded energy and to solve Problem 2. We built on the resilient reachability condition of Chapter 3 to calculate the minimal degree of overactuation necessary for a system to be resilient to the loss of control over any single one of its actuators. We also synthesized a control law achieving resilient reachability for linear systems.

In chapter 5 we switched gear to study the resilience of linear systems with component bounded inputs and we introduced the notion of quantitative resilience. Relying on our works [32], [38], we used linear optimal control to design an efficient method to calculate the quantitative resilience of driftless linear systems. This chapter's objective was then to address Problem 3.

In chapter 6, we described the proof of the Maximax Minimax Quotient Theorem. This optimization result was used in Chapter 5 to calculate the quantitative resilience of driftless systems. This proof was published in our work [36] and relied on a geometrical approach of input selection. We also proved the existence of a solution to this optimization problem with the Berge Maximum theorem [1].

In chapter 7 we extended resilience theory to general linear systems with drift by addressing Problems 1 and 3. We drew from our works [33], [37] and relied on differential games and linear control theories to establish necessary and sufficient conditions for the resilience of general linear systems. We also calculated analytical bounds on the quantitative resilience of these systems.

In chapter 8 based on our work [39], we took our first step towards Problem 4. We started by investigating more complex mission scenarios than resilient reachability by deriving sufficient conditions for resilient trajectory tracking. Then, we extended resilience theory to linear systems with actuation delays to remove the assumption of instantaneous knowledge of the undesirable inputs by the controller. Finally, we derived a sufficient resilience condition for systems with nonlinear dynamics.

In chapter 9 we extended resilience analysis to linear networks suffering partial loss of control authority. We mostly studied how an unresilient subsystem suffering from a partial loss of control authority can affect the stabilizability of the rest of the network. This chapter also contributed to address Problem 4 by extending the framework of resilience theory to linear networks.

In chapter 10 we investigated the resilience of an orbital inspection mission to the loss of control authority over a thruster of the inspecting spacecraft. This chapter drew from our works [34], [39] and contributed to Problem 4 by extending resilience theory to the nonlinear dynamics of a spacecraft. For these nonlinear dynamics we also built a resilient trajectory tracking controller with guaranteed performance.

11.2 Conclusion

We set out to address four central problems of interest stated in Section 1.2. Let us now assess whether we successfully solved these problems.

Problem 1 wondered about the conditions for a target set to be resiliently reachable. In Chapters 3 and 4, we established necessary and sufficient conditions for targets to be resiliently reachable by linear systems with energy bounded inputs thanks to the perturbed reachability theory of [24]. In Chapters 5, 7 and 8, we established resilient reachability conditions for systems with amplitude bounded inputs and respectively driftless, linear and nonlinear dynamics thanks to the differential games theory of [23].

Problem 2 asked how to design systems that are resilient to a loss of control authority over any one of their actuators. We answered this question in Chapter 4 by proving that at least $2n+1$ actuators are required for a n-dimensional driftless system to be resilient. We also poked at the design problem of resilience to the loss of control over any two actuators, but this problem is considerably more complex for the reasons detailed in Chapter 4.

Problem 3 shed light on the difficult optimization problem prompted by the calculation of quantitative resilience. In Chapter 5 we established a method to calculate analytically the quantitative resilience of driftless linear systems thanks to the Maximax Minimax Quotient Theorem of Chapter 6. Since an exact calculation of quantitative resilience is impossible for general linear systems, Chapter 7 instead established analytical bounds on their nominal and malfunctioning reach times T_N^* and T_M^* in order to bound their quantitative resilience.

Problem 4 asked the wider question of how to extend the scope of resilience theory. In Chapter 8 we investigated resilient trajectory tracking, added actuation delay to the controller and established some simple resilience conditions for nonlinear dynamics. In Chapter 9, we further extended the scope of our theory by studying the resilience of linear networks. Finally, in Chapter 10 we showed that resilience theory can be applied to nonlinear spacecraft dynamics.

On the whole, we thoroughly solved Problems 1, 2 and 3, while establishing major milestones for the much broader Problem 4. Indeed, this last problem of interest cannot be answered thoroughly and prompted numerous avenues for future work that we will now discuss.

11.3 Future work directions

Following the work accomplished in this dissertation, we now give recommendations to be considered for future research on the topic of resilience of control systems to partial loss of control authority over actuators.

Concerning the quantitative resilience of driftless linear systems, the main results of Chapter 5, namely Theorems 11 and 12 only apply to the loss of control over a single actuator. This limitation comes in fact from the Maximax Minimax Quotient Theorem of Chapter 6 where one of the optimization sets must be unidimensional. Future work should then investigate how to generalize the Maximax Minimax Quotient Theorem to enable the calculation of quantitative resilience of systems enduring a simultaneous loss of multiple actuators.

We also want to extend our notion of resilience from the system's state to its output. This would allow to assess the resilience of some systems with respect to some of their states instead of considering all the states together. For instance, after a quadcopter loses control over one of its propeller, it becomes underactuated and cannot resiliently reach arbitrary states of position and orientation. However, if we relinquish control over its yaw angle, such a quadcopter can in fact resiliently reach arbitrary positions, pitch, and roll angles according to the work [8]. With an output gathering these crucial states, this quadcopter would then be output resilient.

Adding an actuation delay to the controller as in Chapter 8 allowed to make our resilience framework more realistic. The next step in this direction would be to consider the effects of measurements and process disturbances. By adding some noise to the measure of $w(t)$ transmitted to the controller, we could evaluate the robustness of resilient controllers with respect to their knowledge of the undesirable inputs. Additionally, if the controller does not know exactly the dynamics of the system, it might not be able to accurately thwart the undesirable inputs. Thus, we should also investigate the robustness of resilient controllers with respect to their knowledge of the system dynamics.

Finally, the major and most complex avenue for future work is to further extend resilience theory to nonlinear systems. Indeed, we only established a sufficient condition for resilience in Chapter 8. Once a complementary necessary condition is found, resilience theory will be able to study more realistic systems with more accurate dynamics. For instance, the dynamics of the octocopter of Chapter 5 would conserve all their nonlinear complexity and coupling between position and rotations. In Chapter 9, we would be able to study more realistic networks, especially power networks. Finally, nonlinear resilience theory would allow to extend the approach of Chapter 10 to spacecraft dynamics combining position and attitude for a more realistic treatment.

A first step towards nonlinear resilience theory would be to verify our intuition concerning whether the reverse implication of Theorem 26 holds. To do so, one could select a simple nonlinear system and find optimal strategies for u and w either by hand or using machine learning. If this approach leads to a counterexample it would avoid wasted theoretical efforts. Otherwise, it would provide a numerical example for this novel nonlinear resilience theory.

References

[1] C. Aliprantis and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*. New York: Springer, 2006. DOI: 10.1007/3-540-29587-9.

[2] J. Slay and M. Miller, "Lessons learned from the Maroochy water breach," in *Critical Infrastructure Protection*, Springer, 2008, pp. 73–82. DOI: 10.1007/978-0-387-75462-8_6.

[3] S. Pushpak, A. Diwadkar, and U. Vaidya, "Vulnerability analysis of dynamical power networks to stochastic link failure attacks," in *18th International Conference on Hybrid Systems: Computation and Control*, 2015, pp. 219–226. DOI: 10.1145/2728606.2728614.

[4] J.-B. Bouvier, S. P. Nandanoori, M. Ornik, and S. Kundu, "Distributed transient safety verification via robust control invariant sets: A microgrid application," in *2022 American Control Conference*, 2022, pp. 2202–2207. DOI: 10.23919/ACC53348.2022.9867323.

[5] A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proceedings of the 3rd conference on Hot topics in security*, USENIX Association, 2008. DOI: 10.5555/1496671.1496677.

[6] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014. DOI: 10.1109/TAC.2014.2303233.

[7] C. Badue, R. Guidolini, R. V. Carneiro, *et al.*, "Self-driving cars: A survey," *Expert Systems with Applications*, vol. 165, p. 113 816, 2021. DOI: 10.1016/j.eswa.2020.113816.

[8] A. Freddi, A. Lanzon, and S. Longhi, "A feedback linearization approach to fault tolerance in quadrotor vehicles," *IFAC proceedings volumes*, vol. 44, no. 1, pp. 5413–5418, 2011. DOI: 10.3182/20110828-6-IT-1002.02016.

[9] C. Belcastro, J. Foster, G. Shah, *et al.*, "Aircraft loss of control problem analysis and research toward a holistic solution," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 4, pp. 733–775, 2017. DOI: 10.2514/1.G002815.

[10] Y. Yu, H. Wang, and N. Li, "Fault-tolerant control for over-actuated hypersonic reentry vehicle subject to multiple disturbances and actuator faults," *Aerospace Science and Technology*, vol. 87, pp. 230–243, 2019. DOI: 10.1016/j.ast.2019.02.024.

[11] J. Davis, J. Mayberry, and J. Penn, "On-orbit servicing: Inspection repair refuel upgrade and assembly of satellites in space," *The Aerospace Corporation, report*, 2019.

[12] M. Vavrina, E. Skelton, K. DeWeese, B. Naasz, D. Gaylor, and C. D'souza, "Safe rendezvous trajectory design for the Restore-L mission," in *29th AAS/AIAA Space Flight Mechanics Meeting*, 2019, pp. 3649–3668.

[13] M. Bartels, "Russia says 'software failure' caused thruster misfire at space station," *space.com*, 2021. [Online]. Available: https://www.space.com/space-station-nauka-arrival-thruster-fire-update.

[14] X. Yu and J. Jiang, "Hybrid fault-tolerant flight control system design against partial actuator failures," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 4, pp. 871–886, 2011. DOI: 10.1109/TCST.2011.2159606.

[15] U. Vaidya and M. Fardad, "On optimal sensor placement for mitigation of vulnerabilities to cyber attacks in large-scale networks," in *2013 European Control Conference*, 2013, pp. 3548–3553. DOI: 10.23919/ECC.2013.6669723.

[16] J. Davidson, F. Lallman, and T. Bundick, "Real-time adaptive control allocation applied to a high performance aircraft," in *5th SIAM Conference on Control and Its Applications*, 2001. [Online]. Available: https://dl.acm.org/doi/book/10.5555/887951.

[17] A. A. Amin and K. M. Hasan, "A review of fault tolerant control systems: Advancements and applications," *Measurement*, vol. 143, pp. 58–68, 2019. DOI: 10.1016/j.measurement.2019.04.083.

[18] B. Xiao, Q. Hu, and P. Shi, "Attitude stabilization of spacecrafts under actuator saturation and partial loss of control effectiveness," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2251–2263, 2013. DOI: 10.1109/TCST.2012.2236327.

[19] R. Suich and R. Patterson, "How much redundancy: Some cost considerations, including examples for spacecraft systems," NASA Technical Memorandum 103197, Lewis Research Center, Cleveland, Ohio, Tech. Rep., 1990. [Online]. Available: https://www.osti.gov/biblio/5894363 (visited on 03/12/2021).

[20] K. Zhou and J. C. Doyle, *Essentials of robust control*. Prentice Hall, 1998.

[21] B. Anderson and A. Dehghani, "Challenges of adaptive control–past, permanent and future," *Annual Reviews in Control*, vol. 32, pp. 123–135, 2008. DOI: 10.1016/j.arcontrol.2008.06.001.

[22] M. Breitner, "The genesis of differential games in light of Isaacs' contributions," *Journal of Optimization Theory and Applications*, vol. 124, no. 3, pp. 523–559, 2005. DOI: 10.1007/s10957-004-1173-0.

[23] O. Hájek, "Duality for differential games and optimal control," *Mathematical Systems Theory*, vol. 8, no. 1, pp. 1–7, 1974. DOI: 10.1007/BF01761702.

[24] M. Delfour and S. Mitter, "Reachability of perturbed systems and min sup problems," *SIAM Journal on Control and Optimization*, vol. 7, no. 4, pp. 521–533, 1969. DOI: 10.1137/0307038.

[25] J.-P. Richard, "Time-delay systems: An overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, 2003. DOI: 10.1016/S0005-1098(03)00167-5.

[26] R. J. Patton, F. J. Uppal, S. Simani, and B. Polle, "Robust fdi applied to thruster faults of a satellite system," *Control Engineering Practice*, vol. 18, no. 9, pp. 1093–1109, 2010. DOI: 10.1016/j.conengprac.2009.04.011.

[27] D. Henry, "Fault diagnosis of Microscope satellite thrusters using H_∞/H_- filters," *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 3, pp. 699–711, 2008. DOI: 10.2514/1.31003.

[28] H. Romero, S. Salazar, A. Sanchez, and R. Lozano, “A new UAV configuration having eight rotors: Dynamical model and real-time control,” in *46th IEEE Conference on Decision and Control*, 2007, pp. 6418–6423. DOI: 10.1109/CDC.2007.4434776.

[29] M. Athans, “The status of optimal control theory and applications for deterministic systems,” *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 580–596, 1966. DOI: 10.1109/TAC.1966.1098353.

[30] Y. Sakawa, “Solution of linear pursuit-evasion games,” *SIAM Journal on Control*, vol. 8, no. 1, pp. 100–112, 1970. DOI: 10.1137/0308006.

[31] J.-B. Bouvier and M. Ornik, “Resilient reachability for linear systems,” in *21st IFAC World Congress*, 2020, pp. 4409–4414. DOI: 10.1016/j.ifacol.2020.12.372.

[32] J.-B. Bouvier, K. Xu, and M. Ornik, “Quantitative resilience of linear driftless systems,” in *SIAM Conference on Control and its Applications*, 2021, pp. 32–39. DOI: 10.1137/1.9781611976847.5.

[33] J.-B. Bouvier and M. Ornik, “Quantitative resilience of linear systems,” in *20th European Control Conference*, 2022, pp. 485–490. DOI: 10.23919/ECC55457.2022.9838147.

[34] J.-B. Bouvier, H. Panag, R. Woollands, and M. Ornik, “Resilience of orbital inspections to partial loss of control authority over the chaser satellite,” in *2022 AAS/AIAA Astrodynamics Specialist Conference*, 2022. [Online]. Available: https://mornik.web.illinois.edu/wp-content/uploads/BPWO22.pdf.

[35] J.-B. Bouvier and M. Ornik, “Designing resilient linear systems,” *IEEE Transactions on Automatic Control*, vol. 67, no. 9, pp. 4832–4837, 2022. DOI: 10.1109/TAC.2022.3163242.

[36] ——, “The maximax minimax quotient theorem,” *Journal of Optimization Theory and Applications*, vol. 192, pp. 1084–1101, 2022. DOI: 10.1007/s10957-022-02008-z.

[37] ——, “Resilience of linear systems to partial loss of control authority,” *Automatica*, p. 110 985, 2023. DOI: 10.1016/j.automatica.2023.110985.

[38] J.-B. Bouvier, K. Xu, and M. Ornik, “Quantitative resilience of generalized integrators,” *IEEE Transactions on Automatic Control*, 2023. [Online]. Available: https://arxiv.org/abs/2111.04163.

[39] J.-B. Bouvier, H. Panag, R. Woollands, and M. Ornik, “Resilient trajectory tracking to partial loss of control authority over actuators with actuation delay,” *not submitted yet*, 2023.

[40] J.-B. Bouvier and M. Ornik, “Resilience of linear networks,” *not submitted yet*, 2023.

[41] D. Bertsekas, “Infinite-time reachability of state-space regions by using feedback control,” *IEEE Transactions on Automatic Control*, vol. 17, no. 5, pp. 604–612, 1972. DOI: 10.1109/TAC.1972.1100085.

[42] A. Kurzhanski and P. Varaiya, “Reachability analysis for uncertain systems-the ellipsoidal technique,” *Dynamics of Continuous Discrete and Impulsive Systems Series B*, vol. 9, pp. 347–368, 2002.

[43] D. Bertsekas and I. Rhodes, “On the minimax reachability of target sets and target tubes,” *Automatica*, vol. 7, pp. 233–247, 1971. DOI: 10.1016/0005-1098(71)90066-5.

[44] L. Y. Wang and J.-F. Zhang, “Fundamental limitations and differences of robust and adaptive control,” in *2001 American Control Conference*, vol. 6, 2001, pp. 4802–4807. DOI: 10.1109/ACC.2001.945742.

[45] X. Tang, G. Tao, and S. Joshi, “Adaptive actuator failure compensation for nonlinear MIMO systems with an aircraft control application,” *Automatica*, vol. 43, pp. 1869–1883, 2007. DOI: 10.1016/j.automatica.2007.03.019.

[46] W. Wang and C. Wen, "Adaptive actuator failure compensation control of uncertain nonlinear systems with guaranteed transient performance," *Automatica*, vol. 46, pp. 2082–2091, 2010. DOI: 10.1016/j.automatica.2010.09.006.

[47] S. S. Tohidi, Y. Yildiz, and I. Kolmanovsky, "Fault tolerant control for over-actuated systems: An adaptive correction approach," in *2016 American Control Conference*, 2016, pp. 2530–2535. DOI: 10.1109/ACC.2016.7525297.

[48] G. Tao, S. Chen, and S. Joshi, "An adaptive actuator failure compensation controller using output feedback," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 506–511, 2002. DOI: 10.1109/9.989150.

[49] C. L. Lewis and A. A. Maciejewski, "Fault tolerant operation of kinematically redundant manipulators for locked joint failures," *IEEE Transactions on Robotics and Automation*, vol. 13, no. 4, pp. 622–629, 1997. DOI: 10.1109/70.611335.

[50] A. Aitouche and B. O. Bouamama, "Fault tolerant control with respect to actuator failures, application to steam generator process," *European Symposium on Computer Aided Process Engineering*, pp. 1471–1476, 2005. DOI: 10.1016/S1570-7946(05)80087-2.

[51] D. A. Marsillach, S. Di Cairano, and A. Weiss, "Abort-safe spacecraft rendezvous in case of partial thrust failure," in *59th Conference on Decision and Control*, 2020, pp. 1490–1495. DOI: 10.1109/CDC42340.2020.9303782.

[52] L. Breger and J. How, "Safe trajectories for autonomous rendezvous of spacecraft," *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 5, pp. 1478–1489, 2008. DOI: 10.2514/1.29590.

[53] M. Bucić, M. Ornik, and U. Topcu, "Graph-based controller synthesis for safety-constrained, resilient systems," in *56th Annual Allerton Conference on Communication, Control, and Computing*, 2018, pp. 297–304. DOI: 10.1109/ALLERTON.2018.8635905.

[54] W. Schmitendorf and B. Elenbogen, "Constrained max-min controllability," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 731–733, 1982. DOI: 10.1109/TAC.1982.1102966.

[55] S. Sinha, S. P. Nandanoori, T. Ramachandran, C. Bakker, and A. Singhal, "Data-driven resilience characterization of control dynamical systems," in *2022 American Control Conference*, 2022, pp. 2186–2193. DOI: 10.23919/ACC53348.2022.9867785.

[56] S. Shin, S. Lee, D. Judi, *et al.*, "A systematic review of quantitative resilience measures for water infrastructure systems," *Water*, vol. 10, no. 2, pp. 164–189, 2018. DOI: 10.3390/w10020164.

[57] J. T. Kim, J. Park, J. Kim, and P. H. Seong, "Development of a quantitative resilience model for nuclear power plants," *Annals of Nuclear Energy*, vol. 122, pp. 175–184, 2018. DOI: 10.1016/j.anucene.2018.08.042.

[58] D. Henry and J. E. Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering & System Safety*, vol. 99, pp. 114–122, 2012. DOI: 10.1016/j.ress.2011.09.002.

[59] R. Brockett, "Nonlinear systems and differential geometry," *Proceedings of the IEEE*, vol. 64, no. 1, pp. 61–72, 1976. DOI: 10.1109/PROC.1976.10067.

[60] A. Isidori, *Nonlinear Control Systems, An Introduction*. Springer Verlag, 1989. DOI: 10.1007/BFb0006368.

[61] A. Girard, C. Le Guernic, and O. Maler, "Efficient computation of reachable sets of linear time-invariant systems with inputs," in *International Workshop on Hybrid Systems: Computation and Control*, Springer, 2006, pp. 257–271. DOI: 10.1007/11730637_21.

[62] M. Althoff, N. Kochdumper, and M. Wetzlinger, "CORA 2020 manual," *TU Munich*, 2020. [Online]. Available: https://tumcps.github.io/CORA/data/Cora2020Manual.pdf.

[63] A. Girard and C. L. Guernic, "Efficient reachability analysis for linear systems using support functions," in *17th IFAC World Congress*, 2008, pp. 8966–8971. DOI: 10.3182/20080706-5-KR-1001.01514.

[64] A. Kurzhanski and P. Varaiya, "Ellipsoidal techniques for reachability analysis: Internal approximation," in *Systems & Control Letters*, 2000, pp. 201–211. DOI: 10.1016/S0167-6911(00)00059-1.

[65] H. N. V. Pico and D. Aliprantis, "Reachability analysis of linear dynamic systems with constant, arbitrary, and Lipschitz continuous inputs," *Automatica*, vol. 95, pp. 293–305, 2018. DOI: 10.1016/j.automatica.2018.05.026.

[66] S. Raković, E. Kerrigan, D. Mayne, and J. Lygeros, "Reachability analysis of discrete-time systems with disturbances," *IEEE Transactions on Automatic Control*, vol. 51, no. 4, pp. 546–561, 2006. DOI: 10.1109/TAC.2006.872835.

[67] J. LaSalle, "Time optimal control systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 45, no. 4, pp. 573–577, 1959. DOI: 10.1073/pnas.45.4.573.

[68] E. Rechtschaffen, "Unique winning policies for linear differential pursuit games," *Journal of Optimization Theory and Applications*, vol. 29, no. 4, pp. 629–658, 1979. DOI: 10.1007/BF00934455.

[69] L. Neustadt, "Synthesizing time optimal control systems," *Journal of Mathematical Analysis and Applications*, vol. 1, no. 4, pp. 484–493, 1960. DOI: 10.1016/0022-247X(60)90015-9.

[70] Y.-C. Ho, "A successive approximation technique for optimal control systems subject to input saturation," *Journal of Basic Engineering*, vol. 84, no. 1, pp. 33–37, 1962. DOI: 10.1115/1.3657263.

[71] J. Eaton, "An iterative solution to time-optimal control," *Journal of Mathematical Analysis and Applications*, vol. 5, no. 2, pp. 329–344, 1962. DOI: 10.1016/S0022-247X(62)80015-8.

[72] L. Neustadt, "The existence of optimal controls in the absence of convexity conditions," *Journal of Mathematical Analysis and Applications*, vol. 7, pp. 110–117, 1963. DOI: 10.1016/0022-247X(63)90081-7.

[73] T. Babunashvili, "The synthesis of linear optimal systems," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 2, no. 2, pp. 261–265, 1964. DOI: 10.1137/0302023.

[74] T. Fujisawa and Y. Yasuda, "An iterative procedure for solving the time-optimal regulator problem," *SIAM Journal on Control*, vol. 5, no. 4, pp. 501–512, 1967. DOI: 10.1137/0305029.

[75] F. Grognard and R. Sepulchre, "Global analysis of a continuous-time flow which computes time-optimal switchings," in *40th IEEE Conference on Decision and Control*, 2001, pp. 3826–3831. DOI: 10.1109/CDC.2001.980460.

[76] ——, "Computation of time-optimal switchings for linear systems with complex poles," in *2003 European Control Conference*, 2003, pp. 2190–2195. DOI: 10.23919/ECC.2003.7085292.

[77] M. Romano and F. Curti, "Time-optimal control of linear time invariant systems between two arbitrary states," *Automatica*, vol. 120, p. 109 151, 2020. DOI: 10.1016/j.automatica.2020.109151.

[78] D. Liberzon, *Calculus of Variations and Optimal Control Theory: a Concise Introduction*. Princeton University Press, 2011.

[79] Y.-C. Ho and S. Baron, "A minimal time intercept problem," *IEEE Transactions on Automatic Control*, vol. 10, no. 2, pp. 200–200, 1965. DOI: 10.1109/TAC.1965.1098111.

[80] E. Rechtschaffen, "Equivalences between differential games and optimal controls," *Journal of Optimization Theory and Applications*, vol. 18, no. 1, pp. 73–79, 1976. DOI: 10.1007/BF00933795.

[81] A. Marzollo and A. Pascoletti, "On the reachability of a given set under disturbances," *Control and Cybernetics*, vol. 2, no. 3, pp. 99–106, 1973. [Online]. Available: http://oxygene.ibspan.waw.pl:3000/contents/export?filename=1973-3-4-09_marzollo_pescoletti.pdf.

[82] M. Althoff, "On computing the Minkowski difference of zonotopes," *arXiv preprint*, 2015. [Online]. Available: https://arxiv.org/pdf/1512.02794.pdf.

[83] I. Kolmanovsky and E. Gilbert, "Theory and computation of disturbance invariant sets for discrete-time linear systems," *Mathematical Problems in Engineering*, vol. 4, no. 4, pp. 317–367, 1998. DOI: 10.1155/S1024123X98000866.

[84] L. Montejano, "Some results about Minkowski addition and difference," *Mathematika*, vol. 43, no. 2, pp. 265–273, 1996. DOI: 10.1112/S0025579300011761.

[85] Y.-C. Ho, "Review of the book Differential Games by R. Isaacs," *IEEE Transactions on Automatic Control*, vol. 10, pp. 501–503, 1965. DOI: 10.1109/TAC.1965.1098221.

[86] W. Borgest and P. Varaiya, "Target function approach to linear pursuit problems," *IEEE Transactions on Automatic Control*, vol. 16, no. 5, pp. 449–459, 1971. DOI: 10.1109/TAC.1971.1099786.

[87] N. Z. Leon Petrosyan, *Game Theory*. World Scientific, 1996. DOI: 10.1142/2875.

[88] I. Mitchell and C. Tomlin, "Overapproximating reachable sets by hamilton-jacobi projections," *Journal of Scientific Computing*, vol. 19, pp. 323–346, 2003. DOI: 10.1023/A:1025364227563.

[89] M. Heymann, M. Pachter, and R. Stern, "Weak and strong max-min controllability," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 612–613, 1976. DOI: 10.1109/TAC.1976.1101275.

[90] ——, "Max-min control problems: A system theoretic approach," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 455–463, 1976. DOI: 10.1109/TAC.1976.1101315.

[91] R. Brammer, "Controllability in linear autonomous systems with positive controllers," *SIAM Journal on Control*, vol. 10, no. 2, pp. 339–353, 1972. DOI: 10.1137/0310026.

[92] L. Wu, Y. Ke, and B. Chen, "Systematic modeling of rotor dynamics for small unmanned aerial vehicles," in *International Micro Air Vehicle Competition and Conference*, 2016, pp. 284–290. DOI: 10.1142/S2301385018400010.

[93] Z. Artstein, "Linear systems with delayed controls: A reduction," *IEEE Transactions on Automatic control*, vol. 27, no. 4, pp. 869–879, 1982. DOI: 10.1109/TAC.1982.1103023.

[94] V. Léchappé, E. Moulay, F. Plestan, A. Glumineau, and A. Chriette, "New predictive scheme for the control of LTI systems with input delay and unknown disturbances," *Automatica*, vol. 52, pp. 179–184, 2015. DOI: 10.1016/j.automatica.2014.11.003.

[95] J. Conway, *A Course in Functional Analysis*. Springer, 1990. DOI: 10.1007/978-1-4757-3828-5.

[96] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Springer, 2016. DOI: 10.1007/978-3-319-32552-1.

[97] H. Tuy, "Global minimization of a difference of two convex functions," *Mathematical Programming Study*, vol. 30, pp. 150–182, 1987. DOI: 10.1007/BFb0121159.

[98] P. D. Tao and L. T. H. An, "Convex analysis approach to d.c. programming: Theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997. [Online]. Available: http://journals.math.ac.vn/acta/pdf/9701289.pdf.

[99] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003. DOI: 10.1162/08997660360581958.

[100] G. Golub and C. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 2013. [Online]. Available: https://www.press.jhu.edu/books/title/10678/matrix-computations.

[101] P. A. Vela, K. A. Morgansent, and J. W. Burdick, "Underwater locomotion from oscillatory shape deformations," in *41st IEEE Conference on Decision and Control*, vol. 2, 2002, pp. 2074–2080. DOI: 10.1109/CDC.2002.1184835.

[102] J. Yu, C. Wang, and G. Xie, "Coordination of multiple robotic fish with applications to underwater robot competition," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 2, pp. 1280–1288, 2016. DOI: 10.1109/TIE.2015.2425359.

[103] M. Gu and S. Eisenstat, "Downdating the singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 16, no. 3, pp. 793–810, 1995. DOI: 10.1137/S0895479893251472.

[104] A. Hedayat, W. D. Wallis, *et al.*, "Hadamard matrices and their applications," *The Annals of Statistics*, vol. 6, no. 6, pp. 1184–1238, 1978. [Online]. Available: https://www.jstor.org/stable/2958712.

[105] C. Van Loan, "The sensitivity of the matrix exponential," *SIAM Journal on Numerical Analysis*, vol. 14, no. 6, pp. 971–981, 1977. DOI: 10.1137/0714065.

[106] U. N. Lars Forssell, "ADMIRE the aero-data model in a research environment version 4.0, model description," FOI - Swedish Defence Research Agency, Tech. Rep., 2005. [Online]. Available: https://www.foi.se/rest-api/report/FOI-R--1624--SE.

[107] O. Härkegård and T. Glad, "Resolving actuator redundancy - optimal control vs. control allocation," *Automatica*, vol. 41, pp. 137–144, 2005. DOI: 10.1016/j.automatica.2004.09.007.

[108] A. Khelassi, P. Weber, and D. Theilliol, "Reconfigurable control design for over-actuated systems based on reliability indicators," in *Conference on Control and Fault-Tolerant Systems*, 2010, pp. 365–370. DOI: 10.1109/SYSTOL.2010.5675957.

[109] W. Durham, K. Bordignon, and R. Beck, *Aircraft Control Allocation*. John Wiley and Sons, 2017. DOI: 10.1002/9781118827789.

[110] V. Adir and A. Stoica, "Integral LQR control of a star-shaped octorotor," *Incas Bulletin*, vol. 4, no. 2, pp. 3–18, 2012. DOI: 10.13111/2066-8201.2012.4.2.1.

[111] H. Sussmann, "A bang-bang theorem with bounds on the number of switchings," *SIAM Journal on Control and Optimization*, vol. 17, no. 5, pp. 629–651, 1979. DOI: 10.1137/0317045.

[112] G. Aronsson, "Global controllability and bang-bang steering of certain nonlinear systems," *SIAM Journal on Control*, vol. 11, no. 4, pp. 607–619, 1973. DOI: 10.1137/0311047.

[113] K. Glashoff and E. Sachs, "On theoretical and numerical aspects of the bang-bang-principle," *Numerische Mathematik*, vol. 29, no. 1, pp. 93–113, 1977. DOI: 10.1007/BF01389316.

[114] G. M. Ziegler, *Lectures on Polytopes*. Springer Science & Business Media, 2012, vol. 152. DOI: 10.1007/978-1-4613-8431-1.

[115] D. Kolosa, "Implementing a linear quadratic spacecraft attitude control system," M.S. thesis, Western Michigan University, 2015. [Online]. Available: https://scholarworks.wmich.edu/masters_theses/661 (visited on 05/03/2021).

[116] J. S. Hudson and D. J. Scheeres, "Reduction of low-thrust continuous controls for trajectory dynamics," *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 3, pp. 780–787, 2009. DOI: 10.2514/1.40619.

[117] D.-Z. Du and P. M. Pardalos, *Minimax and Applications*. Springer Science & Business Media, 2013, vol. 4. DOI: 10.1007/978-1-4613-3557-3.

[118] R. Hettich and K. Kortanek, "Semi-infinite programming: Theory, methods, and applications," *SIAM Review*, vol. 35, no. 3, pp. 380–429, 1993. DOI: 10.1137/1035089.

[119] M. Posner and C.-T. Wu, "Linear max-min programming," *Mathematical Programming*, vol. 20, no. 1, pp. 166–172, 1981. DOI: 10.1007/BF01589343.

[120] C. Weibel, "Minkowski sums of polytopes: Combinatorics and computation," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2007. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=65bcdeee4cc525f619047e629bf7a8a5ad676fdb.

[121] R. Kalman and J. Bertram, "Control system analysis and design via the "second method" of Lyapunov: Continuous-time systems," *Journal of Basic Engineering*, vol. 82, no. 2, pp. 371–393, 1960. DOI: 10.1115/1.3662604.

[122] H. K. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.

[123] S. G. Krantz, *A handbook of real variables: with applications to differential equations and Fourier analysis*. Springer Science & Business Media, 2011. DOI: 10.1007/978-0-8176-8128-9.

[124] S. H. Trapnes, "Optimal Temperature Control of Rooms," M.S. thesis, Norwegian University of Science and Technology, 2012. [Online]. Available: https://folk.ntnu.no/skoge/diplom/prosjekt12/trapnes/control_of_floor_heating_process.pdf.

[125] M. Moszynska, *Selected Topics in Convex Geometry*. Springer, 2006. DOI: 10.1007/0-8176-4451-2.

[126] The White House, *Presidential Policy Directive 21: Critical infrastructure security and resilience*, Washington, USA, 2013. [Online]. Available: https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil.

[127] M. Ornik and J.-B. Bouvier, "Assured system-level resilience for guaranteed disaster response," in *2022 IEEE International Smart Cities Conference*, 2022, pp. 1–4. DOI: 10.1109/ISC255366.2022.9922438.

[128] L. Shen, Y. Tang, and L. C. Tang, "Understanding key factors affecting power systems resilience," *Reliability Engineering & System Safety*, vol. 212, 2021. DOI: 10.1016/j.ress.2021.107621.

[129] W. Neelen and R. van Duijn, "Hacking traffic lights," in *DEF CON 28 Safe Mode*, 2020. [Online]. Available: https://securityboulevard.com/2020/08/def-con-28-safe-mode-wesley-neelens-and-rik-van-duijns-hacking-traffic-lights/.

[130] D. Marelli, M. Zamani, M. Fu, and B. Ninness, “Distributed Kalman filter in a network of linear systems,” *Systems & Control Letters*, vol. 116, pp. 71–77, 2018. DOI: 10.1016/j.sysconle.2018.04.005.

[131] E. Sontag, “An algebraic approach to bounded controllability of linear systems,” *International Journal of Control*, vol. 39, no. 1, pp. 181–188, 1984. DOI: 10.1080/00207178408933158.

[132] B. N. Datta, *Numerical Methods for Linear Control Systems*. Elsevier, 2004. DOI: 10.1007/978-1-4612-4120-1_4.

[133] L. Qiu and E. Davison, “An improved bound on the real stability radius,” in *1992 American Control Conference*, IEEE, 1992, pp. 588–589. DOI: 10.23919/ACC.1992.4792134.

[134] W. M. Wonham, *Linear Multivariable Control: a Geometric Approach*, 3rd ed. Springer, 1985. DOI: 10.1007/978-1-4684-0068-7.

[135] D. C. Woffinden, “On-orbit satellite inspection: Navigation and Δv analysis,” M.S. thesis, Massachusetts Institute of Technology, 2004. [Online]. Available: https://dspace.mit.edu/handle/1721.1/28862.

[136] N. M. Horri, K. U. Kristiansen, P. Palmer, and M. Roberts, “Relative attitude dynamics and control for a satellite inspection mission,” *Acta Astronautica*, vol. 71, pp. 109–118, 2012. DOI: 10.1016/j.actaastro.2011.07.029.

[137] J. Diaz and M. Abderrahim, “Visual inspection system for autonomous robotic on-orbit satellite servicing,” in *9th ESA Workshop on Advanced Space Technologies for Robotics and Automation*, 2006. [Online]. Available: http://robotics.estec.esa.int/ASTRA/Astra2006/Papers/ASTRA2006-1.4.2.04.pdf.

[138] N. Ortolano, D. Geller, and A. Avery, “Autonomous optimal trajectory planning for orbital rendezvous, satellite inspection, and final approach based on convex optimization,” *The Journal of the Astronautical Sciences*, pp. 444–479, 2021. DOI: 10.1007/s40295-021-00260-5.

[139] P. Dumazert, F. Marchandise, L. Jolivet, D. Estublier, and N. Cornu, “PPS-1350-G qualification status,” in *40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, 2004. DOI: 10.2514/6.2004-3604.