Scalable Techniques for the Computation of Viable and Reachable Sets

Safety Guarantees for High-Dimensional Linear Time-Invariant Systems

by

Shahab Kaynama

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2012

Abstract

Reachability analysis and viability theory are key in providing guarantees of safety and proving the existence of safety-preserving controllers for constrained dynamical systems. The minimal reachable tube and (by duality) the viability kernel are the only constructs that can be used for this purpose. Unfortunately, current numerical schemes that compute these constructs suffer from a complexity that is exponential in the dimension of the state, rendering them impractical for systems of dimension greater than three or four.

In this thesis we propose two separate approaches that improve the scalability of the computation of the minimal reachable tube and the viability kernel for high-dimensional systems. The first approach is based on structure decomposition and aims to facilitate the use of computationally intensive yet versatile and powerful tools for higher-dimensional linear time-invariant (LTI) systems. Within the structure decomposition framework we present two techniques—Schur-based and Riccati-based decompositions—that impose an appropriate structure on the system which is then exploited for the computation of our desired constructs in lower-dimensional subspaces.

The second approach is based on set-theoretic methods and draws a new connection between the viability kernel and maximal reachable sets. Existing tools that compute the maximal reachable sets are efficient and scalable with polynomial complexity in time and space. As such, these scalable techniques can now be used to compute our desired constructs and therefore provide guarantees of safety for high-dimensional systems. Based on this new connection between the viability kernel and maximal reachable sets we propose a scalable algorithm using ellipsoidal techniques for reachability. We show that this algorithm can efficiently compute a conservative under-

approximation of the viability kernel (or the discriminating kernel when uncertainties are present) for LTI systems. We then propose a permissive state-feedback control strategy that is capable of preserving safety despite bounded input authority and possibly unknown disturbances or model uncertainties for high-dimensional systems.

We demonstrate the results of both of our approaches on a number of practical examples including a problem of safety in control of anesthesia and a problem of aerodynamic flight envelope protection.

Preface

Part of the work presented in this thesis has been published in or submitted to several international conferences and scientific journals.

Results from Chapter 3 were published in:

[i] S. Kaynama and M. Oishi, "Complexity reduction through a Schur-based decomposition for reachability analysis of linear time-invariant systems," *International Journal of Control*, vol. 84, no. 1, pp. 165–179, January 2011

[ii] S. Kaynama and M. Oishi, "Schur-based decomposition for reachability analysis of linear time-invariant systems," in *Proc. IEEE Conference on Decision and Control*, Shanghai, China, December 16–18, 2009, pp. 69–74

Results from Chapter 4 will appear in:

[iii] S. Kaynama and M. Oishi, "A modified Riccati transformation for complexity reduction in reachability analysis of linear time-invariant systems," *IEEE Transactions on Automatic Control* (accepted as Full Paper)

Results from Chapter 5 were published in:

[iv] S. Kaynama, J. Maidens, M. Oishi, I. M. Mitchell, and G. A. Dumont, "Computing the viability kernel using maximal reachable sets," in *Proc. Hybrid Systems: Computation and Control*, Beijing, China, April 17–19, 2012, pp. 55–63

Appendix C and Section 5.3.2 are partially based on the results that appeared in:

[v] S. Kaynama, M. Oishi, I. M. Mitchell, and G. A. Dumont, "The continual reachability set and its computation using maximal reachability techniques," in *Proc. Joint IEEE Conference on Decision and Control, and European Control Conference*, Orlando, FL, December 12–15, 2011, pp. 6110–6115

[vi] S. Kaynama, M. Oishi, I. M. Mitchell, and G. A. Dumont, "Fixed-complexity piecewise ellipsoidal representation of the continual reachability set based on ellipsoidal techniques," to appear in *Proc. American Control Conference*, Montreal, QC, June 27–29, 2012

In all papers I was the lead investigator. I determined the problem to solve, conducted the theoretical research and verified the results via simulations. I was responsible for the bulk of the writing of the papers. My supervisors and co-authors Professors M. Oishi, I. M. Mitchell, and G. A. Dumont provided technical and editing feedback. The paper [iv] is a collaborative work with J. Maidens. I formulated the problem, formed the hypothesis of expressing the viability kernel in terms of maximal reachable sets, and presented the solution for the discrete-time case with help and advice from Professor Mitchell. Mr. Maidens offered the solution for the continuous-time case (Propositions 5.1 and 5.2), while I helped with the proofs and the presentation of the results. I developed the computational algorithms and verified the methods via simulations. Professors Oishi, Mitchell, and Dumont provided technical and editing feedback throughout the work.

Table of Contents

List of Tables

List of Figures

Acknowledgments

If I have seen further, it is by standing
on the shoulders of giants.
—Sir Isaac Newton

This thesis would not have been possible without the support of a number of people. I consider myself extremely lucky to have worked under the incredible supervision and mentorship of Professors Meeko Oishi, Ian Mitchell, and Guy Dumont. Their expertise in the field, their readiness to provide consistent feedback, and their willingness to allow me the freedom to work however, wherever, and on whatever topic I find fascinating have been instrumental in my development as a researcher. I would not be where I am today had it not been for their continuous support in every which way I can think of. And for all that I am forever grateful.

I would like to express my gratitude to the rest of my examination committee Professors Ryozo Nagamune, Philip Loewen, Tim Salcudean, Shahriar Mirabbasi, Zinovy Reichstein, and Dušan Stipanović (U. of Illinois at Urbana-Champaign) for the time and effort spent in considering my candidacy. I am especially indebted to Professor Nagamune for providing great feedback on my research throughout these past few years.

I have had wonderful collaborations with John Maidens—collaborations that I hope continue beyond our time at UBC. I have no doubt that a bright future awaits him. I would also like to take this opportunity to thank Dr. Alex A. Kurzhanskiy (UC Berkeley) for always generously and diligently answering my many questions regarding Ellipsoidal Toolbox.

I have had the pleasure of being part of a great research group. Nikolai, Neda, Ahmad, Pouyan, and Pouria have all made my stay at UBC an inter-

estingly nonlinear experience. I thank Nikolai Matni (Caltech) in particular for all our intellectually stimulating conversations (both before and after his departure from UBC) and our mutual fascination by the beauty of the field of controls. He has been both a fantastic friend and a great resource to bounce ideas off of. The Control Systems Reading Group meetings were inspiring and informative: Special thanks to Professors Bhushan Gopaluni, Farrokh Sassani, Ryozo Nagamune, and Jin-Oh Hahn (U. of Alberta) and the rest of the team, particularly Ehsan, Daniel, Devin, Marius, and Klaske.

Many scholars at UBC, UMIST, and EMU have had lasting influence on my understanding of controls and mathematics. I would like to particularly thank Professor Runyi Yu (EMU) whose in-depth knowledge of the field, genuine care for his students' success, and consistent support of my academic endeavors throughout the years have helped shape what I am today.

Last but not least, I thank my family—Mom, Dad, Kaveh (and Kimia), and Sina—for their never-ending and unconditional love, encouragement, and moral support. Thank you for teaching me the importance of education and the gratification of success. I am who I am because of you. My wonderful wife, Blerina, to whom I wholeheartedly dedicate this thesis, has been incredibly supportive of me and my work throughout the past almost decade—since our very first date at Café Blue Note! Had it not been for her selfless sacrifices and her faith in me, this work would simply not have come to existence. Thank you sweetheart for helping me up when days were dark and cheering me on when they were not. I also thank Deni for putting up with me over the past couple of years, and for being the most well-behaved and disciplined teenager I have known.

This work has been financially supported by NSERC Discovery Grants #327387 and #298211, NSERC Collaborative Health Research Project #CHRPJ-350866-08, and the Institute for Computing, Information and Cognitive Systems (ICICS).

To Blerina.

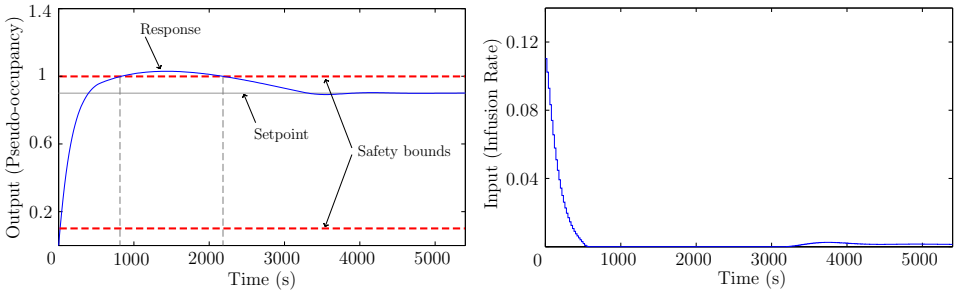
Chapter 1

Introduction

1.1 Motivation

Constrained dynamical systems have deservingly received a tremendous amount of attention among researchers in both academia and industry due to the presence of safety constraints and hard bounds that appear in many practical scenarios. Providing guarantees of constraint satisfaction and facilitating appropriate synthesis of constraint-satisfying controllers therefore is highly desirable, particularly in safety-critical applications.

One example of such an application with safety constraints and bounded input authority is the closed-loop control of anesthesia [26]; safety (state or output constraints) may be defined in terms of prespecified therapeutic bounds on plasma concentration of the anesthetics, and the input (drug infusion rate) is physically bounded by the actuator limitations. To ensure safety of the patient and obtain regulatory certificates, guarantees of safety of the system can play an important role. While we postpone further discussions on the safety aspects of the control of depth of anesthesia to Sections 5.3.2 and 6.3.3, we shall point out that naive design of a controller—even when all constraints have been accounted for—may result in violation of the safety constraints. For example, Figure 1.1 shows the response of a patient using a model predictive controller (MPC) that has been designed with both input and safety (output in this case) constraints in mind. Additional pre-design analysis such as formulation of a terminal constraint set for the receding horizon optimization problem and/or other tricks and techniques that enforce invariance, (strong) feasibility, and other properties of the closed-loop system are warranted. The reader is referred to e.g. [11, 61, 84] for a thorough treatment of such techniques as well as other application areas where



(a) Response of the closed-loop system.

(b) Synthesized control law.

Figure 1.1: Control of drug plasma concentration in a high-responder patient (#80) undergoing 1.5 hr surgery using an inadequately-designed MPC. The input (drug infusion rate) is bounded above and below ($u \in [0, 0.8]$). Clearly, the hard constraint on the output ($y \in [0.1, 1]$) indicating the safety requirement is violated for nearly 20 mins of the surgery; The receding horizon optimization becomes infeasible with respect to its initial condition, resulting in "softening" of the safety constraint.

constraint satisfaction in MPC is desirable.

Another example is the so-called flight or aircraft envelope protection problem [81, 83, 116, 117], where the safety constraints are defined as the aircraft's aerodynamic envelope and consequently the flight management system must ensure that certain combinations of states are avoided at all times to prevent stalling or other undesirable behaviors.

Other application domains in which safety must be maintained despite bounded inputs include aircraft autolanders [8], collision avoidance [25, 28, 46, 89, 96], automated highway systems [80], control of under-actuated underwater vehicles [98], stockout prevention of constrained storage systems in manufacturing processes [13], management of a marine renewable resource [9], and safety in semi-autonomous vehicles [119], to name a few.

1.2 Reachability Analysis and Viability Theory

Reachability analysis and viability theory provide solid frameworks for control synthesis and trajectory analysis of constrained dynamical systems in a set-valued fashion (cf. [5, 12, 22, 67, 69, 118]) and have been utilized in diverse applications ranging from those listed above to the control of uncertain oscillatory systems [23] and verification of temporal properties of a toggle circuit [44], etc.

Reachability analysis identifies the set of states forward (backward) reachable by a constrained dynamical system from a given initial (target) set of states. Reachability analysis distinguishes itself from what *simulations* can achieve in the sense that with simulations a single trajectory or execution of the system corresponding to a single initial state is computed at a time, whereas with reachability analysis all points belonging to all possible trajectories or executions are computed at once from all possible initial states. Properties of the system inferred from such set-valued computations are therefore universal and hold true for the entire set of initial states. Similarly, viability theory provides a set-valued insight into the behavior of the trajectories inside a given constraint set. For example, the *viability kernel* is the set of initial states for which there exists at least one trajectory or execution of the input-constrained system that respects the state constraint for all time.

The notions of *maximal* and *minimal* reachability analysis were introduced in [86]. Their corresponding constructs differ in how the time variable and the bounded input are quantified. In formation of the maximal reachability construct, the input tries to steer as many states as possible to the target set. In formation of the minimal reachability construct, the trajectories reach the target set regardless of the input applied. Based on these differences, the maximal and minimal *reachable sets* and *tubes* (the set of states traversed by the trajectories over the time horizon [70, 86]) are formed. The objects generated by each of these constructs have unique attributes: The maximal reachability constructs can be used to synthesize input policies that steer the trajectories of the system to the target (or,

to analyze how the trajectories behave under uncertainty and/or external disturbance). The minimal reachability constructs, on the other hand, can be used to synthesize inputs that keep the trajectories of the system away from the target set. The viability kernel and the minimal reachable tube are duals of one another [18, 79], while the *invariance kernel* (the set of states that remain in the constraint set for all possible inputs for all time) is the dual of the maximal reachable tube [5, 69, 79].

It is shown in [86] and [5] that the minimal reachable tube and the viability kernel are the *only* constructs that can be used to prove safety/viability of the system and to synthesize inputs (controllers) that preserve this safety. As such it is highly desirable to be able to compute these constructs for possibly high-dimensional safety-critical constrained systems for which guarantees of safety and prevention of constraint violation are crucially important.

1.3 Computational Techniques and the "Curse of Dimensionality"

In this thesis we are concerned with *backward* constructs generated by reachability analysis and viability theory. That is, we seek to compute the set of initial states that is formed under the input-constrained system for a given target/terminal or constraint set of states. In general an exact computation of the reachability or viability constructs is extremely difficult if not impossible. Instead, approximations of these sets are computed. Such computations have historically been subject to Bellman's "curse of dimensionality"; their complexity increases rapidly with the dimension of the continuous states [4].

Algorithms that approximate the backward constructs can be divided into two main categories [86]: *Eulerian methods* (e.g., [18, 33, 89, 106]) are capable of computing the viability kernel and by duality, the minimal reachable tube. Although versatile in terms of ability to handle complex dynamics and constraints, these algorithms rely on gridding the state space and therefore their computational complexity increases exponentially with the dimension of the state, rendering them impractical for systems of di-

mension higher than three or four. The second category of algorithms are *Lagrangian methods* (e.g., [30, 39, 40, 50, 70, 73, 78]) that follow trajectories and compute the maximal reachable sets and tube in a scalable and efficient manner. These algorithms take advantage of compact set representations (e.g., ellipsoids and zonotopes) and/or in general the convexity of all constraints.[1] Their computational complexity is therefore usually polynomial in time and space, making them suitable for application to high-dimensional systems.

1.4 The Goal and Contributions of the Thesis

In many applications such as those involving safety-critical systems the ability to compute the minimal reachable tube and the viability kernel can be paramount: Guarantees of safety and synthesis of safety-preserving controllers can only be obtained using these constructs. However, if the system is of even moderate dimension, the Eulerian methods that to this date have exclusively[2] been used to compute these constructs cannot be employed. This argument is the cornerstone of our research.

This thesis presents our efforts to address the scalability aspect of the curse of dimensionality to enable the computation of the minimal reachable tube and the viability kernel for higher-dimensional systems. We do so using two separate approaches:

- Complexity reduction via *structure decomposition*, our first approach,

[1] The convexity requirement may be circumvented if a non-convex constraint can be represented by the union of appropriate closed and bounded convex sets. In such a case the computations are performed using a few convex initial sets, which in turn affects the computational time only linearly in the number of initial sets. In the general case regardless of the convexity issues, the resulting set computed by Lagrangian methods over the entire time horizon is a close approximation of the (possibly non-convex) maximal reachable tube.

[2] One exception is the method of polytopes [76], a Lagrangian method developed from within the MPC community, for discrete-time LTI systems that allows for the computation of the controlled-invariant subset (viability kernel) when all constraints are polytopes. However, similarly to Eulerian methods, this technique suffers from an exponential complexity for most systems: The number of vertices of the polytope increases rapidly with each successive Minkowski sum. The convex-hull operation performed by the technique is also known to be computationally demanding [93].

aims to facilitate the use of powerful Eulerian methods on higher-dimensional continuous-time continuously-valued linear time-invariant (LTI) systems (and by extension, hybrid systems with LTI continuous dynamics). As such, many of the benefits that Eulerian methods offer (e.g., ability to handle arbitrarily-shaped possibly unbounded non-convex constraints, or the synthesis of safety-preserving control laws) can now be taken advantage of for relatively higher-dimensional LTI systems.

- Complexity reduction via *set-theoretic methods*, our second approach, bridges the gap between maximal and minimal constructs and aims to facilitate the use of the scalable and efficient Lagrangian methods for the computation of the viability kernel for continuous- and discrete-time (possibly nonlinear) systems. Thanks to these results we propose an efficient and scalable piecewise ellipsoidal algorithm that not only enables the approximation of the viability kernel for high-dimensional LTI systems, but also facilities a scalable synthesis of safety-preserving controllers.

1.4.1 Complexity Reduction via Structure Decomposition

Decomposing the system into lower dimensional subsystems is an approach (referred to as *structure decomposition*) that has been utilized in the past in e.g. [88, 91, 113] as a means to reduce the computational complexity in reachability analysis. While applicable to nonlinear systems, these techniques assume that the system itself presents a certain structure that can be exploited. In this thesis we propose a number of techniques, based on the structure decomposition framework, applicable to LTI systems of generic form. We assume no initial structures. Our techniques (under certain conditions) will *impose* the desired structure on the system which is then exploited for decomposition purposes and ultimately for reduction of complexity in reachability analysis.

The decomposition allows for the computation of the reachable tube in lower dimensions, thus reducing the computational burden significantly.

We accomplish this through transformation of the system into appropriate coordinate spaces in which reachability analysis could be performed in lower-dimensional subspaces and is guaranteed to yield an over-approximation of the actual minimal reachable tube in that space. By performing the analysis in lower dimensions we obtain significant reduction in the computational costs, albeit at the expense of over-approximation.

While also applicable to maximal reachable tube computation, our proposed techniques are of primary benefit to the computation of the minimal reachable tube (or by duality, the viability kernel) for which Eulerian methods have traditionally been used exclusively.

1.4.2 Complexity Reduction via Set-Theoretic Methods

Through our second approach we will show that the viability kernel can be expressed as a nested sequence of maximal reachable sets. By bridging the gap between the viability kernel and the maximal reachable sets we pave the way for more efficient computation of the viability kernel through the use of Lagrangian algorithms. Significant reduction in the computational costs can be achieved since instead of a single calculation with exponential complexity one can perform a series of calculations with polynomial complexity.

Based on the well-studied ellipsoidal techniques for maximal reachability analysis [70] we propose an algorithm that computes a guaranteed piecewise ellipsoidal under-approximation of the viability kernel for LTI systems. We show that the proposed algorithm is efficient and scalable and can be used to address the safety needs of high-dimensional safety-critical systems (such as the anesthesia automation or the flight envelope protection problems mentioned at the beginning of this chapter).

The presented connection between the maximal reachable sets and the viability kernel facilitates a more scalable computation of the viability kernel (and by duality, the minimal reachable tube) as well as the respective safety-preserving control laws.

1.5 Related Work

Complexity reduction for reachability analysis has been addressed by a number of researchers. Methods to compute the reachability constructs (maximal or minimal) for higher-dimensional systems can be divided into three main categories. First are techniques that take advantage of certain representations of sets [30, 40, 64, 65, 74, 76, 78, 108]. Second are techniques that make use of model approximation [42, 47], hybridization [3], projection [44, 91] and structure decomposition [48, 113, 121]. Finally, third are methods that combine the approaches from the first two categories; For example, the proposed technique in [49] employs Krylov subspace projection along with low-dimensional polytopes to compute the maximal reachable sets/tube for large-scale affine systems. Related works on complexity reduction for the computation of the minimal reachable tube and the viability kernel are surveyed next.

Efficient techniques for synthesis of maximal reachability controllers that steer the system to a given target set have been extensively studied in the past. As the main objective of this thesis is safety and safety-preserving control, we refrain from surveying these techniques and only provide a few references: [39, 40, 64, 68, 69, 94]. Instead, here we will discuss relevant techniques that are capable of directly synthesizing safety-preserving control laws.

1.5.1 Complexity Reduction for Computation of Minimal Reachable Tube and Viability Kernel

A projection scheme in [91] based on Hamilton-Jacobi (HJ) partial differential equations (PDEs) over-approximates the projection of the actual minimal reachable tube in lower dimensional subspaces, with the unmodeled dimensions treated as a disturbance. Similarly, [113] decomposes a full-order nonlinear system into either disjoint or overlapping subsystems and solves multiple HJ PDEs in lower dimensions. The computed minimal reachable tube for each subsystem is an over-approximation of the projection of the full-order minimal reachable tube onto the subsystem's subspace.

More recently, in [88] a mixed implicit-explicit HJ formulation of the minimal reachable tube is presented for nonlinear systems whose state vector contains states that are integrators of other states. It is shown that the computational complexity of this new formulation is linear in the number of integrator states, while still exponential in the dimension of the rest of the state space.

In [21] an approximate dynamic programming technique is proposed that, although still grid-based, enables a more efficient computation of the viability kernel. The viability kernel (similarly to [79]) is expressed as the zero sub-level set of the value function of the corresponding optimal control problem. The authors assume that the value function, which is a viscosity solution of a HJB PDE, is differentiable everywhere on the constraint set. The HJB PDE is then discretized and the resulting value function is numerically computed on a grid using a function approximator such as the k-nearest neighbor algorithm. An error bound on the (over-)approximation of the viability kernel is provided. The approximation converges to the true kernel in the limit, as the number of grid points goes to infinity.

In [25] it is shown that for systems with *order-preserving* dynamics when the control set is a direct product of two compact intervals in $\mathbb{R}$ (i.e. the control set is a two-dimensional, axis-aligned rectangle), the minimal reachable tube equals the intersection of two reachable tubes each with a constant input. The constant inputs take value on the opposite vertices of the input set where the maximum of one interval meets the minimum of the other. The reachable tubes with constant inputs can be computed efficiently resulting in an efficient computation of the original minimal reachable tube. This scheme also yields an efficient synthesis of safety-preserving control laws.

Another related approach is the search for a barrier certificate [99] (a Lyapunov-like function) for the system $\dot{x} = f(x, u)$, $x \in \mathcal{X}$, $u \in \mathcal{U}$, that forms a separating surface between any two given sets $\mathcal{A}$ and $\mathcal{B}$. If there exists a function $B\colon \mathcal{X} \to \mathbb{R}$ such that $B(x) \leq 0\ \forall x \in \mathcal{A}$, $B(x) > 0\ \forall x \in \mathcal{B}$, and $L_f B(x) \leq 0\ \forall (x, u) \in \mathcal{X} \times \mathcal{U}$ along the zero level set of B (here L_f denotes the Lie derivative along f), then no trajectories will ever go from $\mathcal{A}$ to $\mathcal{B}$. This technique can be adapted to analytically formulate the boundary

of the *infinite-horizon* viability kernel as well—if it is non-empty. The Lie derivative on the surface of the candidate certificate must now satisfy $\exists u \in \mathcal{U}$ $\forall x \in \mathcal{X}$ $L_f B(x) \leq 0$. It is shown in [99] that for systems with polynomial vector fields and semi-algebraic constraints (e.g. polynomial inequalities), efficient techniques based on Sum of Squares and semi-definite programming can be used to find the barrier certificate.[3]

1.5.2 Safety-Preserving Control Synthesis

Safety-preserving controllers, the control policies associated with viability kernels and minimal reachable tubes, are capable of keeping the trajectories of the system within the safe region of the state space. Their synthesis has therefore received significant attention among researchers as a means to guarantee the safety/viability of constrained dynamical systems [2, 5, 11, 81, 116] (cf. [7, 12, 89] for more recent expositions).

A classical approach is to use the information of the shape of the computed set (viability kernel or minimal reachable tube) by choosing the safety-preserving optimal control laws based on the contingent cone [14] or the proximal normal [5, 18] at every point on the boundary of the set as per Nagumo's theorem and its generalizations (cf. [5]). Similarly, a given viability kernel can be used as a feasible or terminal constraint set to guarantee (strong) feasibility of the receding horizon optimization problem in an MPC framework [11, 61, 84, 100, 101].[4] This results in (recursive) constraint satisfaction of the closed-loop system and therefore the generated control laws can be regarded as safety-preserving. In both cases discussed above the computational complexity of synthesizing such control laws is at best equal to that of computing the corresponding viability kernel or minimal reachable

[3] This method cannot be used to formulate the *finite-horizon* viability kernel which may be useful when e.g. the infinite-horizon kernel is empty. Moreover, there are no guarantees that a barrier certificate can be found (even with simple, stable linear dynamics for which a Lyapunov function can be systematically constructed).

[4] While alternative techniques exist that eliminate the need for terminal constraint sets for feasibility (such as ensuring that the horizon of the receding horizon optimization problem is "sufficiently" large), in this thesis we are only concerned with computing the viability kernel and the minimal reachable tube as well as their corresponding safety controllers and will not cover such techniques.

tube.

Another approach is related to the previously described barrier certificate technique. The technique has been extended in [102] to synthesize safety-preserving controllers using density functions and convex optimization methods for polynomial nonlinear systems with semi-algebraic constraints and simple magnitude bounded inputs. The pros and cons of this approach are similar to those of computing a barrier certificate with the exception that for the synthesis problem a gridding of the state space is also required. Unfortunately, this gridding renders the technique intractable in high dimensions.

A classification technique based on support vector machines (SVMs) is presented in [24] that approximates the viability kernel and yields an analytical expression of its boundary. A sequential minimal optimization algorithm is solved to compute the SVM that in turn forms a barrier function [95] (not to be confused with a barrier certificate mentioned above) on or close to the boundary of the viability kernel in the discretized space. While the method successfully reduces the computational time for the synthesis of control laws when the dimension of the input space is high, its applicability to systems with high-dimensional state space is limited as it relies on the same state space gridding approach used in the classical techniques e.g. [106]. Furthermore, the method does not provide any guarantees that the synthesized control laws are safety-preserving.

The notion of approximate bisimulation [41] can be used to construct a discrete abstraction of the continuous state space such that the observed behavior of the corresponding abstract system is "close" to that of the original continuous system. Girard *et al.* in a series of papers [17, 37, 38] use this notion to construct safety-preserving controllers for approximately bisimilar discrete abstractions and prove that the controller is *correct-by-design* for the original systems. The technique is applied to incrementally stable switched systems (for which approximately bisimilar discrete abstractions of arbitrary precision can be constructed) with autonomous or affine dynamics, and safety-preserving switched controllers are synthesized. The abstraction, however, relies on sampling of time and space (i.e. gridding) and therefore

its applicability appears to be limited to low-dimensional systems.

1.6 Organization of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 provides necessary preliminaries and formally formulates the objectives of the thesis. Chapter 3 describes our first technique, Schur-based decomposition, within the structure decomposition framework. Our second decomposition technique, Riccati-based decomposition, is presented in Chapter 4 building upon the results of Chapter 3 and providing a more in-depth treatment of the structure decomposition framework for complexity reduction in reachability analysis. These two chapters are based on the assumption that the system dynamics are LTI and aim to facilitate the applicability of computationally intensive Eulerian methods for higher-dimensional LTI systems. Chapter 5 describes a set-theoretic approach that enables the use of efficient Lagrangian methods for the computation of the viability kernel, circumventing entirely the need to employ computationally intensive Eulerian methods. An efficient algorithm is presented that approximates the viability kernel of LTI systems in a scalable fashion. Chapter 6 extends these results to systems with disturbances and then formulates a scalable and robust safety-preserving feedback control strategy for LTI systems. Chapter 7 summarizes the thesis reiterating its major contributions, and provides directions for future research.

Finally, supplementary materials are provided in the Appendices: Appendix A contains the proof of Proposition 3.4 (page 38) and elaborates on Assumptions 3.1 (page 34) and 4.2 (page 54) used in Chapters 3 and 4. Appendix B contains the proofs of Propositions 4.2 (page 60) and 4.3 (page 61) and also provides an upper-bound on the condition number of the proposed modified Riccati transformation matrix in Chapter 4. Appendix C is complementary to the reachability constructs used within the body of the thesis and presents additional backward constructs that are formed under constrained dynamical systems. Their connections to one another and to the constructs used within the thesis are formalized, aiming to help the reader

attain a more complete picture of the contributions of this thesis.

1.7 Basic Notation

The quantifiers $\exists$ and $\forall$ are existential and universal, respectively. We denote the $N \times N$ identity matrix by I_N and the inner product by $\langle \cdot, \cdot \rangle$. For brevity, $\|\cdot\|$ denotes the infinity norm. For a constant matrix $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ the induced norm is

$$\|A\| := \sup_{v \in \mathbb{R}^n,\, v \neq 0} \frac{\|Av\|}{\|v\|} = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|. \tag{1.1}$$

For a Lebesgue measurable function $f \colon \mathbb{R} \to \mathbb{R}^n$ defined over an interval $[t_a, t_b]$ we denote

$$\|f\| := \|f(\cdot)\|_{\mathcal{L}_\infty[t_a, t_b]} = \sup_{t \in [t_a, t_b]} \|f(t)\| < \infty. \tag{1.2}$$

The *Hausdorff distance* between any two nonempty subsets $\mathcal{X}$ and $\mathcal{Y}$ of a metric space $(\mathbb{R}^n, d)$ is defined as

$$\operatorname{dist}_H(\mathcal{X}, \mathcal{Y}) := \max\left\{ \sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} d(x, y),\ \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} d(x, y) \right\}. \tag{1.3}$$

The *erosion* of $\mathcal{X}$ by $\mathcal{Y}$ (also known as the *Pontryagin difference* between $\mathcal{X}$ and $\mathcal{Y}$) is defined as

$$\mathcal{X} \ominus \mathcal{Y} := \{x \mid x + y \in \mathcal{X} \ \ \forall y \in \mathcal{Y}\}. \tag{1.4}$$

The *Minkowski sum* of $\mathcal{X}$ and $\mathcal{Y}$ is defined as

$$\mathcal{X} \oplus \mathcal{Y} := \{x + y \mid x \in \mathcal{X},\ y \in \mathcal{Y}\}. \tag{1.5}$$

The *projection* of a set $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$ onto $\mathcal{X}$ is defined as

$$\operatorname{Proj}_{\mathcal{X}}(\mathcal{A}) := \{x \in \mathcal{X} \mid \exists y \in \mathcal{Y} \ \text{ s.t. } \ (x, y) \in \mathcal{A}\}. \tag{1.6}$$

We denote by $\mathring{\mathcal{X}}$ the interior of $\mathcal{X}$, by $\partial\mathcal{X}$ its boundary, and by $\operatorname{cl}\mathcal{X} := \mathring{\mathcal{X}} \cup \partial\mathcal{X}$ its closure. The set $\mathcal{B}(\delta)$ denotes a norm-ball of radius $\delta \in \mathbb{R}^+$ about the origin in $\mathbb{R}^n$. The set $\mathcal{X}^c$ denotes the complement of $\mathcal{X}$ in $\mathbb{R}^n$.

Chapter 2

Preliminaries and Problem Formulation

2.1 Backward Constructs for Constrained Dynamical Systems

Consider a continuously-valued dynamical system

$$\mathcal{L}(x(t)) = f(x(t), u(t)), \quad x(0) = x_0 \tag{2.1}$$

with state space $\mathcal{X} := \mathbb{R}^n$ (a finite-dimensional vector space), state vector $x(t) \in \mathcal{X}$, and input $u(t) \in \mathcal{U}$ where $\mathcal{U}$ is a compact and convex subset of $\mathbb{R}^m$. Depending on whether the system evolves in continuous time ($t \in \mathbb{R}^+$) or discrete time ($t \in \mathbb{Z}^+$), $\mathcal{L}(\cdot)$ denotes the derivative operator or the unit forward shift operator, respectively. The vector field $f\colon \mathcal{X} \times \mathcal{U} \to \mathcal{X}$ is assumed to be continuous in its arguments in the discrete-time case, and Lipschitz in x and continuous in u in the continuous-time case. Let

$$\mathscr{U}_{[0,t]} := \{u\colon [0,t] \to \mathbb{R}^m \text{ measurable}, \;\; u(t) \in \mathcal{U} \text{ a.e.}\}. \tag{2.2}$$

With an arbitrary, finite time horizon $\tau > 0$, for every $t \in [0, \tau]$, $x_0 \in \mathcal{X}$, and $u(\cdot) \in \mathscr{U}_{[0,t]}$, there exists a unique trajectory $x^u_{x_0}\colon [0,t] \to \mathcal{X}$ that satisfies the initial condition $x^u_{x_0}(0) = x_0$ and the differential/difference equation (2.1). When clear from the context, we shall drop the subscript and superscript from the trajectory notation.

For a nonempty state constraint/target set $\mathcal{K} \subset \mathcal{X}$ this thesis is primarily concerned with the following *backward* constructs, their implications, and

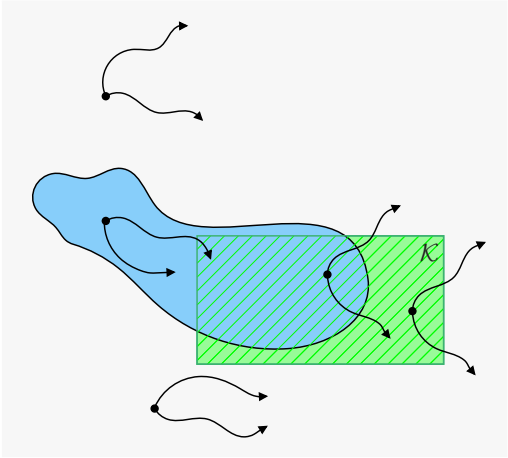


Figure 2.1: The maximal reachable set $Reach_t^\sharp(\mathcal{K},\mathcal{U})$ (blue/plain) for the target set $\mathcal{K}$ (green/patterned). A few sample initial conditions and trajectories are also shown.

their connections to one another. (For other backward constructs please see Appendix C.)

Definition 2.1 (Maximal Reachable Set)**.** *The maximal reachable set at time t is the set of initial states for which there exists an input such that the trajectories emanating from those states reach $\mathcal{K}$ exactly at time t (Figure 2.1):*

$$Reach_t^\sharp(\mathcal{K},\mathcal{U}) := \{x_0 \in \mathcal{X} \mid \exists u(\cdot) \in \mathscr{U}_{[0,t]},\ x_{x_0}^u(t) \in \mathcal{K}\}. \tag{2.3}$$

Definition 2.2 (Maximal Reachable Tube)**.** *The maximal reachable tube*[1] *over the horizon $[0,\tau]$ is the set of initial states for which there exists an input such that the trajectories emanating from those states reach $\mathcal{K}$ at some time $t \in [0,\tau]$ (Figure 2.2):*

$$Reach_{[0,\tau]}^\sharp(\mathcal{K},\mathcal{U}) := \{x_0 \in \mathcal{X} \mid \exists u(\cdot) \in \mathscr{U}_{[0,\tau]},\ \exists t \in [0,\tau],\ x_{x_0}^u(t) \in \mathcal{K}\}. \tag{2.4}$$

[1] Also known as the *possible victory domain* [18], the *attainability tube* [70], or the *capture basin* [6].

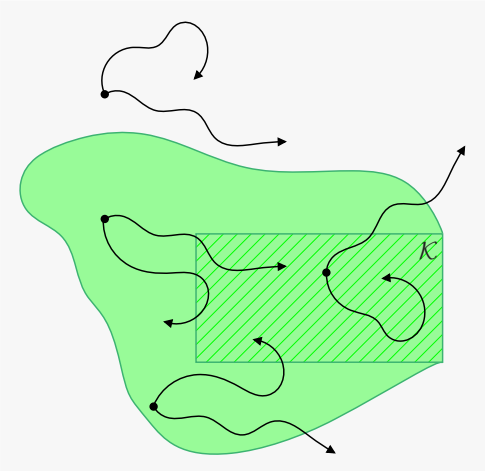


Figure 2.2: The maximal reachable tube $Reach^{\sharp}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ (also includes $\mathcal{K}$ itself). A few sample initial conditions and trajectories are also shown.

Definition 2.3 (Minimal Reachable Tube)**.** *The minimal reachable tube*[2] *over the horizon* $[0,\tau]$ *is the set of initial states for which for every input there exits a time* $t \in [0,\tau]$ *such that the trajectories emanating from those states reach* $\mathcal{K}$ *at* t *(Figure 2.3):*

$$Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}) := \{x_0 \in \mathcal{X} \mid \forall u(\cdot) \in \mathscr{U}_{[0,\tau]},\ \exists t \in [0,\tau],\ x^u_{x_0}(t) \in \mathcal{K}\}. \quad (2.5)$$

Definition 2.4 (Viability Kernel)**.** *The (finite-horizon) viability kernel*[3] *of* $\mathcal{K}$ *is the set of all initial states in* $\mathcal{K}$ *for which there exists an input such that the trajectories emanating from those states remain in* $\mathcal{K}$ *for all time* $t \in [0,\tau]$ *(Figure 2.4):*

$$Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) := \{x_0 \in \mathcal{X} \mid \exists u(\cdot) \in \mathscr{U}_{[0,\tau]},\ \forall t \in [0,\tau],\ x^u_{x_0}(t) \in \mathcal{K}\}. \quad (2.6)$$

What differentiates the above constructs from one another is the *type*

[2] Also known as the *certain victory domain* [18], or the *capture set* in the differential games literature [81] (not to be confused with the capture basin).

[3] The infinite-horizon viability kernel $Viab_{\mathbb{R}+}(\mathcal{K},\mathcal{U})$ is also known as the *maximal controlled-invariant subset* [11].

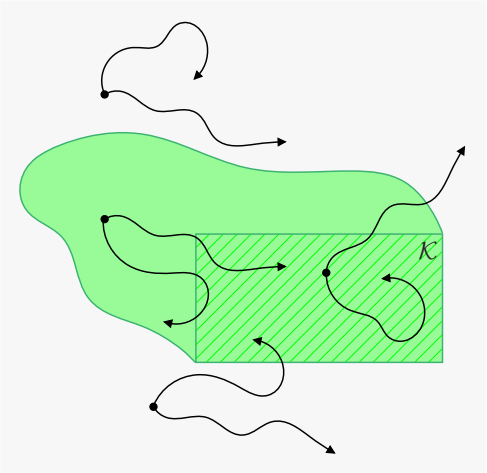


Figure 2.3: The minimal reachable tube $Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ (also includes $\mathcal{K}$ itself). A few sample initial conditions and trajectories are also shown.

and *order* of quantifiers operating on the time and input variables. These seemingly subtle differences generate fundamentally distinct sets (with properties that are unique to each of them). In particular, the following inclusions hold.

Proposition 2.1.

$$Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq \mathcal{K} \subseteq Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Reach^{\sharp}_{[0,\tau]}(\mathcal{K},\mathcal{U}). \tag{2.7}$$

Proof. That $Viab_{[0,\tau]}(\mathcal{K}) \subseteq \mathcal{K}$ is well-known [5]. To show $\mathcal{K} \subseteq Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ take $x_0 \in \mathcal{K}$ and let $\tau = 0$. Thus $x^u_{x_0}(0) = x_0$ for any $u(\cdot) \in \mathscr{U}_{[0,\tau]}$ which implies $x_0 \in Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U})$. To prove $Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Reach^{\sharp}_{[0,\tau]}(\mathcal{K},\mathcal{U})$, take $x_0 \in Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U})$. We have that $\forall u(\cdot) \in \mathscr{U}_{[0,\tau]}\ \exists t \in [0,\tau]\ \ x^u_{x_0}(t) \in \mathcal{K} \Longrightarrow \exists u(\cdot) \in \mathscr{U}_{[0,\tau]}\ \exists t \in [0,\tau]\ \ x^u_{x_0}(t) \in \mathcal{K} \Longleftrightarrow x_0 \in Reach^{\sharp}_{[0,\tau]}(\mathcal{K},\mathcal{U})$. □

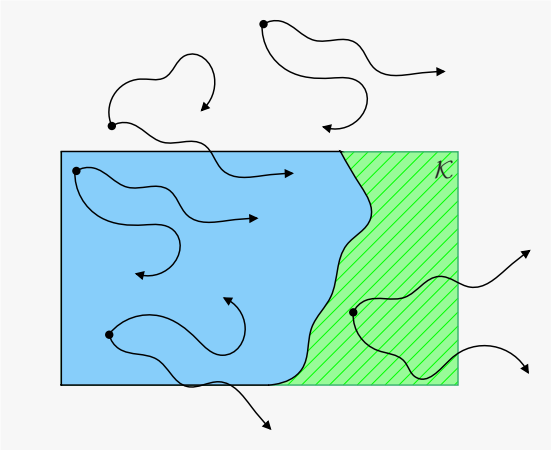


Figure 2.4: The (finite-horizon) viability kernel $Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U})$ (blue/plain). A few sample initial conditions and trajectories are also shown.

2.2 Significance of Minimal Reachable Tube and Viability Kernel

What is so significant about computing the minimal reachable tube and the viability kernel? The short answer is "guarantees of safety" and "safety-preserving control synthesis". To see this let us begin with maximal reachability constructs.

The maximal reachable tube and sets can be used to synthesize input policies that steer the trajectories of the system to the target, or if the target is deemed "unsafe", they can be used to analyze the set of states backward reachable in the worst case by the system under a bounded disturbance or uncertainty.

The recently developed Lagrangian methods (e.g. [30, 39, 40, 50, 70, 73]) approximate the maximal reachable tube by first fixing the time variable and computing the corresponding maximal reachable set, and then taking the

union of these sets over the time horizon. It is easy to verify that

$$Reach^{\sharp}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \equiv \bigcup_{t\in[0,\tau]} \left\{x_0 \in \mathcal{X} \mid \exists u(\cdot) \in \mathscr{U}_{[0,t]},\, x^u_{x_0}(t) \in \mathcal{K}\right\} \tag{2.8}$$

$$= \bigcup_{t\in[0,\tau]} Reach^{\sharp}_t(\mathcal{K},\mathcal{U}). \tag{2.9}$$

The equality holds since both quantifiers operating on the input and time variables in (2.4) are existential and therefore their order can be interchanged. This connection, which has been recognized and extensively used since the earlier works in the literature (e.g. [20, 22, 69]), was also formalized in [86] and proven using the Hamilton-Jacobi-Bellman framework in [79]. The Lagrangian methods are scalable and computationally efficient (with polynomial complexity in both time and space).

On the other hand, the minimal reachable tube and the viability kernel can be used to synthesize *safety-preserving* inputs that keep the trajectories of the system away from the unsafe target set, or contained within a safe constraint set. In fact, it is shown in [86] and [5] that the minimal reachable tube, and by duality the viability kernel, are the *only* constructs that can be employed to prove the existence of an input which guarantees safety of the system.

Notice that, as shown in [18] and as a dual of the results in [79], we have

$$Viab_{[0,\tau]}(\mathcal{K}^c,\mathcal{U}) = (Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}))^c. \tag{2.10}$$

Since the viability kernel and the minimal reachable tube are duals of one another, they need not be treated separately. In this thesis we will initially focus on computing the minimal reachable tube for an unsafe target $\mathcal{K}$ in Chapters 3 and 4, and then shift focus to the case in which constraint $\mathcal{K}$ is deemed safe and therefore compute the viability kernel in the remaining chapters. Note however that approximating the minimal reachable tube and the viability kernel are not mutually exclusive—a method that facilitates an under-approximation of the viability kernel of $\mathcal{K}$ automatically provides a

means for the over-approximation of the minimal reachable tube for $\mathcal{K}^c$.[4]

In computing the minimal reachable tube the input is universally quantified. Furthermore, the time variable must be quantified only *after* the input is quantified. Interchanging the order of quantifiers here will change the meaning of the set and is therefore not possible. In fact, as also shown in [86],

$$\begin{aligned} \mathit{Reach}^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \supseteq \big\{ & x_0 \in \mathcal{X} \mid \exists t \in [0,\tau], \\ & \forall u(\cdot) \in \mathscr{U}_{[0,t]},\, x^u_{x_0}(t) \in \mathcal{K} \big\} \\ = & \bigcup_{t\in[0,\tau]} \big\{ x_0 \in \mathcal{X} \mid \forall u(\cdot) \in \mathscr{U}_{[0,t]},\, x^u_{x_0}(t) \in \mathcal{K} \big\}. \quad (2.11) \end{aligned}$$

Among Lagrangian methods, the technique in [72] has been extended to handle universally quantified inputs. However, since the time variable is quantified first, according to (2.11) the resulting set is an under-approximation of the unsafe minimal reachable tube.

The powerful Eulerian methods (e.g. [18, 33, 89, 106]), in addition to computing the maximal reachable set, are capable of directly computing the minimal reachable tube and the viability kernel. However, they rely on gridding the state space and are therefore computationally intensive. Although versatile in terms of ability to handle various types of dynamics and constraints, the applicability of these techniques has been historically limited to systems of low dimensionality (up to 3D or 4D in practice) due to their exponential complexity.

2.3 Main Goal and Problem Formulation

Given the fact that for guarantees of safety and the synthesis of safety-preserving controllers the minimal reachable tube and the viability kernel can play an extremely important role, we seek to devise efficient techniques that enable a more scalable computation of these constructs. To this end,

[4] Under-approximation of the viability kernel and over-approximation of the minimal reachable tube are the correct forms of approximation; See Section 2.7.

we present two separate approaches. The first approach, the structure decomposition techniques that will be presented in Chapters 3 and 4, aims to enable the use of powerful Eulerian methods on higher-dimensional LTI systems. The second approach, which is based on set-theoretic techniques, aims to facilitate the application of efficient and scalable Lagrangian methods for the computation of the viability kernel and safety-preserving controllers by drawing a connection between minimal and maximal reachability constructs. These objectives are defined next.

2.4 Approach I: Structure Decomposition

2.4.1 Objective and Problem Statement

Consider the case in which (2.1) is a continuous-time LTI system

$$\dot{x} = Ax + Bu \tag{2.12}$$

described by matrix notation

$$\mathcal{S} := \left[\begin{array}{c|c} A & B \end{array}\right] \tag{2.13}$$

with $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times p}$, and the constrained input $u(t) \in \mathcal{U}$ is the *control* input.

Problem 2.1 (Objective I)**.** *Find an appropriate basis transformation for* (2.12) *such that in the new coordinates the system can be decomposed into lower-dimensional, decoupled or weakly-coupled subsystems for which reachability analysis can be performed independently and thus more efficiently.*

2.4.2 Definitions and Preliminaries

A linear transformation of the system $\mathcal{S}$ in (2.13) using a nonsingular matrix $T \in \mathbb{R}^{n\times n}$ is defined as $\mathcal{S}' = T^{-1}(\mathcal{S}) := \left[\begin{array}{c|c} T^{-1}AT & T^{-1}B \end{array}\right]$. A linear transformation of a set $\mathcal{V} \subseteq \mathbb{R}^n$ under the same mapping is defined as $\mathcal{Y} = T^{-1}\mathcal{V} := \{y \in \mathbb{R}^n \mid y = T^{-1}v,\ v \in \mathcal{V}\}$.

Definition 2.5 (Unidirectionally Coupled)**.** *The LTI system that consists of two subsystems*

$$\dot{x}_1 = A_1 x_1 + \Delta\, x_2, \tag{2.14}$$

$$\dot{x}_2 = A_2 x_2 \tag{2.15}$$

with $A_1 \in \mathbb{R}^{k\times k}$, $A_2 \in \mathbb{R}^{(n-k)\times(n-k)}$, $\Delta \in \mathbb{R}^{k\times(n-k)}$, $x_1(t) \in \mathbb{R}^k$, *and* $x_2(t) \in \mathbb{R}^{n-k}$, *is said to be* unidirectionally coupled *since the trajectories of* (2.14) *are affected by those of* (2.15)*, while* (2.15) *evolves independently from* (2.14)*. The worst-case unidirectional coupling can thus be characterized by the maximum row sum* $\|\Delta\|$*.*

Definition 2.6 (Unidirectionally Weakly-Coupled)**.** *Let there be a nonsingular transformation matrix* $T \in \mathbb{R}^{n\times n}$*, such that* $[z_1^T, z_2^T]^T = T^{-1}[x_1^T, x_2^T]^T$*, and*

$$\dot{z}_1 = A_1 z_1 + \widetilde{\Delta}\, z_2 \tag{2.16}$$

$$\dot{z}_2 = A_2 z_2. \tag{2.17}$$

Then (2.16) *and* (2.17) *are said to be* unidirectionally weakly-coupled *(in comparison to* (2.14) *and* (2.15)*) if*

$$\|\widetilde{\Delta}\| \leq \|\Delta\|. \tag{2.18}$$

Definition 2.7 (Disjoint Input)**.** *Let there be a nonsingular transformation matrix* $T \in \mathbb{R}^{n\times n}$ *and a coordinate space* $z = T^{-1}x$ *in which* (2.12) *can be partitioned into* N *subsystems as*

$$\dot{z}_i = \widetilde{A}_i z_i + g_i(u), \quad i = 1, \ldots, N, \tag{2.19}$$

with $g_i(u) := \widetilde{B}_i u$*. The input* $u = [u_1, \ldots, u_p]^T \in \mathcal{U} \subset \mathbb{R}^p$ *is* disjoint *across these subsystems if* $\forall s \in \{1, \ldots, p\}$, $\forall i, j \in \{1, \ldots, N\}$, $i \neq j$,

$$\frac{\partial g_i(u)}{\partial u_s} \neq 0 \;\rightarrow\; \frac{\partial g_j(u)}{\partial u_s} = 0. \tag{2.20}$$

Example 2.1. *Consider the subsystem trajectories of* $\dot{z}_i = \widetilde{A}_i z_i + \widetilde{B}_i u$ *with* $\widetilde{B}_i = [\widetilde{B}_{i1}, \widetilde{B}_{i2}, \widetilde{B}_{i3}]$, $u = [u_1, u_2, u_3]^T$, *and* $i \in \{1, 2\}$. *The input* u *is disjoint if for example*

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} e^{\widetilde{A}_1(t-t_0)} & \mathbf{0} \\ \mathbf{0} & e^{\widetilde{A}_2(t-t_0)} \end{bmatrix} \begin{bmatrix} z_1(t_0) \\ z_2(t_0) \end{bmatrix} + \int_{t_0}^{t} \begin{bmatrix} e^{\widetilde{A}_1(t-r)} & \mathbf{0} \\ \mathbf{0} & e^{\widetilde{A}_2(t-r)} \end{bmatrix} \begin{bmatrix} \widetilde{B}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widetilde{B}_{22} & \widetilde{B}_{23} \end{bmatrix} \begin{bmatrix} u_1(r) \\ u_2(r) \\ u_3(r) \end{bmatrix} dr. \quad (2.21)$$

This ensures that no input from the input vector u *occurs in both subsystems. On the other hand, the input in the following equation is non-disjoint since* u_2 *influences the trajectories of both subsystems:*

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} e^{\widetilde{A}_1(t-t_0)} & \mathbf{0} \\ \mathbf{0} & e^{\widetilde{A}_2(t-t_0)} \end{bmatrix} \begin{bmatrix} z_1(t_0) \\ z_2(t_0) \end{bmatrix} + \int_{t_0}^{t} \begin{bmatrix} e^{\widetilde{A}_1(t-r)} & \mathbf{0} \\ \mathbf{0} & e^{\widetilde{A}_2(t-r)} \end{bmatrix} \begin{bmatrix} \widetilde{B}_{11} & \widetilde{B}_{12} & \mathbf{0} \\ \mathbf{0} & \widetilde{B}_{22} & \widetilde{B}_{23} \end{bmatrix} \begin{bmatrix} u_1(r) \\ \boxed{u_2(r)} \\ u_3(r) \end{bmatrix} dr. \quad (2.22)$$

Definition 2.8 (ETUC)**.** *A subsystem is said to be* externally-trivially-uncontrollable *(ETUC) if it possesses a null input matrix.*

Finally, consider the following two lemmas.

Lemma 2.1 ([123, Lem. 2.7])**.** *The Sylvester equation*

$$EX + XF + H = \mathbf{0}, \quad (2.23)$$

with $E \in \mathbb{R}^{k \times k}$, $F \in \mathbb{R}^{m \times m}$, *and* $H \in \mathbb{R}^{k \times m}$, *has a solution* $X \in \mathbb{R}^{k \times m}$ *if and only if*

$$\operatorname{rank} \begin{bmatrix} (F^T \otimes I_k) + (I_m \otimes E) & -\operatorname{vec}(H) \end{bmatrix} = \operatorname{rank} \begin{bmatrix} (F^T \otimes I_k) + (I_m \otimes E) \end{bmatrix}, \quad (2.24)$$

where $\otimes$ *denotes the Kronecker product and* $\mathrm{vec}(H)$ *is a vector formed by stacking the columns of* H *below one another. This solution is unique if and only if the eigenvalue sum* $\lambda_i(E)+\lambda_j(F) \neq 0$, $\forall i \in \{1,...,k\}$, $\forall j \in \{1,...,m\}$.

Lemma 2.2 (Real Schur Form [43, Thm's 7.1.3 and 7.4.1], [114, 5R])**.** *For any real matrix* $M \in \mathbb{R}^{n\times n}$ *there exists an orthogonal matrix* $U \in \mathbb{R}^{n\times n}$ *such that* $U^T M U = \widetilde{M}$ *is real upper quasi-triangular, and the eigenvalues of* M *are the eigenvalues of the block diagonals (each of dimension 2 or less) of* $\widetilde{M}$*. Furthermore, the matrix* U *can be chosen to order the eigenvalues arbitrarily.*

Remark 2.1. *There always exists a partitioning of* $\widetilde{M}$ *such that* $\widetilde{M} = \begin{bmatrix} \widetilde{M}_{11} & \widetilde{M}_{12} \\ \mathbf{0} & \widetilde{M}_{22} \end{bmatrix}$*. The size of the partitions can be chosen as desired, so long as each block diagonal entry (maximum size* 2×2*) of* $\widetilde{M}$ *is completely covered by exactly one of the blocks on the diagonal of the partitioned matrix.*

2.5 Approach II: Set-Theoretic Methods

2.5.1 Objective and Problem Statement

For the generic, continuously-valued system (2.1) we define our second objective as follows.

Problem 2.2 (Objective II)**.** *Express the viability kernel* $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ *in terms of the maximal reachable sets* $Reach_t^\sharp(\mathcal{K},\mathcal{U})$*,* $t \in [0,\tau]$ *to enable the use of Lagrangian methods for the computation of the viability kernel and the synthesis of safety-preserving controllers.*

2.5.2 Definitions and Preliminaries

In this section we will present the definitions required for the treatment of the case in which (2.1) is a continuous-time system (cf. Section 5.1.1) of the form

$$\dot{x} = f(x,u). \tag{2.25}$$

Definition 2.9. *We say that a vector field $f\colon \mathcal{X} \times \mathcal{U} \to \mathcal{X}$ is bounded on $\mathcal{K}$ if there exists a norm $\|\cdot\|\colon \mathcal{X} \to \mathbb{R}^+$ and a real number $M > 0$ such that for all $x \in \mathcal{K}$ and $u \in \mathcal{U}$ we have $\|f(x,u)\| \leq M$.*

Definition 2.10. *A partition $P = \{t_0, t_1, \ldots, t_n\}$ of $[0,\tau]$ is a set of distinct points $t_0, t_1, \ldots, t_n \in [0,\tau]$ with $t_0 = 0$, $t_n = \tau$ and $t_0 < t_1 < \cdots < t_n$. Further, we denote*

- *the number n of intervals $[t_{k-1}, t_k]$ in P by $|P|$,*
- *the size of the largest interval by $\|P\| := \max_{k=1}^{|P|}\{t_{k+1} - t_k\}$, and*
- *the set of all partitions of $[0,\tau]$ by $\mathscr{P}([0,\tau])$.*

Definition 2.11. *For a signal $u\colon [0,\tau] \to \mathcal{U}$ and a partition $P = \{t_0, \ldots, t_n\}$ of $[0,\tau]$, define the tokenization of u corresponding to P as the set of functions $\{u_k\colon [0, t_k - t_{k-1}] \to \mathcal{U}\}_k$ such that*

$$u_k(t) = u(t + t_{k-1}). \tag{2.26}$$

Conversely, for a set of functions $\{u_k\colon [0, t_k - t_{k-1}] \to \mathcal{U}\}_k$, define their concatenation $u\colon [0,\tau] \to \mathcal{U}$ as

$$u(t) = u_k(t - t_{k-1}), \quad t \in [t_{k-1}, t_k]. \tag{2.27}$$

Definition 2.12. *The $\|\cdot\|$-distance of a point $x \in \mathcal{X}$ from a nonempty set $\mathcal{A} \subset \mathcal{X}$ is defined as*

$$\operatorname{dist}(x, \mathcal{A}) := \inf_{a \in \mathcal{A}} \|x - a\|. \tag{2.28}$$

For a fixed set $\mathcal{A}$, the map $x \mapsto \operatorname{dist}(x, \mathcal{A})$ is continuous.

2.6 Robust Reachability Analysis: Systems with Competing Inputs

While the majority of this thesis deals with systems with a single input, there are instances where we will make use of the differential game framework and

assume adversarial inputs. For example, an "artificial" disturbance input is considered in Section 3.3 to perform a robust reachability analysis for one of the subsystems of the original deterministic system. Another example is when we extend the results presented in Chapter 5 to compute the robust version of the viability kernel (known as the *discriminating kernel* [18]) for systems with competing inputs in Chapter 6. Therefore in this section we will briefly lay out the definitions and preliminaries concerning systems with not only control input, but also uncertainties and/or external disturbances.

Consider the continuous-time system

$$\dot{x}(t) = f(x(t), u(t), v(t)), \quad x(0) = x_0 \tag{2.29}$$

with disturbance input $v(t) \in \mathcal{V}$ where $\mathcal{V}$ is a compact convex subset of $\mathbb{R}^{m_v}$. The vector field $f\colon \mathcal{X} \times \mathcal{U} \times \mathcal{V} \to \mathcal{X}$ is assumed to be Lipschitz in x and continuous in both u and v. Let $\mathscr{U}_{[0,t]}$ be as in (2.2) and define

$$\mathscr{V}_{[0,t]} := \{v\colon [0,t] \to \mathbb{R}^{m_v} \text{ measurable}, \;\; v(t) \in \mathcal{V} \text{ a.e.}\}. \tag{2.30}$$

For every $t \in [0,\tau]$, $x_0 \in \mathcal{X}$, $u(\cdot) \in \mathscr{U}_{[0,t]}$, and $v(\cdot) \in \mathscr{V}_{[0,t]}$, there exists a unique trajectory $x_{x_0}^{u,v}\colon [0,t] \to \mathcal{X}$ that satisfies the initial condition $x_{x_0}^{u,v}(0) = x_0$ and the differential equation (2.29). As before, when clear from the context we shall drop the subscript and superscript from the trajectory notation.

We assume that the disturbance input v is unknown but takes values on the (bounded) set $\mathcal{V}$. Note that $v \in \mathcal{V}$ can also be used to capture any (unknown but bounded) uncertainties in the model.

In a differential game setting the information pattern between the players (i.e. control input u and disturbance input v) is important and must be specified. The control input follows a feedback strategy, i.e. $u(t) = \hat{u}(x(t), t) \in \mathcal{U}$ $\forall t \in [0,\tau]$. We assume *non-anticipative* strategies for the disturbance input. A map $\rho\colon \mathscr{U}_{[0,t]} \to \mathscr{V}_{[0,t]}$ is non-anticipative for v if for every $u(\cdot), u'(\cdot) \in \mathscr{U}_{[0,t]}$, $u(s) = u'(s)$ for a.e. $s \in [0,t]$ implies $\rho[u](s) = \rho[u'](s)$ for a.e. $s \in [0,t]$ [27]. This results in a conservative formulation of our desired constructs by

giving v a slight advantage over u. (cf. [89] for more detail.)

Definition 2.13 (Robust Maximal Reachable Set)**.** *The robust maximal reachable set at time t is the set of initial states for which there exists a control for every disturbance such that the trajectories emanating from those states reach $\mathcal{K}$ exactly at time t:*

$$Reach_t^{\sharp}(\mathcal{K},\mathcal{U},\mathcal{V}) := \{x_0 \in \mathcal{X} \mid \exists u(\cdot) \in \mathscr{U}_{[0,t]}, \\ \forall \rho\colon \mathscr{U}_{[0,t]} \to \mathscr{V}_{[0,t]},\, x_{x_0}^{u,\rho[u]}(t) \in \mathcal{K}\}. \quad (2.31)$$

Definition 2.14 (Robust Minimal Reachable Tube)**.** *The robust minimal reachable tube over the horizon $[0,\tau]$ is the set of initial states for which there exists a disturbance for every control such that the trajectories emanating from those states reach $\mathcal{K}$ at some time $t \in [0,\tau]$:*

$$Reach_{[0,\tau]}^{\flat}(\mathcal{K},\mathcal{U},\mathcal{V}) := \{x_0 \in \mathcal{X} \mid \exists \rho\colon \mathscr{U}_{[0,\tau]} \to \mathscr{V}_{[0,\tau]}, \\ \forall u(\cdot) \in \mathscr{U}_{[0,\tau]},\, \exists t \in [0,\tau],\, x_{x_0}^{u,\rho[u]}(t) \in \mathcal{K}\}. \quad (2.32)$$

Definition 2.15 (Discriminating Kernel)**.** *The (finite-horizon) discriminating kernel*[5] *of $\mathcal{K}$ is the set of all initial states in $\mathcal{K}$ for which there exists a control such that the trajectories emanating from those states remain within $\mathcal{K}$ for every disturbance for all time $t \in [0,\tau]$:*

$$Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V}) := \big\{x_0 \in \mathcal{X} \mid \exists u(\cdot) \in \mathscr{U}_{[0,\tau]},\, \forall \rho\colon \mathscr{U}_{[0,\tau]} \to \mathscr{V}_{[0,\tau]}, \\ \forall t \in [0,\tau],\, x_{x_0}^{u,\rho[u]}(t) \in \mathcal{K}\big\}. \quad (2.33)$$

Note that with $\mathcal{V} = \{0\}$ the discriminating kernel reduces to the viability kernel under the deterministic system $\dot{x} = f(x,u)$:

$$Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\{0\}) \equiv Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}). \quad (2.34)$$

Finally, we will assume that the Isaac's condition holds and therefore

[5] The infinite-horizon discriminating kernel $Disc_{\mathbb{R}^+}(\mathcal{K},\mathcal{U},\mathcal{V})$ is also known as the *robust maximal controlled-invariant subset* [11].

the players' order can be interchanged. As such, the discriminating kernel and the robust minimal reachable tube are duals of one another:

$$Disc_{[0,\tau]}(\mathcal{K}^c, \mathcal{U}, \mathcal{V}) = (Reach^{\flat}_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V}))^c. \tag{2.35}$$

2.7 To Over- or Under-Approximate?

Any approximations of the minimal reachable tube (or its robust version) must be an *over-approximation* since the target set $\mathcal{K}$ for this construct is generally deemed unsafe. The minimal reachable tube is the set of states that can become unsafe regardless of the control input applied. An under-approximation of this set would exclude states for which such property holds, falsely labeling them as safe.

In contrast, the viability (or the discriminating) kernel must be *under-approximated* to ensure that every trajectory initiating from this set stays viable in $\mathcal{K}$. The constraint set $\mathcal{K}$ for this construct is generally associated with safety. The states that belong to the viability kernel must retain the property that for each of them there exists at least one control policy that can keep the trajectory of the system in $\mathcal{K}$. Over-approximating the viability kernel would include states for which this property does not hold, falsely labeling them as safe.

The approximative techniques we present in this thesis will ensure that the minimal reachable tube is always over-approximated and that the viability kernel is guaranteed to be under-approximated.

Chapter 3

Schur-Based Structure Decomposition[1]

The first decomposition technique we present to address Problem 2.1 is inspired by a model reduction algorithm for systems with unstable modes [82, 110]. Our method decomposes LTI systems into either completely decoupled or weakly-coupled subsystems. Reachability analysis can be performed on each resulting subsystem independently. Back projecting and intersecting each of the lower-dimensional reachable tubes provides an over-approximation of the actual minimal reachable tube. A Sylvester equation (or an optimization problem) is solved in order to eliminate (or minimize) the coupling between the subsystems. Additional constraints are imposed when the control input is non-disjoint across subsystems, to prevent under-approximation of the (unsafe) minimal reachable tube. In addition, at the end of this section we will also provide conditions under which a subspace reachable tube remains unchanged for all time and show how this can be used in conjunction with the proposed Schur-based decomposition technique to yield an even further reduction of complexity for a class of systems.

Outline Via Lemma 2.2, as in [104], we obtain an upper block triangular A-matrix for (2.13). We then perform a second similarity transformation and obtain a decoupled (or weakly-coupled) block diagonal matrix by solving a Sylvester equation (or an optimization problem). Therefore, we effectively decompose the system into two either completely decoupled or unidirectionally weakly-coupled subsystems. In the case where the decom-

[1] A version of this chapter has been published in [57, 58].

position is decoupled, the reachable tube is computed separately for each isolated subsystem. When the decomposed subsystems are unidirectionally weakly-coupled, the reachable tube is computed independently for the isolated subsystem, whereas for the remaining subsystem, the effect of coupling is accounted for by treating the coupling terms as disturbance and performing reachability with competing inputs. For both decoupled and unidirectionally weakly-coupled decompositions, the intersection of back projections of the lower dimensional reachable tubes is an over-approximation of the actual reachable tube in the transformed coordinate space. When the control input across the decomposed subsystems is non-disjoint, a constrained optimization problem is solved in order to make one of the subsystems ETUC.

In the following analysis, we assume a partitioning of (2.13) that results in exactly two subsystems. However, the proposed method is generalizable to N subsystems by applying the same decomposition algorithm to each subsystem iteratively. A higher number of subsystems (i.e. iterated decomposition) may result in a more conservative over-approximation of the actual reachable tube.

For $k < n$, we now apply Lemma 2.2 with transformation matrix $U \in \mathbb{R}^{n\times n}$ to (2.13) to obtain

$$\left[\begin{array}{cc|c} \widetilde{A}_{11} & \widetilde{A}_{12} & \widetilde{B}_1 \\ \mathbf{0} & \widetilde{A}_{22} & \widetilde{B}_2 \end{array}\right] \tag{3.1}$$

with $\widetilde{A}_{11} \in \mathbb{R}^{k\times k}$, $\widetilde{A}_{12} \in \mathbb{R}^{k\times(n-k)}$, $\widetilde{A}_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$, $\widetilde{B}_1 \in \mathbb{R}^{k\times p}$, and $\widetilde{B}_2 \in \mathbb{R}^{(n-k)\times p}$.

3.1 Disjoint Control Input

Consider the case in which the control input is disjoint across candidate subsystems.

Proposition 3.1. *If there exists a solution $X \in \mathbb{R}^{k\times(n-k)}$ to the Sylvester equation*

$$\widetilde{A}_{11}X - X\widetilde{A}_{22} + \widetilde{A}_{12} = \mathbf{0}, \tag{3.2}$$

then a transformation

$$W = \begin{bmatrix} I_k & X \\ \mathbf{0} & I_{n-k} \end{bmatrix} \in \mathbb{R}^{n \times n} \tag{3.3}$$

makes (3.1) *completely decoupled.*

Proof. cf. [82, 104, 110]. Applying the (invertible) transformation W to (3.1), we obtain

$$\left[\begin{array}{cc|c} \widetilde{A}_{11} & \widetilde{A}_{11}X - X\widetilde{A}_{22} + \widetilde{A}_{12} & \widehat{B}_1 \\ \mathbf{0} & \widetilde{A}_{22} & \widehat{B}_2 \end{array}\right] = \left[\begin{array}{cc|c} \widetilde{A}_{11} & \mathbf{0} & \widehat{B}_1 \\ \mathbf{0} & \widetilde{A}_{22} & \widehat{B}_2 \end{array}\right]. \tag{3.4}$$

□

Notice that the resulting subsystems $\left[\begin{array}{c|c} \widetilde{A}_{11} & \widehat{B}_1 \end{array}\right]$ and $\left[\begin{array}{c|c} \widetilde{A}_{22} & \widehat{B}_2 \end{array}\right]$ have been effectively decoupled through the coordinate transformation $z = T^{-1}x$, $T = UW$. As we shall see in Section 3.3, reachability analysis (in this transformed coordinate space) can then be performed on each lower-dimensional subsystem separately.

Now consider the case in which there is no solution to the Sylvester equation (3.2).

Proposition 3.2. *If* (3.2) *does not have a solution, then the transformation* (3.3) *with*

$$X = \underset{Q \in \mathbb{R}^{k \times (n-k)}}{\arg\min} \; \|\widetilde{A}_{11}Q - Q\widetilde{A}_{22} + \widetilde{A}_{12}\| \tag{3.5}$$

results in unidirectionally weakly-coupled subsystems w.r.t. (3.1)*.*

Proof. Consider $A_c := \widetilde{A}_{11}X - X\widetilde{A}_{22} + \widetilde{A}_{12} \neq \mathbf{0}$ in (3.4). It is clear that in the transformed coordinate space characterized by $z = (UW)^{-1}x$, the state vector z_2 evolves independently of z_1 since $\dot{z}_2 = \widetilde{A}_{22}z_2 + \widehat{B}_2u_2$. However, z_1 is affected by z_2 through A_c. That is, we have $\dot{z}_1 = \widetilde{A}_{11}z_1 + \widehat{B}_1u_1 + A_cz_2$. (Note that u_i, $i \in \{1, 2\}$, is the effective portion of the input vector u for the ith subsystem.) Minimization of the infinity norm of A_c therefore translates into minimizing, i.e. *weakening*, the worst-case unidirectional coupling of

z_1 with z_2. To see this, let $X^* = \arg\min\{\|\widetilde{A}_{11}Q - Q\widetilde{A}_{22} + \widetilde{A}_{12}\| \mid Q \in \mathbb{R}^{k\times(n-k)}\}$. Then the hypothesis $\|\widetilde{A}_{12}\| < \|\widetilde{A}_{11}X^* - X^*\widetilde{A}_{22} + \widetilde{A}_{12}\|$ would imply that $X^* = \mathbf{0}$ can never be a solution. Since there are no constraints in (3.5) imposing this restriction, by contradiction we conclude that $\|\widetilde{A}_{11}X^* - X^*\widetilde{A}_{22} + \widetilde{A}_{12}\| \leq \|\widetilde{A}_{12}\|$. Therefore, according to Definition 2.6, the resulting subsystems are unidirectionally weakly-coupled. □

Remark 3.1. *The objective function of* (3.5) *is convex and therefore a solution always exists.*

Remark 3.2. *The main rationale behind minimizing the* infinity norm *of the unidirectional coupling term (and thus, obtaining unidirectionally-weakly coupled subsystems) is that for the purpose of reachability analysis, the infinity norm of this term will be used to formulate an upper-bound on the magnitude of the disturbance to the upper subsystem. This will be discussed further in Section 3.3.*

3.2 Non-Disjoint Control Input

Now consider a decomposition in which the control input is non-disjoint. In this case even if the dynamics of the subsystems are completely decoupled, their evolution is tightly paired through a common input. The difficulty arises, for example, when in the reachability computation a control value deemed optimal for one subsystem is in fact non-optimal for the full-order system. Blindly performing reachability for each subsystem separately may result in an under-approximation and additional measures have to be taken to ensure the over-approximation of the actual (unsafe) minimal reachable tube.

One way to remedy this issue is by ensuring that at least one of the subsystems in the transformed coordinate space is ETUC. It is clear that in such a case the (otherwise non-disjoint) control action does not affect the evolution of the reachable tube of the ETUC subsystem. Therefore, an optimal control input for the subsystem with nonzero input matrix is also optimal for the full-order system.

More formally, if either the pair $(\widetilde{A}_{22}, \widehat{B}_2)$ or the pair $(\widetilde{A}_{11}, \widehat{B}_1)$ in (3.4) is made ETUC, reachability analysis can be performed as in the disjoint control input case, separately for each subsystem.

Assumption 3.1. *$\mathscr{C}(\widetilde{B}_1^T) \subseteq \mathscr{C}(\widetilde{B}_2^T)$ with $\mathscr{C}(\cdot)$ the column-space operator.*

Proposition 3.3. *The transformation* (3.3) *with*

$$X = \underset{Q \in \mathbb{R}^{k\times(n-k)}}{\arg\min} \|\widetilde{A}_{11}Q - Q\widetilde{A}_{22} + \widetilde{A}_{12}\| \quad \text{subject to } Q\widetilde{B}_2 = \widetilde{B}_1 \tag{3.6}$$

results in unidirectionally coupled subsystems. Moreover, $(\widetilde{A}_{11}, \widehat{B}_1)$ is ETUC.

Proof. Assumption 3.1 is the necessary and sufficient condition for solvability of the overdetermined equality constraint in (3.6) (cf. Appendix A.1). To see the trivial-uncontrollability of $(\widetilde{A}_{11}, \widehat{B}_1)$ consider $\widehat{B} := W^{-1}\widetilde{B}$ in (3.4). We have

$$\begin{bmatrix} \widehat{B}_1 \\ \widehat{B}_2 \end{bmatrix} := \begin{bmatrix} I_k & -X \\ \mathbf{0} & I_{n-k} \end{bmatrix} \begin{bmatrix} \widetilde{B}_1 \\ \widetilde{B}_2 \end{bmatrix} = \begin{bmatrix} \widetilde{B}_1 - X\widetilde{B}_2 \\ \widetilde{B}_2 \end{bmatrix}. \tag{3.7}$$

Constraining the optimizer in (3.6) to choose from the class of solutions $\left\{X \in \mathbb{R}^{k\times(n-k)} \mid X\widetilde{B}_2 = \widetilde{B}_1\right\}$ simply enforces $\widehat{B}_1 = \mathbf{0}$. □

The resulting subsystems can now be treated as in the disjoint control input case, and hence an over-approximation of the reachable tube in each subspace can be computed.

3.3 Reachability in Lower Dimensions

Denote by

$$\mathbb{S}_1 := \mathbb{R}^k \quad \text{and} \quad \mathbb{S}_2 := \mathbb{R}^{n-k} \tag{3.8}$$

the subspaces of $\mathbb{R}^n$ in which the two subsystems evolve. In the new coordinate space $z = T^{-1}x$, $T := UW$ reachability analysis can be performed on each lower-dimensional subsystem separately:

Algorithm 3.1 Reachability in lower dimensions (Schur-Based)

1: $\mathcal{Z}_\tau \leftarrow T^{-1}\mathcal{K}$
2: $\mathcal{Z}_\tau^i \leftarrow \text{Proj}_{\mathbb{S}_i}(\mathcal{Z}_\tau), \quad \forall i \in \{1,2\}$ ▷ project onto ith subspace
For lower subsystem:
3: $\mathcal{Z}^2_{[0,\tau]} \leftarrow Reach^{\flat}_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{U})$
For upper subsystem:
4: Treat $A_c z_2$ as disturbance ▷ $A_c := \widetilde{A}_{11}X - X\widetilde{A}_{22} + \widetilde{A}_{12}$
5: $\zeta \leftarrow \|z_2\| \equiv \sup_{v \in \mathcal{Z}^2_{[0,\tau]}} \|v\|$
6: Compute upper-bound $\|A_c z_2\| \leq \|A_c\|\zeta$
7: $\mathcal{Z}^1_{[0,\tau]} \overset{\text{consrv.}}{\longleftarrow} Reach^{\flat}_{[0,\tau]}(\mathcal{Z}^1_\tau, \mathcal{U}, \mathcal{B}(\|A_c\|\zeta))$
8: **return**$(\mathcal{Z}^1_{[0,\tau]}, \mathcal{Z}^2_{[0,\tau]})$

Note that steps 4 through 6 of Algorithm 3.1 may or may not be needed depending on whether the subsystems are obtained from Propositions 3.1, 3.2, or 3.3. The following scenarios describe how the input(s) are quantified to construct the subsystem reachable tubes:

S1 (Proposition 3.1 is used): For both $\mathcal{Z}^1_{[0,\tau]}$ and $\mathcal{Z}^2_{[0,\tau]}$, the single input is control and it is universally quantified.

S2 (Proposition 3.2 is used): For $\mathcal{Z}^1_{[0,\tau]}$ the control input is universally quantified while the disturbance input (unidirectional coupling) is existentially quantified. For $\mathcal{Z}^2_{[0,\tau]}$ the single input is control and it is universally quantified.

S3 (Proposition 3.3 is used): For $\mathcal{Z}^1_{[0,\tau]}$ the single input is disturbance (unidirectional coupling) and it is existentially quantified. Note that in this case the robust minimal reachability operator in step 7 of the algorithm is effectively replaced by $Reach^{\sharp}_{[0,\tau]}(\mathcal{Z}^1_\tau, \mathcal{B}(\|A_c\|\zeta))$. For $\mathcal{Z}^2_{[0,\tau]}$ the single input is control and it is universally quantified.

The over-approximation of the actual minimal reachable tube of the full-order system in $\mathcal{X}$ can be obtained using the following lemma.

Lemma 3.1 ([91, 113])**.** *Let $\mathcal{Z}^i_{[0,\tau]}$, $i \in \{1,2\}$, be the computed lower-dimensional over-approximative reachable tube of subsystem i. Then the inverse transformation of the intersection of the back-projection of these sets onto $\mathbb{R}^n$ is a guaranteed over-approximation of the actual full-order reachable tube $Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ of the system* (2.13)*:*

$$T\left((\mathcal{Z}^1_{[0,\tau]} \times \mathbb{S}_2) \cap (\mathbb{S}_1 \times \mathcal{Z}^2_{[0,\tau]})\right) \supseteq Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}). \tag{3.9}$$

3.3.1 Formulating an Upper-Bound on Growth of $\mathcal{Z}^1_{[0,\tau]}$ in Scenario S3

When the subsystems are obtained via Proposition 3.3, the reachable tube in the subspace of the ETUC subsystem is computed without the need for solving a differential game. In fact, for this subsystem the unidirectional coupling is treated as disturbance and, therefore, it is existentially quantified. Consequently, this disturbance together with the dynamics strive to enlarge the reachable (unsafe) set as much as possible. This allows us to formulate an analytic upper-bound on the over-approximation of the reachable tube in this subspace in terms of system and design parameters:

Let $\bar{z}_1 \in \mathcal{Z}^1_\tau$ and suppose $\mathscr{D}_{[0,\tau]}$ is the set of measurable functions from $[0,\tau]$ to $\mathcal{B}(\|A_c\|\zeta)$. There exists an admissible input $\vartheta(\cdot) \in \mathscr{D}_{[0,\tau]}$ such that (using time-reversed dynamics) we have

$$z_1 := e^{-\widetilde{A}_{11}\tau}\bar{z}_1 - \int_0^\tau e^{-\widetilde{A}_{11}(\tau-r)}\vartheta(r)dr, \tag{3.10}$$

$$z_1 \in \mathcal{Z}^1_{[0,\tau]}. \tag{3.11}$$

Bounding the effect of the input on the evolution of the trajectories we

obtain

$$\|z_1 - e^{-\widetilde{A}_{11}\tau}\bar{z}_1\| \leq \int_0^\tau e^{\|\widetilde{A}_{11}\|(\tau-r)}\|A_c\|\zeta dr \tag{3.12}$$

$$= \frac{e^{\|\widetilde{A}_{11}\|\tau} - 1}{\|\widetilde{A}_{11}\|}\|A_c\|\zeta \tag{3.13}$$

$$= \left(\lim_{M\to\infty}\sum_{i=1}^{M}\frac{\tau^i\|\widetilde{A}_{11}\|^{i-1}}{i!}\right)\|A_c\|\zeta \tag{3.14}$$

$$\leq \left(\lim_{M\to\infty}\sum_{i=1}^{M}\frac{\tau^i\big(\sqrt{k}\,\overline{\sigma}(\widetilde{A}_{11})\big)^{i-1}}{i!}\right)\|A_c\|\zeta \tag{3.15}$$

$$=: \mu_{[0,\tau]} \tag{3.16}$$

where $\overline{\sigma}(\cdot)$ is the largest singular value operator, and k is the dimension of the ETUC subsystem. Therefore, an upper-bound for how much $\mathcal{Z}^1_{[0,\tau]}$ can grow in backward time can be written as

$$\mathcal{Z}^1_{[0,\tau]} \subseteq \left(\bigcup_{t\in[0,\tau]} e^{-\widetilde{A}_{11}t}\mathcal{Z}^1_\tau\right) \oplus \mathcal{B}(\mu_{[0,\tau]}) \tag{3.17}$$

in which $\oplus$ denotes the Minkowski sum. In particular, the choice of k, the magnitude of the unidirectional coupling $\|A_c\|$, the supremum of the reachable tube in the lower subspace $\zeta = \sup_{v\in\mathcal{Z}^2_{[0,\tau]}}\|v\|$, and the largest singular value of the upper subsystem $\overline{\sigma}(\widetilde{A}_{11})$ can all affect the conservatism of the reachable tube $\mathcal{Z}^1_{[0,\tau]}$. Moreover, given k and τ, the flexibility of the Schur form in placing the eigenvalues in any order along the block-diagonals of $\widetilde{A}$ can be exploited to make this subsystem evolve with slower dynamics. Through various tests we were able to confirm that doing so could potentially prevent the excessive growth of $\mathcal{Z}^1_{[0,\tau]}$ by influencing both $e^{-\widetilde{A}_{11}t}$ and $\mu_{[0,\tau]}$.

3.4 Further Reduction of Complexity in Reachability for a Class of Unstable Systems

We now demonstrate that for a specific class of unstable LTI systems, the Schur-based decomposition can be used to further reduce the computational burden associated with reachability analysis.

Particularly, we decompose any full-order unstable system into *stable* and *anti-stable* subsystems with disjoint input across them. To do this, we employ the presented Schur-based decomposition while rearranging the order of eigenvalues such that the lower (controlled) subsystem contains only the non-negative eigenvalues and the upper (uncontrolled and possibly perturbed) subsystem contains the strictly-negative ones. As we will show in Proposition 3.4, under certain conditions, reachability analysis in the anti-stable subspace need *not* be performed since the target and the reachable tubes coincide for all time.

Proposition 3.4. *Suppose that for a controlled linear system* (2.12) *the following conditions are satisfied.*

(i) $\mathcal{K}$ *is convex (but possibly arbitrarily shaped) and* $\mathbf{0} \in \mathcal{K}$*;*

(ii) the A-matrix is anti-stable (analytic in the open left-half complex plane) with repeated and real eigenvalues $\lambda_1 = \cdots = \lambda_n \geq 0$*;*

(iii) the algebraic and geometric multiplicities of $\lambda_i(A)$ *are equal.*

Then for any $\tau \in \mathbb{R}^+$*,*

$$Reach^{\flat}_{[0,t]}(\mathcal{K}, \mathcal{U}) = \mathcal{K} \quad \forall t \in [0, \tau]. \tag{3.18}$$

Proof. The proof is provided in Appendix A.2. □

Remark 3.3. *Condition* (i) *is easily generalizable to star-convex sets for which the origin is the convergence point (any line segment from the origin to* $x \in \mathcal{K}$ *is contained in* $\mathcal{K}$*). An example of this is when the states are constrained to* l_p*-space with* $0 < p < 1$*.*

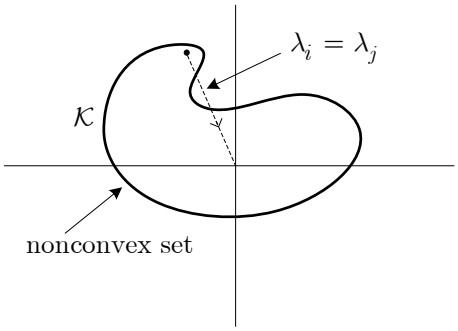


Figure 3.1: Phase-plane of a planar system with a non-convex target set $\mathcal{K}$. Even though conditions (ii) and (iii) are satisfied, the backward reachable tube will grow.

An intuitive 2-dimensional illustration of various cases that would violate conditions in Proposition 3.4 is given in Figures 3.1 and 3.2 where the trajectories are shown in backward time.

Note that although Proposition 3.4 is stated in terms of a general full-order system (and as such, may seem too restrictive), it makes the following assertion:

Corollary 3.1. *If any isolated subsystem of any given unstable system in any coordinate space satisfies the conditions in Proposition 3.4, then the minimal reachable tube for that subsystem remains precisely equal to the target set in the respective subspace.*

Suppose that reachability analysis is to be performed for an unstable system $\dot{x} = Ax + Bu$, $u \in \mathcal{U}$ with k negative and $(n-k)$ non-negative eigenvalues for a target set $\mathcal{K}$. We apply Schur-based decomposition with an appropriately synthesized transformation matrix T to obtain

$$\left[\begin{array}{cc|c} \widetilde{A}_- & A_c & \mathbf{0} \\ \mathbf{0} & \widetilde{A}_+ & \widehat{B}_2 \end{array}\right] \tag{3.19}$$

partitioned such that $\widetilde{A}_+$ and $\widetilde{A}_-$ contain only non-negative and strictly-negative eigenvalues, respectively. If $\widetilde{A}_+$ and $\mathcal{K}$ satisfy the conditions (i),

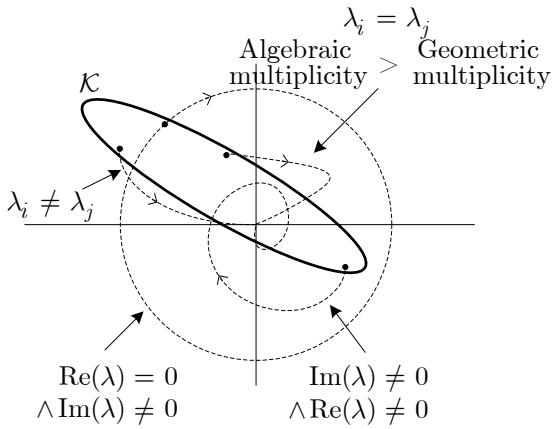


Figure 3.2: Phase-plane with various eigenvalue scenarios that would violate conditions of Proposition 3.4 and thus causing the backward reachable tube to grow.

(ii), and (iii), then according to Corollary 3.1 the reachable tube in the lower subspace does not grow and thus need not be computed. Reachability analysis is performed only for the upper subsystem resulting in further reduction of complexity by avoiding altogether the reachable tube computation in the lower subspace. Specifically, step 3 in Algorithm 3.1 is entirely omitted. The over-approximation of the full-order reachable tube can then be calculated according to (3.9) with $\mathcal{Z}^2_{[0,\tau]} = \mathit{Reach}^{\flat}_{[0,\tau]}(\mathrm{Proj}_{\mathbb{S}_2}(T^{-1}\mathcal{K}), \mathcal{U}) \equiv \mathrm{Proj}_{\mathbb{S}_2}(T^{-1}\mathcal{K})$ and $\mathcal{Z}^1_{[0,\tau]} = \mathit{Reach}^{\sharp}_{[0,\tau]}(\mathrm{Proj}_{\mathbb{S}_1}(T^{-1}\mathcal{K}), \mathcal{B}(\|A_c\|\zeta))$.

Note that linear transformation preserves convexity. Therefore the projection of the transformation of $\mathcal{K}$ onto the lower subspace (i.e. $\mathcal{Z}^2_\tau := \mathrm{Proj}_{\mathbb{S}_2}(T^{-1}\mathcal{K})$) is convex if $\mathcal{K}$ is, and contains the origin if $\mathcal{K}$ does.

3.5 Extension to Switched Linear Systems

The extension of our transformation-based method to switched dynamical systems is fairly straightforward. Consider the hybrid automaton $H = (\mathcal{Q}, \mathcal{X}, f, \mathcal{U}, E, R)$ with discrete modes $\mathcal{Q} = \{q_i\}$, continuous states $x \in \mathcal{X}$,

continuous control inputs $u \in \mathcal{U}$, vector fields

$$\begin{aligned} f\colon \mathcal{Q} \times \mathcal{X} \times \mathcal{U} &\to \mathcal{X}, \\ (q_i, x, u) &\mapsto A_i x + B_i u, \end{aligned} \tag{3.20}$$

edges $E \subseteq \mathcal{Q} \times \mathcal{Q}$, and (identity) reset maps $R\colon E \times \mathcal{X} \to 2^{\mathcal{X}}$.

Let $\mathcal{K}(q_i)$ and $Reach^{\flat}_{[0,\tau]}(q_i, \mathcal{K}(q_i), \mathcal{U})$ denote the target and the reachable tubes in mode $q_i \in \mathcal{Q}$, respectively. Also let T_i be the transformation matrix for mode q_i obtained from the Schur-based decomposition technique described previously. For simplicity of presentation we assume that H has only two modes q_i and q_j such that $(q_i, q_j) \in E$. The backward reachable tube in mode q_j can be directly expressed as

$$T_j Reach^{\flat}_{[0,\tau]} \left(q_j, T_j^{-1} T_i Reach^{\flat}_{[0,\tau]} \left(q_i, T_i^{-1} \mathcal{K}(q_i), \mathcal{U} \right), \mathcal{U} \right). \tag{3.21}$$

Reachability analysis is then performed on lower-dimensional subsystems in each mode according to Algorithm 3.1.

3.6 Numerical Examples

Although complexity reduction through Schur-based decomposition can be used in conjunction with any reachability/viability technique that can accommodate both existentially and universally quantified inputs, we demonstrate the applicability and practicality of our method using a number of examples (up to 8D) that employ the Level Set Toolbox (LS) [90]. While LS has mainly been used for systems of low dimensionality [8], our complexity reduction approach can facilitate the use of LS for higher dimensional systems for which safety-preserving controller synthesis and/or handling of non-convex, arbitrarily-shaped sets is important.

All computations are performed on a dual core Intel-based computer with 2.8 GHz CPU, 6 MB of cache and 3 GB of RAM running single-threaded 32-bit MATLAB 7.5.

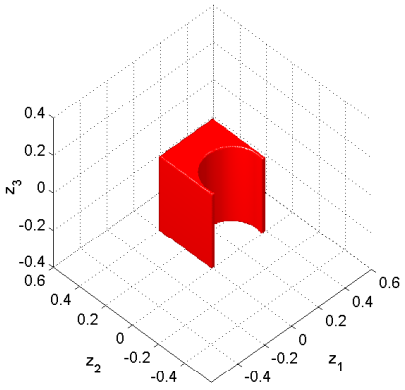


Figure 3.3: The non-convex target set $\mathcal{Z}_\tau$ in the transformed coordinate space.

3.6.1 Arbitrary 3D System

Consider an arbitrary 3D LTI system with

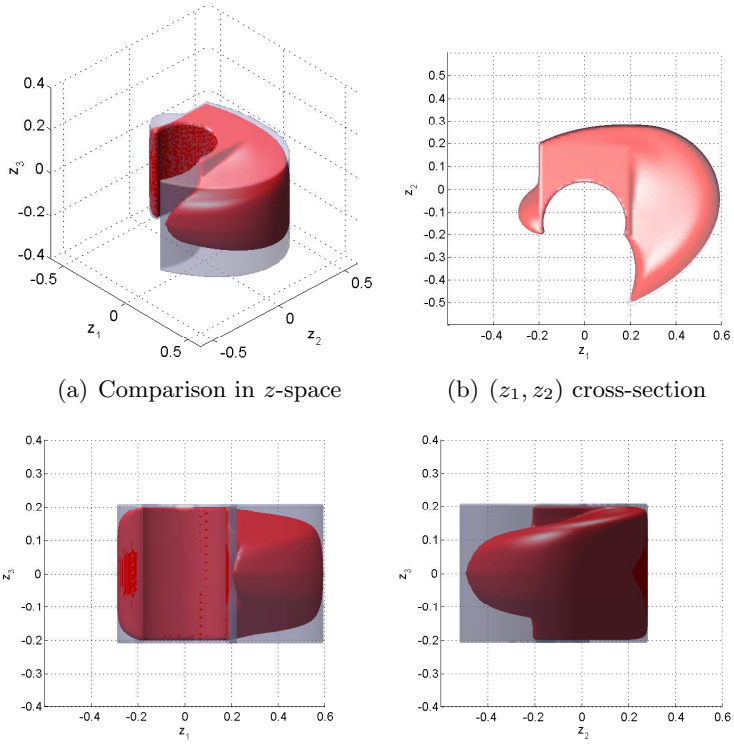
$$A = \begin{bmatrix} -0.5672 & -0.7588 & -0.6282 \\ 3.1364 & -1.1705 & 2.3247 \\ 1.8134 & -1.7689 & -2.6930 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0731 & -0.1639 \\ -0.7377 & -0.3578 \\ 0.1470 & 0.2410 \end{bmatrix}$$

and input $u = [u_1, u_2]^T \in \mathbb{R}^2$, $\|u\| \leq 1.1$. We choose a non-convex target (unsafe) set $\mathcal{K} \subset \mathbb{R}^3$ such that in the transformed coordinate space we have $\mathcal{Z}_\tau = T^{-1}\mathcal{K}$ as shown in Figure 3.3. Here, T is the transformation matrix obtained through Proposition 3.1 that decomposes the system into two subsystems (one 2D and one 1D) with disjoint control across them:

$$T^{-1}AT = \left[\begin{array}{cc:c} -1.6653 & -3.4560 & 0 \\ 1.8706 & -1.4653 & 0 \\ \hdashline 0 & 0 & -1.3000 \end{array}\right], \quad T^{-1}B = \left[\begin{array}{c:c} -0.7530 & 0 \\ 0.0640 & 0 \\ \hdashline 0 & 0.2500 \end{array}\right].$$

Hence, the decoupled subsystems are $\dot{z}_1 = \left[\begin{smallmatrix} -1.6653 & -3.4560 \\ 1.8706 & -1.4653 \end{smallmatrix}\right] z_1 + \left[\begin{smallmatrix} -0.7530 \\ 0.0640 \end{smallmatrix}\right] u_1$ and $\dot{z}_2 = \left[-1.3000\right] z_2 + \left[0.2500\right] u_2$.

We obtain an over-approximation of the actual reachable tube, as shown in Figure 3.4. Reachability calculation is performed over a grid with 101



(a) Comparison in z-space

(b) (z_1, z_2) cross-section

(c) (z_1, z_3) cross-section

(d) (z_2, z_3) cross-section

Figure 3.4: Schur-based over-approximation (transparent light) vs. actual (solid dark) reachable tubes in the transformed coordinate space for Example 3.6.1.

nodes in each dimension for $\tau = 2\,\mathrm{s}$. The computation time for the actual and the Schur-based reachable tubes (including decomposition and projections) were 5823.73 s and 22.87 s, respectively.

3.6.2 4D Aircraft Dynamics

Consider longitudinal aircraft dynamics $\dot{x} = Ax + B\delta_e$,

$$A = \begin{bmatrix} -0.0030 & 0.0390 & 0 & -0.3220 \\ -0.0650 & -0.3190 & 7.7400 & 0 \\ 0.0200 & -0.1010 & -0.4290 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0100 \\ -0.1800 \\ -1.1600 \\ 0 \end{bmatrix}$$

with state $x = [u, v, \dot{\theta}, \theta]^T \in \mathbb{R}^4$ comprised of deviations in aircraft velocity [ft/s] along and perpendicular to body axis, pitch-rate [crad/s], and pitch angle [crad] respectively[2], and with input $\delta_e \in [-13.3°, 13.3°] \subseteq \mathbb{R}$ the elevator deflection. These matrices represent stability derivatives of a Boeing 747 cruising at an altitude of 40 kft with speed 774 ft/s [16].

Define a non-convex target (unsafe) set $\mathcal{K}$ such that in the transformed coordinate space $\mathcal{Z}_\tau = \{z \in \mathbb{R}^4 \mid \|z\| > 0.15,\ z = T^{-1}x,\ x \in \mathcal{K}\}$ where T is the transformation matrix obtained through our method. We first decompose the system into two 2D subsystems. Since the control input is non-disjoint across the resulting subsystems, we use Proposition 3.3 and obtain unidirectionally coupled subsystems, one of which is ETUC (see Appendix A.3). The reachability calculation is performed over a grid with 41 nodes in each dimension for $\tau = 5$ s. The computation time for the actual and the Schur-based minimal reachable tubes (including decomposition and projections) were 28546.80 s and 54.64 s, respectively—a significant reduction.

Since the computed sets are 4D, we plot a series of 3D snapshots of these 4D objects at specific values of z_4 (Figure 3.5). The aircraft flight envelope (safe) is represented by the area inside of the shaded regions.

3.6.3 8D Distillation Column

Consider the dynamic model of a binary distillation column obtained from [111] with

[2] crad $= 0.01$ rad $\approx 0.57°$.

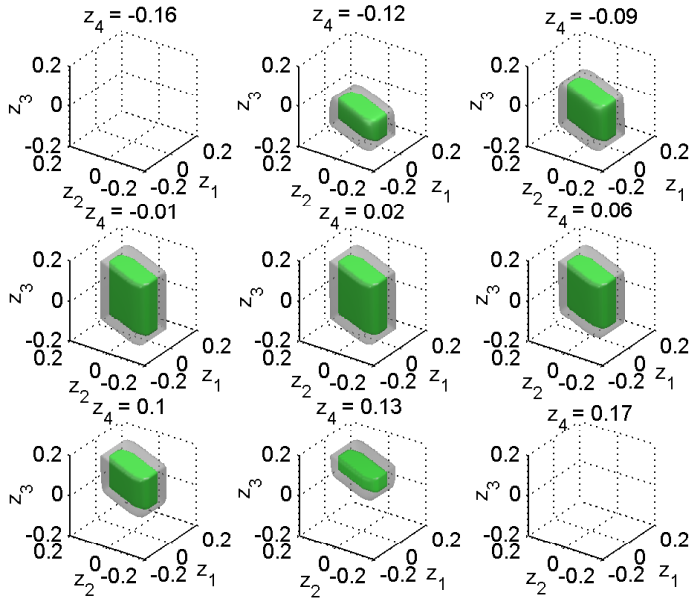


Figure 3.5: Schur-based (solid dark) vs. actual (transparent light) finite-horizon viability kernels (safe) in the transformed coordinate space for Example 3.6.2. The computed minimal reachable tube and its over-approximation are the non-convex complements of these objects.

$$A = \begin{bmatrix} -0.5774 & 3.0567 & 0.0073 & -0.8121 & 0.3034 & -0.3035 & 0.0072 & -0.1542 \\ -2.7290 & -0.7147 & -0.3430 & 1.5321 & 0.6643 & 0.2896 & -0.0013 & 0.0926 \\ 0 & 0 & -0.3891 & -0.9956 & 0.0182 & 0.0235 & 0.0049 & 0.0506 \\ 0 & 0 & 1.3640 & -1.3363 & -0.9037 & -0.4686 & -0.0009 & -0.1887 \\ 0 & 0 & 0 & 0 & -0.7357 & -0.2275 & -0.0082 & -0.0021 \\ 0 & 0 & 0 & 0 & 0 & -0.2259 & 0.0021 & -0.0457 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.0052 & 0.0024 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.0755 \end{bmatrix}$$

$$B = \begin{bmatrix} -0.0335 & -0.4534 & -0.8005 & 0.5497 & 1.2886 & 0.3132 & 0.7117 & 0.0599 \\ -0.1228 & -0.0711 & -0.2612 & -0.1344 & -0.0504 & -0.2249 & -0.6994 & -0.3014 \end{bmatrix}^T.$$

The input $u = [u_1, u_2]^T \in \mathbb{R}^2$ with $u_1, u_2 \in [0, 1]$ is comprised of reflux and boilup flows, respectively. The full-order system with state vector $x \in$

$\mathbb{R}^8$ is first decomposed into two (unidirectionally coupled) 4D subsystems using Proposition 3.3, since the control vector is non-disjoint across the two candidate subsystems. Similarly, each of these 4D subsystems is decomposed into two 2D subsystems. Since the upper 4D subsystem is made ETUC through (3.6), its decomposition is disjoint and therefore Proposition 3.1 is used to obtain the 1st and 2nd (decoupled) 2D subsystems. On the other hand, for the lower 4D subsystem the decomposition results in non-disjoint control input. Therefore Proposition 3.3 is employed and the 3rd and 4th (unidirectionally coupled) 2D subsystems are obtained (see Appendix A.3).

Reachability is first performed on the 3rd and 4th subsystems while taking the effect of unidirectional coupling into account. Next, the reachable tubes of the 1st and 2nd subsystems are computed while treating the effect of the 3rd and 4th subsystems as disturbance. We label the 2D transformed state subspaces as $\widetilde{w}_1 = [w_1, w_2]^T$, $\widetilde{w}_2 = [w_3, w_4]^T$, $\widetilde{q}_1 = [q_1, q_2]^T$, and $\widetilde{q}_2 = [q_3, q_4]^T$. Notice that $\mathbb{R}^4 \ni q = [\widetilde{q}_1^T, \widetilde{q}_2^T]^T = T_3^{-1}\widetilde{z}_2$, $\mathbb{R}^4 \ni w = [\widetilde{w}_1^T, \widetilde{w}_2^T]^T = T_2^{-1}\widetilde{z}_1$, and $\mathbb{R}^8 \ni z = [\widetilde{z}_1^T, \widetilde{z}_2^T]^T = T_1^{-1}x$ with $\widetilde{z}_1, \widetilde{z}_2 \in \mathbb{R}^4$.

We assume that the target (unsafe) set $\mathcal{K} \subset \mathbb{R}^8$ is chosen such that the transformations $T_1^{-1} \in \mathbb{R}^{8\times8}$, $T_2^{-1} \in \mathbb{R}^{4\times4}$, and $T_3^{-1} \in \mathbb{R}^{4\times4}$ result in $\mathcal{W}_\tau := \{w \in \mathbb{R}^4 \mid \|w\| > 20\}$ and $\mathcal{Q}_\tau := \{q \in \mathbb{R}^4 \mid \|q\| > 20\}$. The target sets for the 2D subsystems is simply the projection of $\mathcal{W}_\tau$ and $\mathcal{Q}_\tau$ onto their corresponding subspaces.

Lower dimensional reachability is performed over a grid with 101 nodes in each dimension for $\tau = 6\,\text{s}$. The overall computation time (including decomposition and projections) was 94.31 s. The complement of the shaded regions in Figure 3.6 over-approximate the reachable (unsafe) set in each of the 2D subspaces. The full 8D reachable tube is the intersection of the back-projection of the 2D reachable tubes.

The actual (minimal) reachable tube is not shown since it is prohibitively computationally expensive to compute using any Eulerian method including LS.

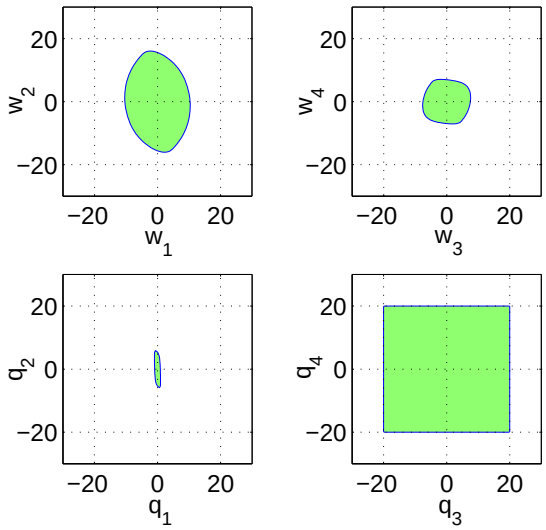


Figure 3.6: The Schur-based viability kernel (safe) of Example 3.6.3 in transformed 2D subspaces.

3.6.4 4D Unstable System (An Example for Section 3.4)

Consider an unstable system [51, Ex. 2.2.1] with

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.1023 & 0 & -0.0085 & 0 \\ 0 & 0 & 0 & 1 \\ -0.0153 & \varepsilon & 0.0993 & 0 \end{bmatrix}, \quad B = 10^{-3} \times \begin{bmatrix} 0 \\ -0.8696 \\ 0 \\ 0.1304 \end{bmatrix}.$$

Let the eigenvalues of the system be slightly perturbed as determined by parameter $\varepsilon \in \mathbb{R}$. With $\varepsilon = 0.0491$ the real anti-stable eigenvalues coincide.

We apply the method described in Section 3.4 and obtain two 2D subsystems (with separated stable and anti-stable eigenvalues) across which the

input is disjoint. The system matrices in the transformed coordinates are

$$T^{-1}AT = \left[\begin{array}{cc:cc} -0.3426 & 0.0354 & -0.6988 & 0.1399 \\ -0.0000 & -0.2912 & 0.9481 & -0.0000 \\ \hdashline 0 & 0 & 0.3150 & -0.0135 \\ 0 & 0 & 0.0003 & 0.3188 \end{array}\right], \; T^{-1}B = 10^{-3} \times \left[\begin{array}{c} 0 \\ 0 \\ \hdashline -0.7621 \\ 0.3426 \end{array}\right].$$

A target (unsafe) set $\mathcal{K}$ is chosen such that $\mathcal{Z}_\tau = \{z \in \mathbb{R}^4 \mid \sqrt{z^T z} \leq 0.2,\ z = T^{-1}x,\ x \in \mathcal{K}\}$, i.e. a small Euclidean ball of radius 0.2 around the origin. The magnitude of the input is bounded by $|u| \leq 1$. Using reachability analysis we attempt to identify the set of initial states that reach $\mathcal{Z}_\tau$ in $\tau = 3$ seconds, regardless of the input applied.

Since all conditions in Proposition 3.4 are satisfied for the lower subsystem, to obtain an over-approximation of the full-order system, we only compute the over-approximation of the reachable tube in its stable subspace. The minimal reachable tube and its over-approximation are shown in Figure 3.7. Reachability was performed over a grid with 41 nodes in each dimension. The overall computation time (including decomposition and projections) was 2.8 s. In comparison, computing the reachable tube of the full-order system would require 1741.6 s.

3.7 Summary and Further Discussions

In this chapter we presented our first decomposition technique, Schur-based decomposition, to facilitate a comparatively more scalable computation of the minimal reachable tube (and by duality the viability kernel) for LTI systems using Eulerian methods.

The decomposition was evaluated in terms of whether the resulting subsystems had disjoint or non-disjoint control inputs. In the event that a Sylvester equation can be solved, the decomposition yields two decoupled subsystems. When the Sylvester equation cannot be solved, its infinity norm minimization yields two weakly coupled subsystems. Additional constraints are considered for the case in which the control input is non-disjoint across

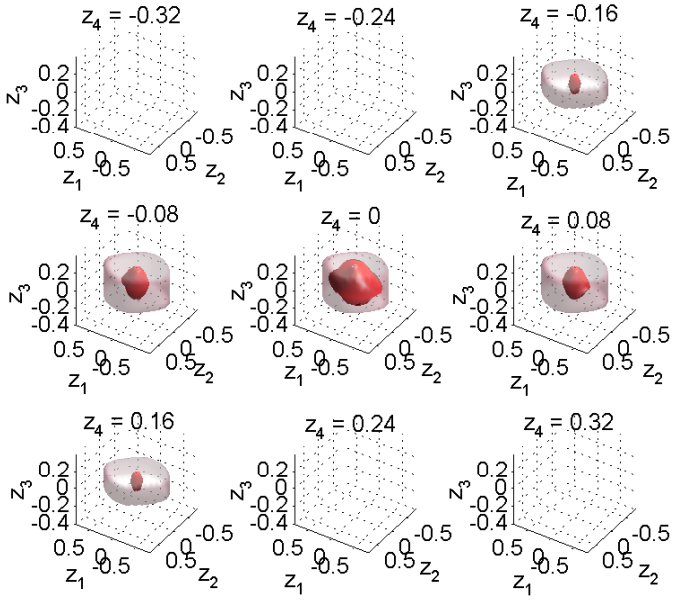


Figure 3.7: 3D snapshots of the actual (solid dark) unsafe reachable tube vs. its over-approximation (transparent light) in the transformed coordinate space for Example 3.6.4. The over-approximation was computed using Schur-based decomposition in conjunction with Proposition 3.4 for only one of the subsystems.

decomposed subsystems. Reachability analysis is then performed on the resulting subsystems independently. We applied this technique to a variety of examples computed with the Level-Set Toolbox, and found computational time significantly reduced when our method was employed. Furthermore, we presented conditions under which the minimal reachable tube and the target set coincide. We then showed that the proposed Schur-based decomposition can be used together with these conditions in order to significantly reduce the computational complexity of reachability analysis for a class of unstable systems.

The introduced conservatism in the over-approximation of the minimal reachable tube can be mitigated to some degree by considering a time-dependent formulation of the disturbance to the upper subsystem and performing reachability in sub-intervals of $[0, \tau]$. This procedure is described

towards the end of the next chapter for our second decomposition technique.

Chapter 4

Riccati-Based Structure Decomposition[1]

Our second proposed decomposition technique that aims to address Problem 2.1 draws upon the so-called Riccati transformation—a two-stage coordinate transformation based on the solutions of a nonsymmetric algebraic Riccati equation (NARE) and a Sylvester equation. This transformation, originally introduced in [19] for decoupling of singularly perturbed systems, was later generalized in [62] to larger classes of *autonomous* LTI systems. An in-depth overview of the application of this transformation in optimal control theory, singular perturbation theory, and asymptotic approximation theory can be found in [112], while more recent advances are given in [32] and [107].

Outline When the transformation results in input that is disjoint across the candidate subsystems, the standard Riccati transformation can be used to decompose the system into two decoupled subsystems. For the case in which the control input is non-disjoint across the decomposed subsystems, we propose a modified Riccati transformation (an extension to the standard Riccati transformation) which, in addition to parameterizing the unidirectional coupling between the subsystems, makes one of the subsystems ETUC. In the new coordinate space reachability analysis can then be performed in lower dimensions for each subsystem separately. The intersection of back projections of the lower dimensional reachable tubes is an overapproximation of the actual reachable tube in the transformed coordinate

[1] A version of this chapter has been published in [56].

space.

In the following analysis we assume a partitioning of (2.13) that results in exactly two subsystems. However, the proposed method can be generalized to N subsystems by applying the same decomposition algorithm iteratively (see Section 4.3).

Let (2.13) be partitioned as

$$\mathcal{S} = \left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \end{array}\right] \tag{4.1}$$

with $A_{11} \in \mathbb{R}^{k\times k}$, $A_{12} \in \mathbb{R}^{k\times(n-k)}$, $A_{21} \in \mathbb{R}^{(n-k)\times k}$, $A_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$, $B_1 \in \mathbb{R}^{k\times p}$, and $B_2 \in \mathbb{R}^{(n-k)\times p}$, for some $k < n$. Now consider the nonsingular transformation matrices

$$T_1 = \begin{bmatrix} I_k & \mathbf{0} \\ -L & I_{n-k} \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{4.2}$$

$$T_2 = \begin{bmatrix} I_k & M \\ \mathbf{0} & I_{n-k} \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{4.3}$$

With $L \in \mathbb{R}^{(n-k)\times k}$ and $M \in \mathbb{R}^{k\times(n-k)}$ that satisfy

$$\text{(NARE:)} \qquad \mathscr{R}(L) := LA_{11} - A_{22}L - LA_{12}L + A_{21} = \mathbf{0}, \tag{4.4}$$

$$\text{(Sylvester:)} \quad \mathscr{S}(M) := \big(A_{11} - A_{12}L\big)M - M\big(A_{22} + LA_{12}\big) + A_{12} = \mathbf{0}, \tag{4.5}$$

the transformed system is

$$\mathcal{S}' = T_1^{-1}(\mathcal{S}) = \left[\begin{array}{cc|c} A_{11} - A_{12}L & A_{12} & B_1 \\ \cancelto{\mathbf{0}}{\mathscr{R}(L)} & A_{22} + LA_{12} & LB_1 + B_2 \end{array}\right], \tag{4.6}$$

$$\mathcal{S}'' = T_2^{-1}(\mathcal{S}') = \left[\begin{array}{cc|c} A_{11} - A_{12}L & \cancelto{\mathbf{0}}{\mathscr{S}(M)} & (I - ML)B_1 - MB_2 \\ \mathbf{0} & A_{22} + LA_{12} & LB_1 + B_2 \end{array}\right]. \tag{4.7}$$

Solutions to (4.4) and (4.5) may not always exist.

4.1 Disjoint Control Input

Consider the resulting subsystems

$$\mathcal{S}_1'' = \left[\begin{array}{c|c} A_{11} - A_{12}L & (I - ML)B_1 - MB_2 \end{array} \right], \tag{4.8}$$

$$\mathcal{S}_2'' = \left[\begin{array}{c|c} A_{22} + LA_{12} & LB_1 + B_2 \end{array} \right]. \tag{4.9}$$

Suppose the control input is disjoint across these dynamically decoupled candidate subsystems (that is, no common input occurs in both subsystems). If the following assumption regarding the input set is satisfied, reachability analysis for each subsystem can be performed independently. Parallelization of reachability calculations in each subspace could further reduce the computational time.

Assumption 4.1. *$\mathcal{U} = \prod_{i=1}^{2} \mathcal{U}_i$ where $\mathcal{U}_i$ is the subset of $\mathbb{R}^p$ from which the portion of the input vector u acting on subsystem i draws its values.*

Assumption 4.1 enures that the inputs acting on the two subsystems are independent of one another. Note that this condition is satisfied for most physical systems where actuators are commonly uncorrelated. In the general case, however, we can under-approximate $\mathcal{U}$ by a hyper-rectangle formed by the direct product of p one-dimensional intervals. The resulting reachable tube in each subspace will be a conservative over-approximation of the actual reachable tube in that subspace since for any given system, a smaller input authority yields a larger (minimal) reachable tube.

4.2 Non-Disjoint Control Input

When the control input across the candidate subsystems is non-disjoint, even if the states of the subsystems are completely decoupled, their evolution is tightly paired through common input. Difficulty arises, for example, when in the reachability computation a control value deemed op-

timal for one subsystem is in fact non-optimal for the full-order system. Blindly performing reachability for each subsystem separately may result in an under-approximation, so appropriate measures must be taken to ensure over-approximation of the actual (unsafe) reachable tube.

One way to remedy this issue is by ensuring that at least one of the subsystems in the transformed coordinate space is ETUC. It is clear that in such a case the (otherwise non-disjoint) control action does not affect the evolution of the ETUC subsystem. Therefore, an optimal input for the subsystem with nonzero input matrix is also optimal for the full-order system. We propose a modified Riccati transformation that ensures that one of the subsystems in the transformed coordinates is ETUC, hence reachability analysis can be performed separately for each subsystem in its corresponding lower dimensional subspace.

Remark 4.1. *With an ETUC subsystem, Assumption 4.1 on the shape of the input set is lifted: The input u acts only on one of the subsystems, therefore the shape of $\mathcal{U}$ becomes irrelevant.*

4.2.1 Transformation 1 (ETUC Subsystem)

Consider a transformation through which the lower subsystem can be made ETUC. That is, in (4.6) for the transformation matrix T_1 we seek an L in $\mathscr{R}(L)$ that is also a solution of $LB_1 + B_2 = \mathbf{0}$.

Assumption 4.2. *$\mathscr{C}(B_2^T) \subseteq \mathscr{C}(B_1^T)$ with $\mathscr{C}(\cdot)$ the column-space operator.*

Lemma 4.1. *Under Assumption 4.2, the class of solutions of $LB_1 = -B_2$ w.r.t. $L \in \mathbb{R}^{(n-k)\times k}$ can be characterized by*

$$\mathcal{L} := \left\{ -B_2 B_1^\dagger + Z - Z B_1 B_1^\dagger, \;\; Z \in \mathbb{R}^{(n-k)\times k} \right\}, \tag{4.10}$$

where $\dagger$ denotes the Moore-Penrose pseudoinverse.

Proof. cf. [103] and [45]. Assumption 4.2 is the necessary and sufficient condition for solvability of $LB_1 = -B_2$. (See Appendix A.1.) □

Substituting (4.10) for L in $\mathscr{R}(L)$ we obtain

$$\begin{aligned}\widehat{\mathscr{R}}(Z) := Z\Pi + \Gamma + Z\Big(A_{12} - B_1B_1^\dagger A_{12}\Big)Z(B_1B_1^\dagger - I) \\ + \Big(A_{22} - B_2B_1^\dagger A_{12}\Big)Z(B_1B_1^\dagger - I),\end{aligned} \tag{4.11}$$

where

$$\Pi = -(B_1B_1^\dagger - I)\Big(A_{11} + A_{12}B_2B_1^\dagger\Big), \tag{4.12}$$

$$\Gamma = \Big(A_{22}B_2B_1^\dagger + A_{21}\Big) - B_2B_1^\dagger\Big(A_{12}B_2B_1^\dagger + A_{11}\Big). \tag{4.13}$$

To eliminate the non-invertible term $(B_1B_1^\dagger - I)$ from the right-hand side of (4.11) we equate $\widehat{\mathscr{R}}(Z)$ to some rank correcting term $\delta\mathscr{F}(Z)$ with

$$\mathscr{F}(Z) := Z\Big(A_{12} - B_1B_1^\dagger A_{12}\Big)Z + \Big(A_{22} - B_2B_1^\dagger A_{12}\Big)Z \tag{4.14}$$

and $\delta \in \mathbb{R}\backslash\{-1, 0\}$ a finite (but possibly large) parameter such that $\big(B_1B_1^\dagger - (\delta+1)I\big)$ is nonsingular:

$$\begin{aligned}\widehat{\mathscr{R}}(Z) &= Z\Pi + \Gamma + Z\Big(A_{12} - B_1B_1^\dagger A_{12}\Big)Z(B_1B_1^\dagger - I) \\ &\qquad + \Big(A_{22} - B_2B_1^\dagger A_{12}\Big)Z(B_1B_1^\dagger - I) \qquad (4.15)\\ &= Z\Pi + \Gamma + \mathscr{F}(Z)(B_1B_1^\dagger - I) \doteq \delta\mathscr{F}(Z). \qquad (4.16)\end{aligned}$$

Simple algebraic manipulation and post-multiplication of $\widehat{\mathscr{R}}(Z) - \delta\mathscr{F}(Z) = \mathbf{0}$ by $\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1}$ then results in a NARE in the variable Z:

$$\mathscr{R}_1(Z) := Z\tilde{A}_{11} - \tilde{A}_{22}Z - Z\tilde{A}_{12}Z + \tilde{A}_{21} = \mathbf{0} \tag{4.17}$$

with $\tilde{A}_{11} = \Pi\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1}$, $\tilde{A}_{21} = \Gamma\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1}$, $\tilde{A}_{12} = \big(B_1B_1^\dagger A_{12} - A_{12}\big)$, and $\tilde{A}_{22} = \big(B_2B_1^\dagger A_{12} - A_{22}\big)$.

Proposition 4.1. *If a root $Z \in \mathbb{R}^{(n-k)\times k}$ of the NARE (4.17) exists, it*

constitutes an $L \in \mathcal{L}$ that simultaneously satisfies

$$LB_1 + B_2 = \mathbf{0}, \tag{4.18a}$$

$$\mathscr{R}(L) = LA_{11} - A_{22}L - LA_{12}L + A_{21} = \delta\mathscr{F}(Z). \tag{4.18b}$$

Proof. By virtue of (4.16), a matrix Z that satisfies (4.17) also satisfies (4.18) via (4.10). □

Remark 4.2. *If $p \geq k$ in partitioning of (4.1), the set $\mathcal{L}$ reduces to the singleton $\{-B_2B_1^{\dagger}\}$ and the method still applies.*

Theorem 4.1. *The transformation (4.2) with $L \in \mathbb{R}^{(n-k)\times k}$ obtained through Proposition 4.1 makes the lower subsystem in (4.1) ETUC. Moreover, the coupling terms are altered such that the effect of the upper subsystem on the evolution of the lower subsystem is parameterized by δ.*

Proof.

$$\mathcal{S}' = T_1^{-1}(\mathcal{S}) = \left[\begin{array}{cc|c} A_{11} - A_{12}L & A_{12} & B_1 \\ LA_{11} - A_{22}L - LA_{12}L + A_{21} & A_{22} + LA_{12} & LB_1 + B_2 \end{array}\right] \tag{4.19}$$

$$= \left[\begin{array}{cc|c} A_{11} - A_{12}L & A_{12} & B_1 \\ \delta\mathscr{F}(Z) & A_{22} + LA_{12} & \mathbf{0} \end{array}\right]. \tag{4.20}$$

□

Remark 4.3. *Note that the imposed δ-parameterization of the off-diagonal term $\delta\mathscr{F}(Z)$ in (4.20) provides an additional degree of freedom in adjusting (minimizing) the coupling of the two subsystems in the new coordinates. This will be discussed further in Section 4.2.3.*

Nonsymmetric Riccati equations have long been an active area of research [31]. To solve (4.17) we draw on the fixed-point algorithm described in [62] and derive the necessary conditions for the existence and uniqueness of a real root Z.

Suppose $\big(B_2B_1^\dagger A_{12} - A_{22}\big)$ is invertible. Define initial values as

$$Z_0 := \big(B_2B_1^\dagger A_{12} - A_{22}\big)^{-1}\Gamma\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1}, \tag{4.21}$$

$$A_0 := \Pi\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1} - \big(B_1B_1^\dagger A_{12} - A_{12}\big)Z_0. \tag{4.22}$$

To find Z we look for

$$D := Z - Z_0 \tag{4.23}$$

by solving

$$\begin{aligned}\widetilde{\mathscr{R}}_1(D) := DA_0 - \Big(B_2B_1^\dagger A_{12} - A_{22} + Z_0\big(B_1B_1^\dagger A_{12} - A_{12}\big)\Big)D \\ - D\big(B_1B_1^\dagger A_{12} - A_{12}\big)D + Z_0A_0 = \mathbf{0}.\end{aligned} \tag{4.24}$$

Lemma 4.2 ([62, Lem. 1])**.** *Suppose* $\big(B_2B_1^\dagger A_{12} - A_{22}\big)$ *is nonsingular. If*

$$\big\|\big(B_2B_1^\dagger A_{12} - A_{22}\big)^{-1}\big\| \le \frac{1}{3\Big(\|A_0\| + \|B_1B_1^\dagger A_{12} - A_{12}\|\|Z_0\|\Big)} \tag{4.25}$$

then (4.24) *has a unique real root* D *that satisfies*

$$0 \le \|D\| \le \frac{2\|A_0\|\|Z_0\|}{\|A_0\| + \|B_1B_1^\dagger A_{12} - A_{12}\|\|Z_0\|} \tag{4.26}$$

and is the fixed-point solution of the contraction $D_{k+1} = \mathscr{P}_1(D_k)$ *given by*

$$\begin{aligned}\mathscr{P}_1(D_k) := \big(B_2B_1^\dagger A_{12} - A_{22}\big)^{-1}\Big(Z_0A_0 + D_kA_0 \\ - Z_0\big(B_1B_1^\dagger A_{12} - A_{12}\big)D_k - D_k\big(B_1B_1^\dagger A_{12} - A_{12}\big)D_k\Big).\end{aligned} \tag{4.27}$$

Remark 4.4. *Similarly to [62], it can be shown that the relative error* $e_k := \|D_k - D\|/\|D\|$ *after* k *iterations is bounded above by*

$$e_k \le \left(3\big\|\big(B_2B_1^\dagger A_{12} - A_{22}\big)^{-1}\big\|\Big(\|A_0\| + \|B_1B_1^\dagger A_{12} - A_{12}\|\|Z_0\|\Big)\right)^k \tag{4.28}$$

and decreases as $|\delta|$ increases since $\|A_0\|$ and $\|Z_0\|$ are inversely related to $|\delta|$.

For a given δ, using $D_0 = \mathbf{0}$ as initial condition we compute D iteratively. The fixed-point solution $D^* = \mathscr{P}_1(D^*)$ is then used to obtain $Z = D^* + Z_0$ which in turn solves $\mathscr{R}_1(Z) = \mathbf{0}$ in (4.17) and results in a matrix L, through (4.10), that satisfies both equations in (4.18).

4.2.2 Transformation 2 (Unidirectionally Coupled Subsystems)

Consider the NARE

$$\begin{aligned}\mathscr{R}_2(M) = \big(A_{11} - A_{12}L\big)M - M\big(A_{22} + LA_{12}\big) \\ - M\big(\delta\mathscr{F}(Z)\big)M + A_{12} = \mathbf{0}.\end{aligned} \tag{4.29}$$

For a given L, δ, and Z, if there exists a solution M that satisfies (4.29), we obtain the following:

Theorem 4.2. *The transformation* (4.3) *with $M \in \mathbb{R}^{k\times(n-k)}$ satisfying NARE* (4.29) *makes the subsystems in* (4.20) *unidirectionally coupled.*

Proof.

$$\mathcal{S}'' = T_2^{-1}(\mathcal{S}') = \left[\begin{array}{cc|c} A_{11} - A_{12}L - M\delta\mathscr{F}(Z) & \cancel{\mathscr{R}_2(M)}\ \mathbf{0} & B_1 \\ \delta\mathscr{F}(Z) & A_{22} + LA_{12} + \delta\mathscr{F}(Z)M & \mathbf{0} \end{array}\right]. \tag{4.30}$$

□

Remark 4.5. *In the transformed coordinates the lower subsystem remains ETUC. Furthermore, the δ-parameterization of the unidirectional coupling between subsystems is also preserved.*

Before further analyzing the unidirectional coupling term $\delta\mathscr{F}(Z)$, let us derive the necessary conditions for the existence and uniqueness of a

solution M to (4.29) to be used with the same convergent iterative procedure described previously.

For a given δ, Z, and L, let $\big(A_{11} - A_{12}L\big)$ be invertible and the initial values be defined as

$$M_0 := -\big(A_{11} - A_{12}L\big)^{-1}A_{12}, \tag{4.31}$$

$$N_0 := A_{22} + LA_{12} + \delta\mathscr{F}(Z)M_0. \tag{4.32}$$

We seek M by forming

$$J := M - M_0 \tag{4.33}$$

and solving

$$\widetilde{\mathscr{R}}_2(J) := JN_0 - \big(A_{11} - A_{12}L - \delta M_0\mathscr{F}(Z)\big)J + \delta J\mathscr{F}(Z)J + M_0N_0 = \mathbf{0}. \tag{4.34}$$

Lemma 4.3 ([62, Lem. 1])**.** *Suppose* $\big(A_{11} - A_{12}L\big)$ *is nonsingular. If*

$$\big\|\big(A_{11} - A_{12}L\big)^{-1}\big\| \leq \frac{1}{3\Big(\|N_0\| + \|\delta\mathscr{F}(Z)\|\|M_0\|\Big)} \tag{4.35}$$

then (4.34) *has a unique real root* J *that satisfies*

$$0 \leq \|J\| \leq \frac{2\|N_0\|\|M_0\|}{\|N_0\| + \|\delta\mathscr{F}(Z)\|\|M_0\|} \tag{4.36}$$

and is the fixed-point solution of the contraction $J_{k+1} = \mathscr{P}_2(J_k)$ *given by*

$$\mathscr{P}_2(J_k) := \big(A_{11} - A_{12}L\big)^{-1}\Big(M_0N_0 + J_kN_0 + \delta M_0\mathscr{F}(Z)J_k + \delta J_k\mathscr{F}(Z)J_k\Big). \tag{4.37}$$

Remark 4.6. *As in [62], we can show that the relative error* $e_k := \|J_k - J\|/\|J\|$ *after* k *iterations is bounded above by*

$$e_k \leq \left(3\big\|\big(A_{11} - A_{12}L\big)^{-1}\big\|\Big(\|N_0\| + \|\delta\mathscr{F}(Z)\|\|M_0\|\Big)\right)^k \tag{4.38}$$

and decreases as $\|\delta\mathscr{F}(Z)\|$, $\|A_{22}\|$, *and* $\big\|\big(A_{11} - A_{12}L\big)^{-1}\big\|$ *decrease. This*

occurs when the ill-conditioning of the A-matrix increases (e.g. in the case of two-time-scale systems; see [63] and the references therein) and δ is chosen such that $\|\delta\mathscr{F}(Z)\|$ is minimized.

Using $J_0 = \mathbf{0}$ as initial condition we compute J iteratively. The fixed-point solution $J^* = \mathscr{P}_2(J^*)$ is then used to obtain $M = J^* + M_0$ which in turn solves $\mathscr{R}_2(M) = \mathbf{0}$ in (4.29).

Note that both conditions (4.25) and (4.35) are conservative and their satisfaction ensures rapid convergence (usually within 2 or 3 iterations). In practice, the right-hand-side of these inequalities can be relaxed up to 10 times in most cases without causing divergence.

4.2.3 The Unidirectional Coupling Term (Choosing δ)

Finally, we analyze the unidirectional coupling term $\delta\mathscr{F}(Z)$ and its behavior with respect to the free parameter δ. Since Z is an implicit function of δ, we adopt the extended notation $\delta\mathscr{F}(Z(\delta))$ to reflect this dependency.

First, we formalize a conservative upper-bound on $\|\delta\mathscr{F}(Z(\delta))\|$ as an explicit function of δ. This assures that the unidirectional coupling remains bounded for almost all admissible values of the free parameter δ.

Proposition 4.2. *The worst-case unidirectional coupling between the two subsystems in the transformed coordinates, i.e. $\|\delta\mathscr{F}(Z(\delta))\|$ in (4.30), is (conservatively) bounded above such that*

$$\|\delta\mathscr{F}(Z(\delta))\| \leq \frac{1}{|\delta|}\left(\frac{|\delta|+1}{|\delta+1|}\right)^2 a + \left(\frac{|\delta|+1}{|\delta+1|}\right) b \quad \forall \delta \in \mathbb{R}\backslash\{-1, 0\}, \tag{4.39}$$

where the constants a and b are independent of δ and are determined by $a := \alpha(b/\beta)^2$, $b := 3\|B_1 B_1^\dagger\|\gamma\beta$, $\gamma := \|\Gamma\|\left\|\left(A_{22} - B_2 B_1^\dagger A_{12}\right)^{-1}\right\|$, $\alpha := \|A_{12} - B_1 B_1^\dagger A_{12}\|$, and $\beta := \|A_{22} - B_2 B_1^\dagger A_{12}\|$.

Proof. The proof is provided in Appendix B.1. □

Now consider inequalities (4.25) and (4.35), which are dependent on δ. Adequately chosen and sufficiently large values of δ help ensure that these

conditions are met. On the other hand, choosing δ exceedingly large defeats the purpose of δ-parameterization of the unidirectional coupling term, since it can be shown that as δ grows, $\|\delta\mathscr{F}(Z(\delta))\|$ approaches a problem-dependent constant that may not necessarily be an extremum point.

Proposition 4.3. $\lim_{\delta\to\pm\infty}\|\delta\mathscr{F}(Z(\delta))\| = \|\Gamma\|$ *with* Γ *given by* (4.13).

Proof. This proof is also provided in Appendix B.1. □

It follows from Proposition 4.3 that $0 \leq \inf_\delta\|\delta\mathscr{F}(Z(\delta))\| \leq \|\Gamma\|$. Therefore naively letting $|\delta| \to \infty$ essentially removes the added flexibility associated with the δ-parameterization in the modified Riccati approach and instead enforces a trivial solution $L = -B_2B_1^\dagger$. While for some systems this solution may yield the smallest possible unidirectional coupling between the resulting subsystems (i.e. a unidirectional coupling with the lowest infinity norm), in most cases a carefully chosen δ not only facilitates the satisfaction of the convergence conditions (4.25) and (4.35), but also further minimizes the worst-case unidirectional coupling. Thus, formulated as an optimization problem, we seek a δ that solves the following:

$$\begin{aligned}&\underset{\delta\in\mathbb{R}\setminus\{-1,0\}}{\text{minimize}} \quad f(\delta) := \|\delta\mathscr{F}(Z(\delta))\| \\ &\text{subject to} \quad (4.25) \text{ and } (4.35).\end{aligned}$$

Note that this is a non-convex problem, and in general, $f(\cdot)$ may be a non-smooth function of δ. However, a global optimum need not be computed. Any suboptimal solution can be used as long as that solution yields a satisfactory degree of unidirectional coupling between the subsystems in the transformed coordinates. In addition, an approximation to the optimum point can be obtained numerically, for example by fine-griding the real line or using the bisection algorithm.

In practice, while the exact shape of the function $f(\cdot)$ is problem-dependent, we have found (but have not proven) that in most cases it exhibits a behavior similar to that of an absolute value proper rational function (over a

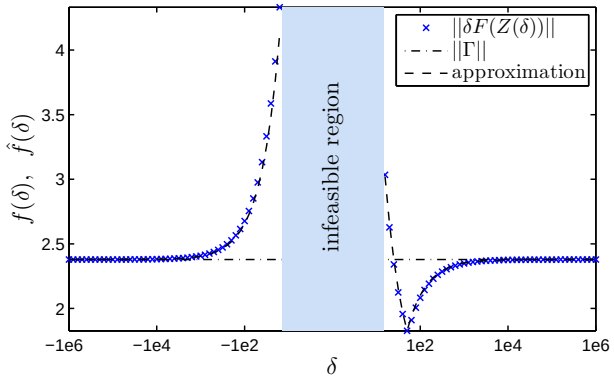


Figure 4.1: The worst-case unidirectional coupling $f(\delta) = \|\delta\mathscr{F}(Z(\delta))\|$ ($\times$'s) and its approximation $\hat{f}(\delta) = |\frac{-27.65}{\delta} + 0.55| + 1.82$ (dashed) computed for Example 4.5.1. The interval $(-15, +15)$ over which (4.25) and (4.35) are violated is labeled as "infeasible region". The asymptote $\lim_{\delta\to\pm\infty} f(\delta) = \|\Gamma\|$ (dash-dotted) is also shown. The minimum of $f(\delta)$ occurs when $\delta \approx +50$.

discontinuous domain) of the form

$$\hat{f}(\delta) = \left|\frac{c_0}{\delta^k} + c_1\right| + c_2 \quad \forall\delta \in \mathcal{D}, \tag{4.40}$$

where $\mathcal{D} \subset \mathbb{R}\backslash\{-1, 0\}$ is the union of the two segments of the real line for which the magnitude of δ is large enough such that (4.25) and (4.35) are both satisfied, $k \in \mathbb{N}$, k : odd, $c_0 = -c_1(\delta^*)^k$, $\delta^* = \arg\min_{\delta\in\mathcal{Y}} f(\delta)$, $c_2 = \min_{\delta\in\mathcal{Y}} f(\delta)$, and $c_1 = \big(\lim_{\delta\to\pm\infty} f(\delta)\big) - c_2 = \|\Gamma\| - c_2$. For example, consider the transformed system in Section 4.5.1. Figure 4.1 shows $f(\delta)$ and its approximation $\hat{f}(\delta) = |-\frac{27.65}{\delta} + 0.55| + 1.82$ evaluated where (4.25) and (4.35) hold.

4.3 Recursive Decomposition

To apply the described decomposition technique in a recursive fashion, consider the resulting subsystems in (4.8) and (4.30). A recursive decomposition when the standard Riccati transformation can be used is straightforward.

Suppose that the modified Riccati transformation is used throughout the process. In deeper level recursions, the decomposition can be applied to the uppermost subsystem since that subsystem is controlled whereas every other subsystem is ETUC. For example, to decompose a 6D system into three 2D subsystems, in the first recursion level, the partitioning can be chosen such that the resulting upper (controlled) subsystem is 4D and the lower (ETUC) subsystem is 2D. In the second recursion level, if the solutions exist, the 4D subsystem is then decomposed into two 2D subsystems.

Note that in the recursive application of the decomposition, when the modified Riccati transformation is employed, all subsystems but one are ETUC. Therefore, this iterated decomposition may result in a more conservative over-approximation of the actual reachable tube.

4.4 Reachability in Lower Dimensions

We will focus mainly on over-approximating the minimal reachable tube that can only be computed using the computationally intensive Eulerian methods as it is these methods that stand to benefit the most from our decomposition approach. Nevertheless, complementary discussions surrounding the computation of the maximal reachable tube are provided in Section 4.5.5.

In the new coordinates $z = T^{-1}x$, $T = T_1T_2$, the subsystem dynamics are governed by

$$\dot{z}_1 = \big(A_{11} - A_{12}L - \delta M\mathscr{F}(Z)\big)z_1 + B_1u, \tag{4.41}$$

$$\dot{z}_2 = \big(A_{22} + LA_{12} + \delta\mathscr{F}(Z)M\big)z_2 + B_2u + \delta\mathscr{F}(Z)z_1 \tag{4.42}$$

with $\delta\mathscr{F}(Z) = \mathbf{0}$ when the standard Riccati transformation yields disjoint input, and $B_2 = \mathbf{0}$ when the modified Riccati transformation is employed. Denote the two subspaces of $\mathbb{R}^n$ in which the subsystems evolve as

$$\mathbb{S}_1 := \mathbb{R}^k \quad \text{and} \quad \mathbb{S}_2 := \mathbb{R}^{n-k}. \tag{4.43}$$

Algorithm 4.1 computes the (minimal) reachable tube in lower dimensions:

Algorithm 4.1 Reachability in lower dimensions (Riccati-Based)

1: $\mathcal{Z}_\tau \leftarrow T^{-1}\mathcal{K}$
2: $\mathcal{Z}_\tau^i \leftarrow \text{Proj}_{\mathbb{S}_i}(\mathcal{Z}_\tau), \quad \forall i \in \{1,2\}$ ▷ project onto ith subspace
For upper subsystem:
3: $\mathcal{Z}^1_{[0,\tau]} \leftarrow Reach^{\flat}_{[0,\tau]}(\mathcal{Z}^1_\tau, \mathcal{U})$
For lower subsystem:
4: Treat $\delta\mathscr{F}(Z)z_1$ as disturbance (existentially quantified)
5: $\zeta \leftarrow \|z_1\| \equiv \sup_{v\in\mathcal{Z}^1_{[0,\tau]}}\|v\|$
6: Compute upper-bound $\|\delta\mathscr{F}(Z)z_1\| \leq \|\delta\mathscr{F}(Z)\|\zeta$
7: $\mathcal{Z}^2_{[0,\tau]} \overset{\text{consrv.}}{\longleftarrow} Reach^{\sharp}_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{B}(\|\delta\mathscr{F}(Z)\|\zeta))$
8: **return**$(\mathcal{Z}^1_{[0,\tau]}, \mathcal{Z}^2_{[0,\tau]})$

When the standard Riccati transformation is used to obtain the subsystems, steps 4–7 of Algorithm 4.1 are simply replaced with $\mathcal{Z}^2_{[0,\tau]} \leftarrow Reach^{\flat}_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{U})$.

Notice that if the error due to projections can be ignored, reachability in the upper subspace is "exact" in the sense that $\mathcal{Z}^1_{[0,\tau]} \equiv \text{Proj}_{\mathbb{S}_1}(T^{-1}Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}))$. This is generally not true in the lower subspace. The unidirectional coupling between the two subsystems is treated as a disturbance to the lower subsystem, resulting in a conservative reachability computation in that subspace. The computed reachable tube in the lower subspace is a guaranteed over-approximation of the projection of the actual reachable tube in that subspace; $Reach^{\sharp}_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{B}(\|\delta\mathscr{F}(Z)\|\zeta)) \supseteq \text{Proj}_{\mathbb{S}_2}(T^{-1}Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}))$.

The parameter $\delta = \delta^*$ is precomputed so as to minimize $\|\delta\mathscr{F}(Z)\|$. However, when the reachability horizon in Step 3 of Algorithm 4.1 is large, the magnitude of the input to the lower subsystem (whose upper-bound is directly proportional to $\sup_{v\in\mathcal{Z}^1_{[0,\tau]}}\|v\|$) may become so large as to warrant reachability in that subspace over sub-intervals of $[0,\tau]$ in a similar fashion to [36]. For $N := \tau/q$, $N \in \mathbb{N}$ time steps each of length $q \in \mathbb{R}^+$, we have

$$\mathcal{Z}^2_{[0,\tau]} = \bigcup_{i=0}^{N-1} \mathcal{Z}^2_{[iq,\,(i+1)q]} \tag{4.44}$$

and, by the semi-group property,

$$\mathcal{Z}^2_{[iq,\,(i+1)q]} = Reach^{\sharp}_{[0,q]}\big(\mathcal{Z}^2_{[(i+1)q,\,(i+2)q]}, \mathcal{B}(\|\delta\mathscr{F}(Z)\|\zeta_{N-1-i})\big), \tag{4.45}$$

where $\mathcal{Z}^2_{[\tau,+\infty)} \doteq \mathcal{Z}^2_\tau$ and ζ_i is the supremum of the reachable tube in the upper subspace at intermediate time steps. This holds since the input to the lower subsystem is a disturbance and therefore all quantifiers in reachability analysis of this subsystem are existential. Recording ζ_i at each time step (rather than at the end of the reachability horizon) allows for a gradual incrementing of the disturbance to the lower subsystem. Thus, using the sequence $\{\zeta_i\}_{i=0}^{N-1}$ and executing Algorithm 4.1 with sub-intervals according to (4.44)–(4.45), we can compute a less conservative reachable tube in the lower subspace.

A guaranteed over-approximation of the actual reachable tube of the full-order system in $\mathcal{X}$ can be obtained using Lemma 3.1 as in the Schur-based case, i.e. via

$$T\left((\mathcal{Z}^1_{[0,\tau]} \times \mathbb{S}_2) \cap (\mathbb{S}_1 \times \mathcal{Z}^2_{[0,\tau]})\right) \supseteq Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}). \tag{4.46}$$

4.4.1 Formulating an Upper-Bound on Conservatism of $\mathcal{Z}^2_{[0,\tau]}$

Consider the computed reachable tube $\mathcal{Z}^2_{[0,\tau]} := Reach^{\sharp}_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{B}(\|\delta\mathscr{F}(Z)\|\zeta))$ in the lower subspace. Take $\bar{z}_2 \in \mathcal{Z}^2_\tau$ and suppose $\mathscr{D}_{[0,\tau]}$ is the set of measurable functions from $[0,\tau]$ to $\mathcal{B}(\|\delta\mathscr{F}(Z)\|\zeta)$. There exists an admissible input $\vartheta(\cdot) \in \mathscr{D}_{[0,\tau]}$ such that in positive time using time-reversed dynamics we have

$$z_2 := e^{-\tau\Omega}\bar{z}_2 - \int_0^\tau e^{-(\tau-r)\Omega}\vartheta(r)dr, \tag{4.47}$$

$$z_2 \in \mathcal{Z}^2_{[0,\tau]} \tag{4.48}$$

with $\Omega := A_{22} + LA_{12} + \delta\mathscr{F}(Z)M$. Therefore, as in [36],

$$\|z_2 - e^{-\tau\Omega}\bar{z}_2\| \leq \int_0^\tau e^{(\tau-r)\|\Omega\|}\|\delta\mathscr{F}(Z)\|\zeta dr \tag{4.49}$$

$$= \frac{e^{\tau\|\Omega\|}-1}{\|\Omega\|}\|\delta\mathscr{F}(Z)\|\zeta. \tag{4.50}$$

Notice that,

$$\frac{e^{\tau\|\Omega\|}-1}{\|\Omega\|}\|\delta\mathscr{F}(Z)\|\zeta = \left(\lim_{M\to\infty}\sum_{i=1}^{M}\frac{\tau^i\|\Omega\|^{i-1}}{i!}\right)\|\delta\mathscr{F}(Z)\|\zeta \tag{4.51}$$

$$\leq \left(\lim_{M\to\infty}\sum_{i=1}^{M}\frac{\tau^i\left(\overline{\sigma}(\Omega)\sqrt{n_k}\right)^{i-1}}{i!}\right)\|\delta\mathscr{F}(Z)\|\zeta \tag{4.52}$$

$$=: \eta_{[0,\tau]} \tag{4.53}$$

with $\overline{\sigma}(\cdot)$ the largest singular value operator, and $n_k := n-k$ the dimension of the ETUC subsystem. Due to linearity and time-invariance of the dynamics, the (possibly non-convex) backward reachable tube can be expressed as

$$\mathcal{Z}^2_{[0,\tau]} \subseteq \left(\bigcup_{t\in[0,\tau]} e^{-\Omega t}\mathcal{Z}^2_\tau\right) \oplus \mathcal{B}(\eta_{[0,\tau]}). \tag{4.54}$$

The right-hand side of (4.54) provides an upper-bound on how much $\mathcal{Z}^2_{[0,\tau]}$ can grow in backward time in terms of the reachability horizon τ, the choice of n_k, the magnitude of the unidirectional coupling $\|\delta\mathscr{F}(Z)\|$, the supremum of the reachable tube in the upper subspace $\zeta = \sup_{v\in\mathcal{Z}^1_{[0,\tau]}}\|v\|$, and the largest singular value $\overline{\sigma}(\Omega)$ of the lower subsystem.

When reachability is performed over sub-intervals, a tighter upper-bound can be formulated by replacing the right-hand side of (4.54) with

$$\overline{\mathcal{Z}}^2_{[0,\tau]} := \bigcup_{i=0}^{N-1}\left(\left(\bigcup_{t\in[iq,(i+1)q]} e^{-\Omega t}\mathcal{Z}^2_{[(i+1)q,(i+2)q]}\right) \oplus \mathcal{B}(\eta_{[iq,(i+1)q]})\right). \tag{4.55}$$

As in [36], it can be shown that the quality of this upper-bound is good in

the Hausdorff distance, in that $\lim_{q\to 0} \text{dist}_H(\mathcal{Z}^2_{[0,\tau]}, \overline{\mathcal{Z}}^2_{[0,\tau]}) \to 0$. Determining whether $\mathcal{Z}^2_{[0,\tau]}$ itself is a close over-approximation to $\text{Proj}_{\mathbb{S}_2}(T^{-1}Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}))$ is much more involved and remains an open problem. We expect that performing reachability over sub-intervals, choosing n_k appropriately, and minimizing the disturbance (uncertainty) magnitude as much as possible could reduce the conservatism considerably.

4.4.2 The Effect of Dimension on Magnitude of Uncertainty

Since the worst-case unidirectional coupling $\|\delta\mathscr{F}(Z)\|$ contributes to the uncertainty (i.e. disturbance input) in the reachability computation for the lower subsystem, we examine the potential affect of the system dimension on a) the magnitude of the unidirectional coupling and b) the amount of time consumed by the decomposition process.

Although entirely dependent on the particular system to which the modified Riccati transformation is applied, the following empirical test can serve as a rough measure: For a given dimension n, we generated 10 randomized systems, applied the decomposition method to each n-dimensional system with $n_k = \frac{n}{2}$, and recorded $\|\delta\mathscr{F}(Z)\|$ and the computational time. We repeated this for $n = 4, 6, 8, 10, 16$, as shown in Figure 4.2. While for higher dimensions, the computational time and the magnitude of the unidirectional coupling show an increasing trend in their average values, there is significant variance. In addition, the time required for the decomposition process, even for the highest dimension, is still negligible compared to the time required for actual reachability computations.

4.4.3 Conservatism Due to Projection

We have assumed that the target set in the transformed coordinates is (or is close to) a direct product of the sets in the subsystems' subspaces, meaning that the error due to projections of the target set onto the lower dimensional subspaces can be ignored. This assumption does not generally hold. If the target set in the new coordinate space is far from being axis-aligned, the projections contribute to the conservatism of the reachable tube over-

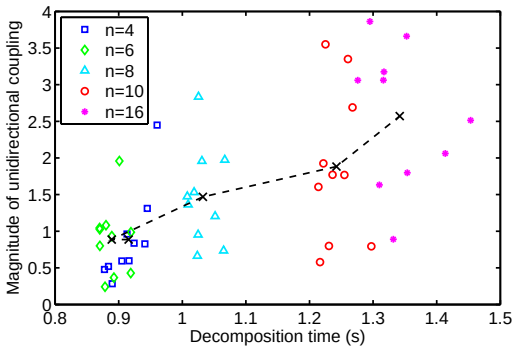


Figure 4.2: The worst-case unidirectional coupling and the time required to compute the transformed matrices for randomly generated systems of dimension $n = 4, 6, 8, 10, 16$ show significant variance. Average values are marked by 'x'.

approximation. A similar argument holds for the back-projection of the subsystem reachable tubes, since we attempt to reconstruct a higher dimensional object from lower dimensional entities. Unfortunately, loss of information is inherent in any projection operation and, to a great extent, cannot be avoided.

4.4.4 Implications of Computing the Riccati-Based Reachable Tube

Safety Verification

Let $\mathcal{I} \subset \mathcal{X}$ be the set of all possible initial states of (2.12). It follows from Definition 2.3 that, for a given unsafe target set $\mathcal{K}$, the system is safe if and only if the backward minimal reachable tube does not intersect $\mathcal{I}$:

$$\mathcal{R}_{[0,\tau]} := Reach^{\flat}_{[0,\tau]}(\mathcal{K}, \mathcal{U}) \cap \mathcal{I} = \emptyset. \tag{4.56}$$

Let $\overline{\mathcal{R}}_{[0,\tau]}$ and $\mathcal{Z}_{[0,\tau]}$ denote the computed Riccati-based reachable tube in the original and transformed coordinates, respectively.

Proposition 4.4. $\mathcal{Z}_{[0,\tau]} \cap T^{-1}\mathcal{I} = \emptyset \iff \overline{\mathcal{R}}_{[0,\tau]} \cap \mathcal{I} = \emptyset \implies \mathcal{R}_{[0,\tau]} \cap \mathcal{I} = \emptyset.$

Proof. $\forall x \in \overline{\mathcal{R}}_{[0,\tau]} \ \forall y \in \mathcal{I} \ (\overline{\mathcal{R}}_{[0,\tau]} \cap \mathcal{I}) = \emptyset \iff x \neq y \iff T^{-1}x \neq T^{-1}y \iff T^{-1}\overline{\mathcal{R}}_{[0,\tau]} \cap T^{-1}\mathcal{I} = \emptyset$. By Lemma 3.1 we have that $\mathcal{R}_{[0,\tau]} \subseteq \overline{\mathcal{R}}_{[0,\tau]}$. Thus, $\overline{\mathcal{R}}_{[0,\tau]} \cap \mathcal{I} = \emptyset \implies \mathcal{R}_{[0,\tau]} \cap \mathcal{I} = \emptyset$. □

Therefore, if the Riccati-based reachable tube does not intersect $\mathcal{I}$ in either coordinates, the system is safe. A simpler (but more conservative) sufficient condition for safety can be given as:

Proposition 4.5. $\bigwedge_{i=1}^{2} \left(\mathcal{Z}^i_{[0,\tau]} \cap \text{Proj}_{\mathbb{S}_i}(T^{-1}\mathcal{I}) = \emptyset \right) \implies \mathcal{R}_{[0,\tau]} \cap \mathcal{I} = \emptyset.$

Proof. $(\mathcal{Z}^1_{[0,\tau]} \cap \text{Proj}_{\mathbb{S}_1}(T^{-1}\mathcal{I}) = \emptyset) \wedge (\mathcal{Z}^2_{[0,\tau]} \cap \text{Proj}_{\mathbb{S}_2}(T^{-1}\mathcal{I}) = \emptyset) \implies (\mathcal{Z}^1_{[0,\tau]} \times \mathbb{S}_2) \cap (\mathbb{S}_1 \times \mathcal{Z}^2_{[0,\tau]}) \cap T^{-1}\mathcal{I} = \emptyset \iff \mathcal{Z}_{[0,\tau]} \cap T^{-1}\mathcal{I} = \emptyset \implies \mathcal{R}_{[0,\tau]} \cap \mathcal{I} = \emptyset.$ □

That is, if the computed Riccati-based reachable tubes for each subsystem in the new coordinates do not intersect the projections of the transformed $\mathcal{I}$, then the system is safe in the original coordinates. This can be used to prove safety using the lower dimensional reachable tubes when reconstructing and storing the full-order state space (or reachable tube) is costly.

Safety-Preserving Control Synthesis

Given the unsafe target set $\mathcal{K}$ and the computed Riccati-based reachable tube $\overline{\mathcal{R}}_{[0,\tau]}$, let $\overline{\mathcal{R}}^c_{[0,\tau]}$ denote an under-approximation of the finite-time viability kernel $Viab_{[0,\tau]}(\mathcal{K}^c, \mathcal{U})$ under $\mathcal{S}$. Then using the optimal control laws precomputed during the reachability analysis (e.g. via [90]), one can construct a feedback controller as in [81] on the boundaries of $\overline{\mathcal{R}}^c_{[0,t]}$, $t \in [0, \tau]$, that keeps the trajectories of $\mathcal{S}$ within $\mathcal{K}^c$ (i.e. within safety) over the horizon $[0, \tau]$. Hence, although possibly conservative, computing the reachable tube through the Riccati-based approach can be a powerful tool to guarantee safety in safety-critical systems that have moderately high dimensions.

4.5 Numerical Examples

We employ Level Set Toolbox (LS) [90] for the computation of the minimal reachable tubes (and the viability kernels). The use of our Riccati-based approach for maximal reachability analysis is discussed and an example is provided that employs Ellipsoidal Toolbox (ET) [73]. All computations are performed on a dual core Intel-based computer with 2.8 GHz CPU, 6 MB of L2 cache and 3 GB of RAM running single-threaded 32-bit MATLAB 7.5.

4.5.1 Arbitrary 4D System

Consider an LTI system $\dot{x} = Ax + Bu$ with

$$A = \begin{bmatrix} 1.5072 & 3.3984 & 0.1300 & -0.0884 \\ 5.0644 & -2.6683 & 0.0227 & 0.1689 \\ 0.1156 & -0.1863 & 0.5686 & 0.2648 \\ -0.0808 & 0.0229 & 0.4915 & 0.5949 \end{bmatrix}, \quad B = \begin{bmatrix} -0.7433 \\ -2.2528 \\ -0.9075 \\ 0.6036 \end{bmatrix}$$

and $u \in \mathbb{R}$, $|u| \leq 0.1$. We define a target (unsafe) set $\mathcal{K}$ such that in the transformed coordinate space $\mathcal{Z}_\tau = \{z \in \mathbb{R}^4 \mid \|z\|_2 \leq 0.2,\, z = T^{-1}x,\, x \in \mathcal{K}\}$ where $\|\cdot\|_2$ is the Euclidean norm and T is the transformation matrix obtained through the presented modified Riccati method.

We decompose this system into two 2D subsystems. Sweeping through the real line, a nearly optimal $\delta^* \approx +50$ that minimizes the worst-case unidirectional coupling between the subsystems is found in less than a second. Equations (4.27) and (4.37) converge to their fixed-point solutions in less than a dozen iterations. The system matrices in the new coordinate space are

$$A'' = \left[\begin{array}{cc:cc} 1.5362 & 3.4622 & 0 & 0 \\ 4.9377 & -2.7030 & 0 & 0 \\ \hdashline -1.8075 & -0.0193 & 0.5727 & 0.2612 \\ 1.2341 & 0.2918 & 0.4858 & 0.5964 \end{array}\right], \quad B'' = \left[\begin{array}{c} -0.7433 \\ -2.2528 \\ \hdashline 0 \\ 0 \end{array}\right].$$

Reachability calculations are performed over a grid with 41 nodes in

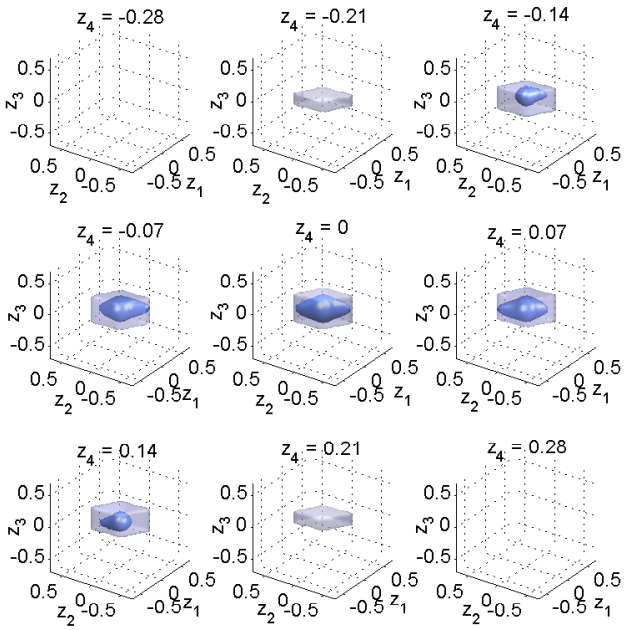


Figure 4.3: Series of 3D snapshots of the Riccati-based over-approximation (transparent light) vs. actual (solid dark) minimal reachable (unsafe) tube in the transformed coordinate space for Example 4.5.1.

each dimension for $\tau = 3\,\mathrm{s}$. The full-order reachable tube (10144.80 s computational time) is over-approximated by the Riccati-based reachable tube (3.98 s computational time, including calculation of δ^*, transformation matrices, the decomposition, and projections) as shown in Figure 4.3.

The loss of accuracy (due to treating the unidirectional coupling as disturbance to the lower subsystems and assembling the full-order reachable tube from projections) can be quantified in terms of the Hausdorff distance between the full-order and the Riccati-based reachable tubes: Since a set in LS is described by a signed distance function whose magnitude at any point in the state space is the minimum distance to the boundary of that set, it is straightforward to compute the Hausdorff distance between any given two sets. For this example the Hausdorff distance is 0.18. In addition, a volumetric measure of the inaccuracy can also be quantified in terms of the

ratio between the number of grid points contained in the difference of the two sets and the number of grid points in the Riccati-based reachable tube. This ratio is found to be 38.4%.

4.5.2 4D Cart with Two Inverted Pendulums

Consider the linearized model of a cart with two separately mounted inverted pendulums from [51, Ex. 2.2.1] with $l_1 = 30$, $l_2 = 35$. The state vector $x \in \mathbb{R}^4$ consists of angular displacement of each inverted pendulum from vertical and the corresponding angular velocities; the control input $u \in \mathbb{R}$ arises from a force applied to the cart such that $|u| \leq 10$. The system matrices are

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.3920 & 0 & -0.0327 & 0 \\ 0 & 0 & 0 & 1 \\ 0.0560 & 0 & 0.2753 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ -0.0033 \\ 0 \\ -0.0005 \end{bmatrix}.$$

We choose a non-convex target (unsafe) set $\mathcal{K}$ such that in the transformed coordinates we have $\mathcal{Z}_\tau = \{z \in \mathbb{R}^4 \mid \|z\| \geq 0.5,\, z = T^{-1}x,\, x \in \mathcal{K}\}$. We seek to identify the set of states for which there exists a bounded control law that keeps the system trajectories contained in $\mathcal{Z}_\tau^c$. The safety-preserving control synthesized through LS provides a guarantee that the pendulums' angular displacement will not exceed an infinity norm ball around their upright positions, despite control saturation. Similar "envelope protection" problems arise in other domains, including aircraft flight management systems and anesthesia automation, among others.

We decompose this system into two 2D subsystems, with the unidirectional coupling determined by the solution $L = -B_2 B_1^\dagger$ regardless of the

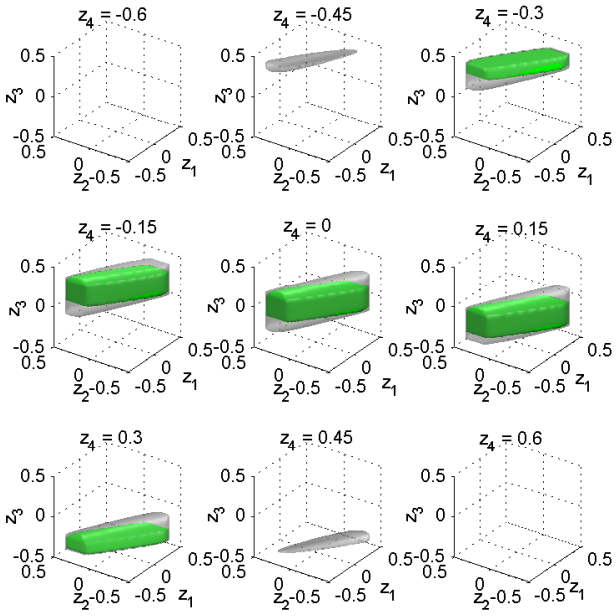


Figure 4.4: Riccati-based (solid dark) vs. actual (transparent light) viability kernels in the transformed coordinate space for Example 4.5.2. The minimal reachable tube and its over-approximation are the non-convex complements of these objects.

value of δ. The system matrices in the new coordinate space become

$$A'' = \left[\begin{array}{cc:cc} 0 & 0.9524 & 0 & 0 \\ 0.3920 & 0 & 0 & 0 \\ \hdashline 0 & 0.1429 & 0 & 1.0500 \\ 0 & 0 & 0.2800 & 0 \end{array}\right], \quad B'' = \left[\begin{array}{c} 0 \\ -0.0033 \\ \hdashline 0 \\ 0 \end{array}\right].$$

Reachability calculations are performed over a grid with 41 nodes in each dimension for $\tau = 3\,\mathrm{s}$. Figure 4.4 shows the computed viability kernels (safe) as the area inside the shaded regions. The computation time for the actual and the transformation-based reachable tubes were 1098.48 s and 4.27 s, respectively. The Hausdorff distance between these two sets is 0.21 and the Riccati-based viability kernel covers 74% of the volume of the full-order set.

4.5.3 Arbitrary 6D System

Consider the system $\dot{x} = Ax + Bu$ with

$$A = \begin{bmatrix} 3.3155 & 0.7768 & 2.4455 & 0.0028 & -0.0094 & -0.0097 \\ 0.6320 & -1.4796 & -2.3001 & -0.0370 & 0.0322 & -0.0112 \\ -0.1047 & -0.3522 & 0.6578 & 0.0282 & -0.0621 & -0.0214 \\ 0.0344 & 0.0360 & 0.0091 & 0.1885 & 0.0518 & 0.3567 \\ -0.0140 & -0.0225 & -0.0012 & 0.2746 & 0.0866 & 0.1567 \\ -0.0106 & -0.0215 & -0.0082 & 0.0814 & 0.0887 & 0.1182 \end{bmatrix},$$

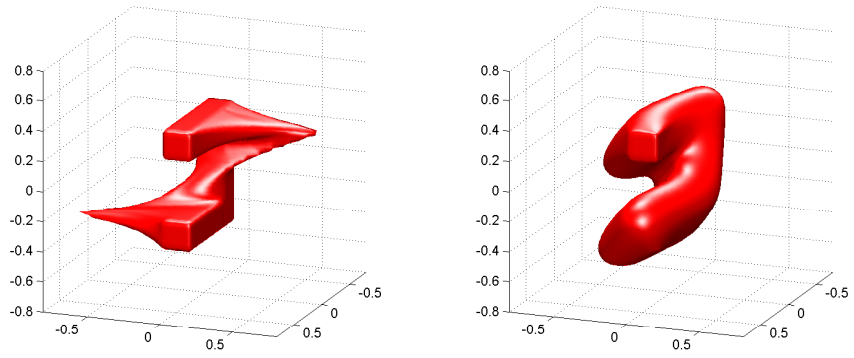
$$B = \begin{bmatrix} 0.1469 & -0.7988 & -2.3854 & -0.0054 & 1.3260 & -0.1623 \\ 0.2657 & 2.4582 & -0.3955 & -0.1403 & -0.1187 & 0.3601 \end{bmatrix}^T$$

and $u \in \mathbb{R}^2$, $\|u\| \leq 0.05$. We decompose this system into two 3D subsystems using the modified Riccati transformation with $\delta^* \approx -199$ (see Appendix B.3). A non-convex target set $\mathcal{K}$ is chosen such that in the transformed coordinates $\mathcal{Z}_\tau^i = \bigcup_{j=1}^3 \mathcal{P}_j$, $\forall i \in \{1, 2\}$, where $\mathcal{P}_j = \{z_i \in \mathbb{R}^3 \mid Cz_i - b_j \leq 0\}$ with

$$C^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \begin{aligned} b_1^T &= \begin{bmatrix} 0.5 & 0.3 & 0.1 & 0.1 & -0.2 & 0.4 \end{bmatrix}, \\ b_2^T &= \begin{bmatrix} -0.1 & 0.3 & 0.1 & 0.1 & 0.4 & 0.4 \end{bmatrix}, \\ b_3^T &= \begin{bmatrix} 0.5 & 0.3 & 0.1 & 0.1 & 0.4 & -0.2 \end{bmatrix}, \end{aligned}$$

and $z = [z_1^T, z_2^T]^T = T^{-1}x$, $x \in \mathcal{K}$.

Reachability computations are performed over a grid with 71 nodes in each dimension for $\tau = 2\,\text{s}$, as shown in Figure 4.5. The full-order 6D (minimal) reachable tube is the intersection of the back-projection of these 3D reachable tubes. The overall computation time was 489.89 s, whereas the actual reachable tube is prohibitively computationally expensive to compute with LS for any meaningful grid resolution. In addition, only 28 MB of RAM was used in the Riccati-based calculations, whereas computation of the full-order reachable tube would require over 4 TB of RAM (well beyond the capabilities of today's technology).



(a) Upper subspace (b) Lower subspace

Figure 4.5: The Riccati-based reachable tube (unsafe) in the transformed coordinates for Example 4.5.3.

4.5.4 Comparison With Schur-Based Decomposition (Chapter 3)

In Chapter 3 we presented a Schur-based decomposition technique that is applicable to almost any LTI system (subject to a mild assumption on the input-to-state map). In contrast, the decomposition method presented here is based on two nonsymmetric algebraic equations. The existence of solutions to these algebraic equations, however, is limited by a number of conditions on system matrices and is therefore heavily problem dependent. Indeed, as pointed out earlier, the conditions are more likely to be satisfied as the ill-conditioning of the original system matrices increases—e.g., for two-time-scale systems. (Figure 4.6 shows the fraction of tests on randomly generated systems for which a solution existed.) However, when the algebraic Riccati equations converge, the resulting subsystems could *potentially* yield less conservative reachable tube over-approximations than in the case of the Schur-based decomposition.

Consider a simple constrained 2D system with $A = \left[\begin{smallmatrix} -2.0228 & 0.9732 \\ -0.3695 & 0.0893 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 0.9600 \\ 0.1372 \end{smallmatrix}\right]$. Applying the modified Riccati transformation results in $A''_{\text{ric}} = \left[\begin{smallmatrix} -1.8360 & 0 \\ -0.0875 & -0.0975 \end{smallmatrix}\right]$, $B''_{\text{ric}} = \left[\begin{smallmatrix} 0.9600 \\ 0 \end{smallmatrix}\right]$. Notice that the "fast" eigenvalue $\lambda_1 = -1.8360$ is assigned to the controlled subsystem and that the mag-

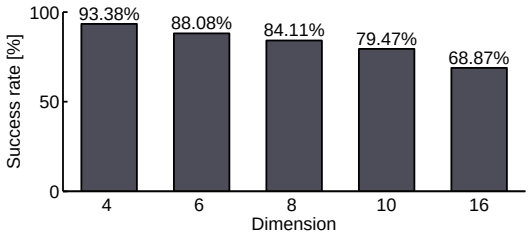


Figure 4.6: Fraction of a randomized test for which a Riccati solution existed. Success rate is shown as a percentage of 150 randomly generated two-time-scale systems for each dimension $n = 4, 6, 8, 10, 16$ with $n_k = \frac{n}{2}$. The A matrix entries of each system are drawn from a normal distribution with mean 0 and standard deviations 2 and 0.2 for A_{11} and A_{22} blocks respectively.

nitude of the unidirectional coupling is $\|\delta\mathscr{F}(Z)\| = 0.0875$. To retain the same eigen-structure (that is, the controlled subsystem be associated with the fast eigenvalue and the uncontrolled and perturbed subsystem with the slow eigenvalue $\lambda_2 = -0.0975$) using the Schur-based decomposition yields $A''_{\text{sch}} = \left[\begin{smallmatrix} -0.0975 & -0.1276 \\ 0 & -1.8360 \end{smallmatrix}\right]$, $B''_{\text{sch}} = \left[\begin{smallmatrix} 0 \\ -0.7948 \end{smallmatrix}\right]$. The corresponding subsystems are subject to a unidirectional coupling that is 46% larger in magnitude than in the Riccati-based case.

Suppose the target set is the Euclidean unit disk. This set is reshaped similarly under both transformations (Figure 4.7). Since the magnitude of the input-to-state map in the Schur-based case is smaller than that in the Riccati-based case, the optimal Hamiltonian for the reachability analysis of its controlled subsystem is also smaller. Therefore the reachable tube of the controlled subsystem in the Schur-based case is larger than in the Riccati-based case. In both decomposition techniques, the supremum of the reachable tube of the controlled subsystem as well as the unidirectional coupling between the subsystems are treated as disturbance in reachability computation of the ETUC subsystems. Since this disturbance has a larger magnitude in the Schur-based case, the overall reachable tube computation is more conservative than in the Riccati-based case. Therefore the Riccati-based decomposition is the superior approach in this example.

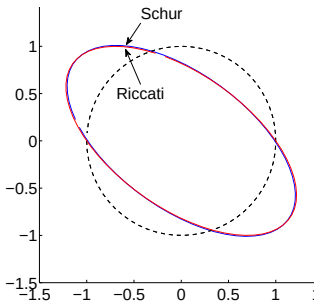


Figure 4.7: The unit disc target set under Riccati-based and Schur-based transformations. The corresponding reachable tube using the Riccati-based approach will be less conservative.

One should note, however, that no universal conclusions can be drawn from this particular example: In general, it is the problem under study (i.e. the system matrices, the shape of the target set in the transformed coordinate space, the effect of projections, etc.) that determines which decomposition method is the better choice.

4.5.5 The Decomposition and Maximal Reachability Analysis

With an unsafe target set, the Riccati-based approach presented here can also be used (in conjunction with both Eulerian and Lagrangian techniques) to over-approximate the *maximal* reachable tube, where the input u to the original system is considered as "disturbance" or "uncertainty" and is existentially quantified. This is done by replacing Step 3 in Algorithm 4.1 with $\mathcal{Z}^1_{[0,\tau]} \leftarrow Reach^\sharp_{[0,\tau]}(\mathcal{Z}^1_\tau, \mathcal{U})$ (and Steps 4–7 with $\mathcal{Z}^2_{[0,\tau]} \leftarrow Reach^\sharp_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{U})$ if the standard Riccati transformation is used to obtain the subsystems). As expected, we have

$$\mathrm{Proj}_{\mathbb{S}_1}(T^{-1} Reach^\sharp_{[0,\tau]}(\mathcal{K}, \mathcal{U})) \equiv Reach^\sharp_{[0,\tau]}(\mathcal{Z}^1_\tau, \mathcal{U}), \tag{4.57}$$

and

$$\mathrm{Proj}_{\mathbb{S}_2}(T^{-1} Reach^\sharp_{[0,\tau]}(\mathcal{K}, \mathcal{U})) \subseteq Reach^\sharp_{[0,\tau]}(\mathcal{Z}^2_\tau, \mathcal{B}(\|\delta\mathscr{F}(Z)\|\zeta)). \tag{4.58}$$

The over-approximation of the actual maximal reachable tube of the full-order system can be obtained similarly to Lemma 3.1 as

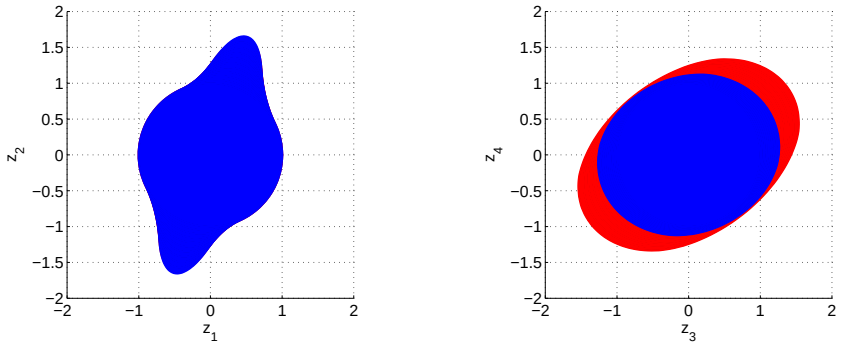
$$T\left((\mathcal{Z}^1_{[0,\tau]} \times \mathbb{S}_2) \cap (\mathbb{S}_1 \times \mathcal{Z}^2_{[0,\tau]})\right) \supseteq Reach^{\sharp}_{[0,\tau]}(\mathcal{K}, \mathcal{U}). \tag{4.59}$$

While Algorithm 4.1 with the above modifications ensures that the full-order (unsafe) maximal reachable tube is over-approximated, the results may be too conservative. Recall that the modified Riccati transformation is best suited to systems that are two-time-scale and ill-conditioned. The fast eigenvalues are assigned to the upper subsystem whereas the slow eigenvalues to the lower subsystem. Such an eigenvalue allocation is advantageous when computing the minimal reachable tube. In maximal reachability analysis, however, since the input is existentially quantified, the reachable tube in the upper subspace can grow significantly. As a result, the input to the lower subsystem, whose upper bound is directly proportional to the supremum of the reachable tube in the upper subspace, may become excessively large. Consequently, the reachable tube in the lower subspace may be overly conservative.

Note that a conservative over-approximation may be justified when the constraints are severely non-convex and/or cannot be adequately approximated by compact convex shapes (or a union of a few such sets), and therefore the computationally intensive Eulerian methods must be used to compute the maximal reachable tube. With convex constraints, however, the Lagrangian methods can be utilized to compute the full-order maximal reachable tube with great accuracy and efficiency. This will also circumvent the errors introduced via projections, etc. as compared to the case in which the Riccati-based approach is employed.

Consider a 4D example $\dot{x} = Ax + Bu$, $u \in \mathbb{R}$, $|u| \leq 0.1$, with

$$A = \begin{bmatrix} 0.5990 & -0.2333 & -0.0244 & 0.0121 \\ -0.4553 & -0.2107 & -0.0076 & -0.0041 \\ -0.0091 & 0.0128 & -0.1749 & -0.1768 \\ 0.0309 & -0.0229 & -0.0105 & -0.0777 \end{bmatrix}, \quad B = \begin{bmatrix} 0.6846 \\ 2.4813 \\ 0.3583 \\ -1.5195 \end{bmatrix}.$$



(a) (z_1, z_2) subspace (b) (z_3, z_4) subspace

Figure 4.8: Projections of the full-order maximal reachable tube (dark/blue) vs. the Riccati-based maximal reachable tube (light/red) in the transformed coordinates, computed using ET. (The two sets are indistinguishable in (z_1, z_2) subspace.)

The target set in the transformed coordinates is the Euclidean unit ball. We use the Riccati-based method to decompose this system into two 2D subsystems (see Appendix B.3). The computation time for this decomposition (including the search for $\delta^* \approx 12.6$) was 0.42 s. To over-approximate the maximal reachable tube of this system we employ ET [73] and perform lower dimensional reachability over sub-intervals for $\tau = 1$ s. Figure 4.8 compares the projections of the full-order reachable tube (computed in 1.08 s) and the lower-dimensional reachable tubes (collectively computed in 4.1 s) in the transformed coordinates. The greater computational time for the lower-dimensional reachability may be associated with the for-loops and other demanding computations involved in a sub-interval calculation.

4.6 Summary and Further Discussions

In this chapter we presented our second decomposition method for reachability analysis of LTI systems based on the Riccati transformation. This decomposition has considerable potential for reducing the computational complexity in reachability calculations, particularly for Eulerian methods.

For the case in which the input was non-disjoint across resulting subsystems, a modified Riccati transformation was proposed. The extension of this transformation-based reachability approach to switched/hybrid systems with LTI continuous dynamics can be easily achieved following the procedure described in Section 3.5.

Severely non-convex constraints and/or the need to compute the minimal reachable tube and the viability kernel as well as their corresponding safety-preserving control laws warrant the use of computationally intensive Eulerian methods. Despite inevitable conservatism e.g. due to the use of projection, our approach provides a means to compute the minimal reachable tube for higher dimensional systems for which these computationally intensive reachability tools were previously not applicable.

It is possible (although uncommon) that the transformation matrix can become poorly-conditioned due to pseudoinverses and numerical algorithms involved, resulting in the target set in the transformed coordinates becoming too severely distorted under the linear map to be of any practical use. An upper-bound on the condition number in terms of the system matrices and the free parameter δ is provided in Appendix B.2. We are currently investigating possible remedies that would ensure a well-conditioned transformation matrix.

Finally, it is worth emphasizing that the proposed modified Riccati transformation has application beyond reachability analysis. The technique can be used in any scenario in which decoupling of the dynamics as well as the input-to-state map is required.

Chapter 5

Set-Theoretic Methods: Lagrangian Algorithms for Viability[1]

In the following two chapters we will address Problem 2.2 and will present our second approach, based on set-theoretic methods, to reduce the complexity of the computation of the minimal reachable tube or the viability kernel. We will make use of the definitions and terminologies introduced in Section 2.5.

Consider a given constraint/target set $\mathcal{K} \subset \mathcal{X}$. While our main focus in Chapters 3 and 4 was on approximating the minimal reachable tube $Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ when $\mathcal{K}$ was deemed *unsafe*, in this chapter we will turn most of our attention to the case in which $\mathcal{K}$ is deemed *safe* and we seek to approximate the viability kernel $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ of this set. As mentioned before, approximating the minimal reachable tube and the viability kernel are not mutually exclusive as these constructs are duals of one another. For example, a method that facilitates an under-approximation of the viability kernel of $\mathcal{K}$ automatically provides a means for the over-approximation of the minimal reachable tube for $\mathcal{K}^c$.

Our approach is based on drawing a connection between the viability kernel and maximal reachable sets. As mentioned extensively in previous chapters, the Eulerian schemes normally used to compute the viability kernel suffer from a complexity that is exponential in the dimension of the states. In contrast, the efficient and scalable Lagrangian methods compute maximal

[1] A version of this chapter has been published in [55].

reachable sets. We will show that under certain conditions these methods can be employed to conservatively approximate the viability kernel for possibly high-dimensional systems. Significant reduction in the computational costs can be achieved since instead of a single calculation with exponential complexity one can perform a series of calculations with polynomial complexity.

5.1 Connection Between the Viability Kernel and Maximal Reachable Sets

We begin the analysis by considering the continuous-time case first and then proceed to the discrete-time case.

5.1.1 Continuous-Time Systems

Consider the case in which (2.1) is the continuous-time nonlinear system

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \quad t \in \mathbb{R}^+. \tag{5.1}$$

We will show that we can approximate $Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U})$ using a nested sequence of sets that are reachable in small sub-intervals of $[0, \tau]$.

Computing an Under-Approximation of the Viability Kernel

Assume that the vector field f is bounded by M in the norm $\|\cdot\|$. Consider a partition $P \in \mathscr{P}([0, \tau])$. We begin by defining an under-approximation of the state constraint set (Figure 5.1(a)):

$$\mathcal{K}_{\downarrow}(P) := \{x \in \mathcal{K} \mid \text{dist}(x, \mathcal{K}^c) \geq M\|P\|\}. \tag{5.2}$$

We under-approximate $\mathcal{K}$ by a distance $M\|P\|$ so as to only consider the system's state at discrete times $t_0, t_1, \ldots, t_n$. At a time t in the interval

$[t_i, t_{i+1}]$, a trajectory $x(\cdot)$ can travel a distance of at most

$$\|x(t) - x(t_i)\| \leq \int_{t_i}^{t} \|\dot{x}(\tau)\| d\tau \leq M(t - t_i) \leq M\|P\| \tag{5.3}$$

from its initial location $x(t_i)$. As we shall see, formulating the subset (5.2) will ensure that the state does not leave $\mathcal{K}$ at any time during $[0, \tau]$.

The set $\mathcal{K}_{\downarrow}(P)$ defines the first step of our recursion. We then define a sequence of $|P|$ sets recursively:

$$K_{|P|}(P) = \mathcal{K}_{\downarrow}(P), \tag{5.4a}$$

$$K_{k-1}(P) = \mathcal{K}_{\downarrow}(P) \cap Reach^{\sharp}_{t_k - t_{k-1}}(K_k(P), \mathcal{U}) \quad \text{for} \quad k \in \{1, \ldots, |P|\}. \tag{5.4b}$$

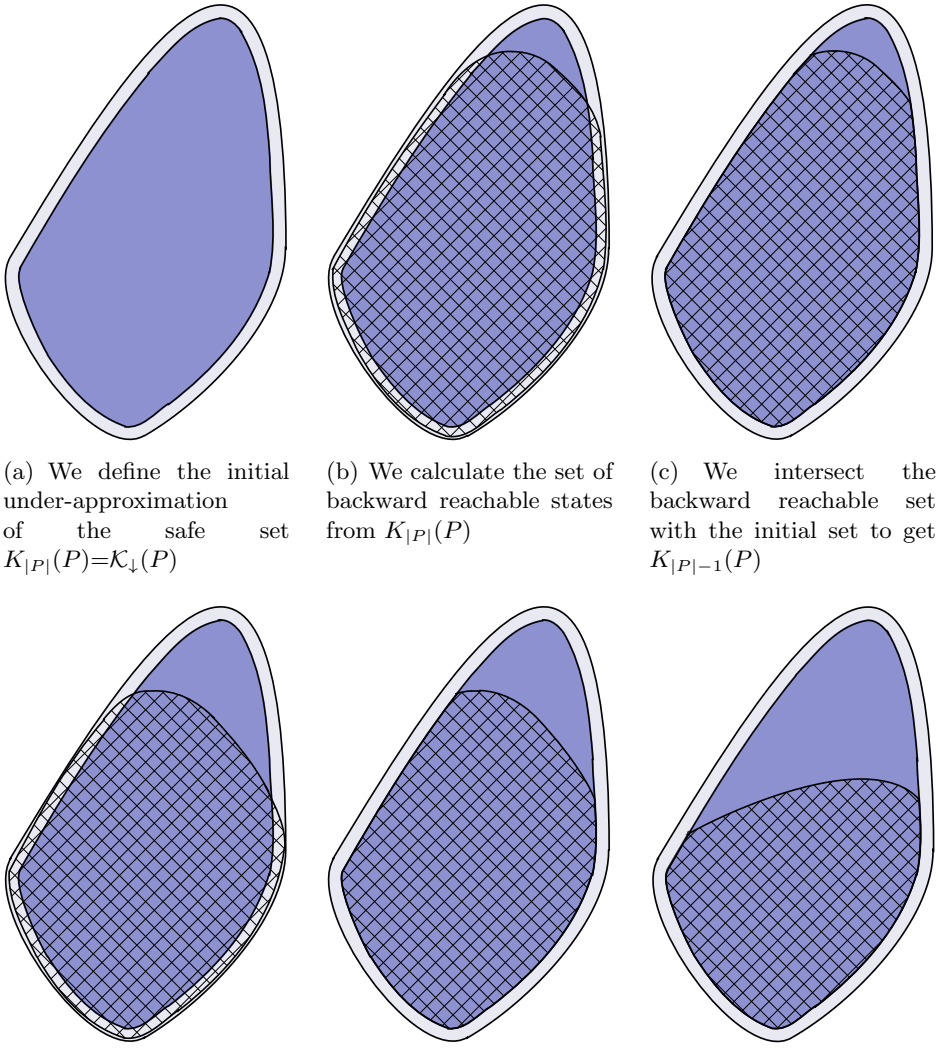
At each time step, we calculate the set of states from which you can reach $K_k(P)$, then intersect this set with the set of safe states (see Figure 5.1). The final set $K_0(P)$ is an approximation of $Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U})$.

Note that the resulting set depends on our choice of a partition P of the time interval $[0, \tau]$. We claim that for any partition P, $K_0(P)$ is an under-approximation.

Proposition 5.1. *Suppose that the vector field $f\colon \mathcal{X} \times \mathcal{U} \to \mathcal{X}$ is bounded on a set $\mathcal{K} \subseteq \mathcal{X}$. Then for any partition P of $[0, \tau]$ the final set $K_0(P)$ defined by the recurrence relation (5.4) satisfies*

$$K_0(P) \subseteq Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U}). \tag{5.5}$$

Proof. Since f is bounded on $\mathcal{K}$, there exists a norm $\|\cdot\|$ and a real number $M > 0$ with $\|f(x, u)\| \leq M$ for all $x \in \mathcal{K}$. Now, fix a partition P of $[0, \tau]$ and take a point $x_0 \in K_0(P)$. By the construction of $K_0(P)$, this means that for each $k = 1, \ldots, |P|$ there is some point $x_k \in K_k(P)$ and an input $u_k : [0, t_k - t_{k-1}] \to \mathcal{U}$ such that x_k can be reached from x_{k-1} at time $t_k - t_{k-1}$ using input u_k. Thus, taking the concatenation of the inputs u_k, we get an input $u : [0, \tau] \to \mathcal{U}$ such that the solution $x : [0, \tau] \to \mathcal{X}$ to the initial value problem $\dot{x} = f(x, u)$, $x(0) = x_0$, satisfies $x(t_k) = x_k \in K_k(P) \subseteq \{x \in \mathcal{K} \mid$



(a) We define the initial under-approximation of the safe set $K_{|P|}(P)$=$\mathcal{K}_{\downarrow}(P)$

(b) We calculate the set of backward reachable states from $K_{|P|}(P)$

(c) We intersect the backward reachable set with the initial set to get $K_{|P|-1}(P)$

(d) Next, we calculate the set of backward reachable states from $K_{|P|-1}(P)$

(e) Again, we intersect the backward reachable set with the initial set to get a new set $K_{|P|-2}(P)$

(f) By repeating this process, we reach an under-approximation $K_0(P)$ of the viability kernel.

Figure 5.1: Iteratively constructing an under-approximation of $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$.

$\text{dist}(x, \mathcal{K}^c) \geq M\|P\|\}$. We claim that this guarantees that $x(t) \in \mathcal{K}$ for all $t \in [0, \tau]$. Indeed, any $t \in [0, \tau)$ lies in some interval $[t_k, t_{k+1})$. Since f is bounded by M, we have

$$\|x(t) - x(t_k)\| \leq M(t - t_k) < M(t_{k+1} - t_k) \leq M\|P\|. \tag{5.6}$$

Further, $x(t_k) \in K_k(P)$ implies that $\text{dist}(x(t_k), \mathcal{K}^c) \geq M\|P\|$. Combining these, we see that

$$\begin{aligned} \text{dist}(x(t), \mathcal{K}^c) &\geq \text{dist}(x(t_k), \mathcal{K}^c) - \|x(t) - x(t_k)\| \\ &> M\|P\| - M\|P\| = 0 \end{aligned} \tag{5.7}$$

and hence $x(t) \in \mathcal{K}$. Thus, $x_0 \in Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U})$. □

Remark 5.1. *For a large enough horizon τ, if $K_{k-1}(P) \equiv K_k(P)$ for some $k \in \{1, \ldots, |P|\}$ with $t_k \in (0, \tau]$, then $K_0(P) = K_k(P)$ approximates the infinite-horizon viability kernel $Viab_{\mathbb{R}^+}(\mathcal{K}, \mathcal{U})$. This set is also known as the maximal controlled-invariant subset [12] (see Section 2.1).*

Precision of the Approximation

The approximation can be made to be arbitrarily precise by choosing a sufficiently fine partition. This is true in the sense that the union of the approximating sets $K_0(P)$ taken over all possible partitions P of $[0, \tau]$ is bounded between the viability kernels of $\mathcal{K}$ and its interior $\mathring{\mathcal{K}}$.

Proposition 5.2. *Suppose that the vector field $f : \mathcal{X} \times \mathcal{U} \to \mathcal{X}$ is bounded on a set $\mathcal{K} \subseteq \mathcal{X}$. Then we have*

$$Viab_{[0,\tau]}(\mathring{\mathcal{K}}, \mathcal{U}) \subseteq \bigcup_{P \in \mathscr{P}([0,\tau])} K_0(P) \subseteq Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U}). \tag{5.8}$$

In particular, when $\mathcal{K}$ is open,

$$\bigcup_{P \in \mathscr{P}([0,\tau])} K_0(P) = Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U}). \tag{5.9}$$

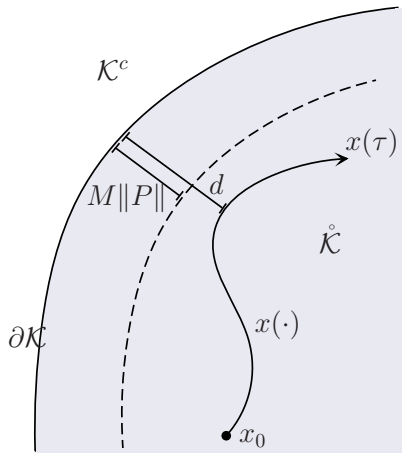


Figure 5.2: Distance $d > 0$ from the boundary set $\partial\mathcal{K}$. Partition P can be chosen such that $M\|P\| < d$.

Proof. The second inclusion in (5.8) follows directly from Proposition 5.1. To prove the first inclusion, take a state $x_0 \in Viab_{[0,\tau]}(\mathring{\mathcal{K}},\mathcal{U})$. There exists an input $u : [0,\tau] \to \mathcal{U}$ such that the solution $x(\cdot)$ to the initial value problem $\dot{x} = f(x,u)$, $x(0) = x_0$, satisfies $x(t) \in \mathring{\mathcal{K}}$ for all $t \in [0,\tau]$. Since $\mathring{\mathcal{K}}$ is open, for any $x \in \mathring{\mathcal{K}}$ we have $\mathrm{dist}(x,\mathcal{K}^c) > 0$. Further, $x\colon [0,\tau] \to \mathcal{X}$ is continuous so the function $t \mapsto \mathrm{dist}(x(t),\mathcal{K}^c)$ is continuous on the compact set $[0,\tau]$. Thus, we can define $d > 0$ to be its minimum value. Now take a partition P of $[0,\tau]$ such that $M\|P\| < d$ (see Figure 5.2). We need to show that $x_0 \in K_0(P)$.

First note that our partition P is chosen such that $\mathrm{dist}(x(t),\mathcal{K}^c) > M\|P\|$ for all $t \in [0,\tau]$. Hence $x(t_k) \in K_{|P|}(P)$ for all $k = 0,\ldots,|P|$. To show that $x(t_{k-1}) \in Reach^{\sharp}_{t_k - t_{k-1}}(K_k(P),\mathcal{U})$ for all $k = 1,\ldots,|P|$, consider the tokenization[2] $\{u_k\}_k$ of the input u corresponding to P. It is easy to verify that for all k, we can reach $x(t_k)$ from $x(t_{k-1})$ at time $t_k - t_{k-1}$ using input u_k. Thus, in particular, we have $x_0 = x(t_0) \in Reach^{\sharp}_{t_1 - t_0}(K_1(P),\mathcal{U})$. So $x_0 \in K_0(P)$. Hence $Viab_{[0,\tau]}(\mathring{\mathcal{K}},\mathcal{U}) \subseteq \bigcup_{P\in\mathscr{P}([0,\tau])} K_0(P)$. □

[2]See Definition 2.11.

5.1.2 Discrete-Time Systems

Consider the case in which (2.1) is the discrete-time nonlinear system

$$x(t+1) = f(x(t), u(t)), \quad x(0) = x_0, \quad t \in \mathbb{Z}^+. \tag{5.10}$$

Computing $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ under this system is a particular case of the results presented in Section 5.1.1. Define a sequence of sets recursively as

$$K_n = \mathcal{K}, \tag{5.11a}$$

$$K_{k-1} = \mathcal{K} \cap Reach_1^\sharp(K_k,\mathcal{U}) \quad \text{for} \quad k \in \{1,\dots,n\}, \tag{5.11b}$$

where $\tau = n$ and $Reach_1^\sharp(\cdot,\cdot)$ is the unit time-step maximal reachable set.

Proposition 5.3. *Let K_0 be the final set obtained from the recurrence relation* (5.11)*. Then,*

$$Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) = K_0. \tag{5.12}$$

Proof. Without loss of generality we assume that the time variable t is integer valued. As a result, the tokenization of the input signal u is a discrete sequence $\{u_k\}_k$ with $u_k := u(t)$ with $t = k-1$ for $k = 1,\dots,n$.

To show $K_0 \subseteq Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$, via recursion (5.11) we have that at each step k there exists u_k such that $x_{k-1} \in K_{k-1}$ reaches $x_k \in K_k$. Thus, $x_0 \in K_0$ implies there exists a concatenation $u(\cdot) = \{u_k\}_k \in \mathscr{U}_{[0,\tau]}$ such that $x(t) \in \mathcal{K}$ for all $t \in [0,\tau]$. Therefore, $x_0 \in Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$.

To show $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq K_0$, take $x_0 \in Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$. There exists $u(\cdot) = \{u_k\}_k$ such that $x(t) \in \mathcal{K}$ for every t. Using the tokenization of $\{u_k\}_k$ we can verify that for some u_k we can reach $x_k := x(t+1)$ from $x_{k-1} := x(t)$. Hence, $x_{k-1} \in Reach_1^\sharp(K_k,\mathcal{U})$ for all $k \in \{1,\dots,n\}$. In particular, for $k = 1$ we have $x_0 := x(0) \in Reach_1^\sharp(K_1,\mathcal{U})$. Thus, $x_0 \in \mathcal{K} \cap Reach_1^\sharp(K_1,\mathcal{U}) = K_0$. □

Remark 5.2. *Note that the above iterative scheme is closely related to the set-valued description of the discrete viability kernel presented in [18, 106]*

and the recursive construction of the controlled-invariant subset for discrete-time systems presented in [11, 12, 61, 120].

5.2 Computational Algorithms

Any technique that is capable of computing the maximal reachable set can now be used to compute the viability kernel. Most currently available Lagrangian methods yield an (under- and/or over-) approximation of the maximal reachable set. The viability kernel should not be over-approximated since an over-approximation would contain initial states for which the viability of the system is inevitably at stake. Thus, to correctly compute $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ all approximations must be in the form of under-approximations.

Every step of the recursions (5.4) and (5.11) involves a reachability computation and an intersection operation. Ideally, the sets that are being intersected should be drawn from classes of shapes that are closed under such an operation, e.g. polytopes. However, the currently available reachability techniques that are based on polytopes (e.g. [76]) do not, in general, scale well with the dimension of the state. Moreover, the scalable reachability techniques, such as the methods of zonotopes [40], ellipsoids [70, 74], and support functions [39], generate sets that may prove to be difficult to transform into a (under-approximating) polytope. For instance, one may compute a polytopic under-approximation of the reachable sets using their support functions based on the approach presented in [77]. However, that approach requires calculation of the facet representation of the resulting polytopes from their vertices before each intersection operation, which is known to be computationally demanding in higher dimensions.

Recently, the authors in [54] introduced an efficient polytopic technique with a fixed complexity called *bounded vertex representation* that is capable of computing under-approximations of the maximal reachable set and the intersection of polytopes. Unfortunately, this method assumes that the dynamics are affine (of the form $Ax+b$) which circumvents the difficulties that arise from linear transformations and Minkowski sums in the reachability

operation.

5.2.1 A Piecewise Ellipsoidal Approach

Here we showcase our results using an efficient algorithm, based on the ellipsoidal techniques [70] implemented in Ellipsoidal Toolbox (ET) [73], that sacrifices accuracy in exchange for scalability. We consider the LTI system

$$\mathcal{L}(x(t)) = Ax(t) + Bu(t) \tag{5.13}$$

with $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. As in (2.1), depending on whether the system evolves in continuous time ($t \in \mathbb{R}^+$) or discrete time ($t \in \mathbb{Z}^+$), $\mathcal{L}(\cdot)$ denotes the derivative operator or the unit forward shift operator, respectively.

An *ellipsoid* in $\mathbb{R}^n$ is defined as

$$\mathcal{E}(q, Q) := \left\{x \in \mathbb{R}^n \mid \langle (x-q), Q^{-1}(x-q)\rangle \leq 1\right\} \tag{5.14}$$

with center $q \in \mathbb{R}^n$ and shape matrix $\mathbb{R}^{n\times n} \ni Q = Q^T \succ 0$. A *piecewise ellipsoidal* set is the union of a finite number of ellipsoids.

Among many advantages, ellipsoidal techniques [70, 73] allow for an efficient computation of *under-approximations* of the maximal reachable sets, making them a particularly attractive choice for the reachability computations involved in our formulation of the viability kernel.

Suppose $\mathcal{K}$ and $\mathcal{U}$ are (or can be closely under-approximated as) compact ellipsoids with nonempty interior. Consider the continuous-time case and the recursion (5.4). (The arguments in the discrete-time case are similar.) Given a partition P and some $k \in \{1, \ldots, |P|\}$, let $K_k(P) = \mathcal{E}(x_\delta, X_\delta) \subset \mathcal{X}$. As in [71], with $\mathcal{N} := \{v \in \mathbb{R}^n \mid \langle v, v\rangle = 1\}$ and $\delta := t_k - t_{k-1}$ we have

$$Reach^\sharp_{\delta - t}(K_k(P), \mathcal{U}) = \bigcup_{\ell_\delta \in \mathcal{N}} \mathcal{E}(x^c(t), X^-_\ell(t)), \quad \forall t \in [0, \delta], \tag{5.15}$$

where $x^c(t)$ and $X^-_\ell(t)$ are the center and the shape matrix of the *internal approximating* ellipsoid at time t that is tangent to $Reach^\sharp_{\delta - t}(K_k(P), \mathcal{U})$ in

the direction $\ell(t) \in \mathbb{R}^n$. For a fixed $\ell(\delta) = \ell_\delta \in \mathcal{N}$, the direction $\ell(t)$ is obtained from the adjoint equation $\dot{\ell}(t) = -A^T\ell(t)$. The center $x^c(t)$ (with $x^c(\delta) = x_\delta$) and the shape matrix $X_\ell^-(t)$ (with $X_\ell^-(\delta) = X_\delta$) are determined from differential equations described in [72]. (cf. [74] for their discrete-time counterparts.)

In practice, only a finite number of directions is used for the maximal reachable set computations. Let $\mathcal{M}$ be a finite subset of $\mathcal{N}$. Then,

$$Reach^\sharp_{\delta-t}(K_k(P),\mathcal{U}) \supseteq \bigcup_{\ell_\delta \in \mathcal{M}} \mathcal{E}(x^c(t), X_\ell^-(t)), \quad \forall t \in [0,\delta]. \tag{5.16}$$

Note that the under-approximation in (5.16) is in general an arbitrarily shaped, non-convex set. Performing our desired operations on this set while maintaining efficiency may be difficult, if not impossible.

Instead consider the final backward reachable set $Reach^\sharp_\delta(K_k(P),\mathcal{U})$ and let $Reach^{\sharp[\ell_\delta]}_\delta(K_k(P),\mathcal{U})$ denote the maximal reachable set corresponding to a single terminal direction $\ell_\delta := \ell(\delta) \in \mathcal{M}$. We have that

$$\begin{aligned} Reach^{\sharp[\ell_\delta]}_\delta(K_k(P),\mathcal{U}) &= \mathcal{E}(x^c(0), X_\ell^-(0)) \\ &\subseteq \bigcup_{\ell_\delta \in \mathcal{M}} \mathcal{E}(x^c(0), X_\ell^-(0)) \\ &\subseteq Reach^\sharp_\delta(K_k(P),\mathcal{U}). \end{aligned} \tag{5.17}$$

Therefore, the reachable set computed for a single direction is an *ellipsoidal subset* of the actual reachable set.

Let $\bigcirc\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ denote a set-valued function that maps a set to its maximum volume inscribed ellipsoid. Algorithms 5.1 and 5.2 compute a piecewise ellipsoidal under-approximation of $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ for continuous-time and discrete-time systems, respectively.

Algorithm 5.1 Piecewise ellipsoidal approximation of $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ (continuous-time)

Choose $P \in \mathscr{P}([0,\tau])$ ▷ Affects precision of approximation.
$K_{|P|}(P) \leftarrow \mathcal{K} \ominus \mathcal{B}(M\|P\|)$ ▷ Find $\{x \in \mathcal{K} \mid \mathrm{dist}(x, \mathcal{K}^c) \geq M\|P\|\}$.
$K_0^*(P) \leftarrow \emptyset$
while $\mathcal{M} \neq \emptyset$ **do**
    $l \leftarrow \ell_\tau \in \mathcal{M}$
    $k \leftarrow |P|$
    while $k \neq 0$ **do**
        if $K_k(P) = \emptyset$ **then**
            $K_0(P) \leftarrow \emptyset$
            break
        end if
        $\mathcal{G} \leftarrow Reach^{\sharp[l]}_{t_k - t_{k-1}}(K_k(P), \mathcal{U})$
            ▷ Compute the ellipsoidal under-approximation of the maximal reach set along the direction l.
        $K_{k-1}(P) \leftarrow \bigcirc(K_{|P|}(P) \cap \mathcal{G})$
            ▷ Find the max volume inscribed ellipsoid in $K_{|P|}(P) \cap \mathcal{G}$.
        $k \leftarrow k - 1$
    end while
    $K_0^*(P) \leftarrow K_0^*(P) \cup K_0(P)$
    $\mathcal{M} \leftarrow \mathcal{M} \backslash \{l\}$
end while
return $(K_0^*(P))$

Algorithm 5.2 Piecewise ellipsoidal approximation of $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ (discrete-time)

$K_n \leftarrow \mathcal{K}$
$K_0^* \leftarrow \emptyset$
while $\mathcal{M} \neq \emptyset$ **do**
    $l \leftarrow \ell_\tau \in \mathcal{M}$
    $k \leftarrow n$
    while $k \neq 0$ **do**
        if $K_k = \emptyset$ **then**
            $K_0 \leftarrow \emptyset$
            break
        end if
        $\mathcal{G} \leftarrow Reach_1^{\sharp[l]}(K_k,\mathcal{U})$
        $K_{k-1} \leftarrow \bigcirc(K_n \cap \mathcal{G})$
        $k \leftarrow k-1$
    end while
    $K_0^* \leftarrow K_0^* \cup K_0$
    $\mathcal{M} \leftarrow \mathcal{M} \backslash \{l\}$
end while
return (K_0^*)

Proposition 5.4. *For a given partition $P \in \mathscr{P}([0,\tau])$, let $K_0^*(P)$ be the set generated by Algorithm 5.1. Then,*

$$K_0^*(P) \subseteq Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}). \tag{5.18}$$

Proof. Let $\widetilde{K}_0(P)$ denote the final set constructed recursively by (5.4). Also, for a fixed direction l, let $K_0^{[l]}(P)$ denote the set produced at the end of each outer loop in Algorithm 5.1. Notice that via (5.17), for every $l \in \mathcal{M}$, $K_0^{[l]}(P) \subseteq \widetilde{K}_0(P)$. Therefore, $\bigcup_{l\in\mathcal{M}} K_0^{[l]}(P) \subseteq \widetilde{K}_0(P)$. Thus, $K_0^*(P) = \bigcup_{l\in\mathcal{M}} K_0^{[l]}(P) \subseteq Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$. □

Remark 5.3. *A similar argument holds for the discrete-time case in Algorithm 5.2, i.e. $K_0^* \subseteq Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$.*

$\bigcirc(\cdot)$: Computing the Maximum Volume Inscribed Ellipsoid

Notice that in the continuous-time case, the sets $\mathcal{Y} := K_{|P|}(P)$ and $\mathcal{G} := Reach^{\sharp[l]}_{t_k - t_{k-1}}(K_k(P), \mathcal{U})$ are compact ellipsoids for every $l \in \mathcal{M}$, $P \in \mathscr{P}([0,\tau])$, and $k \in \{1,\dots,|P|\}$. Similarly in the discrete-time case, $\mathcal{Y} := K_n$ and $\mathcal{G} := Reach^{\sharp[l]}_{1}(K_k, \mathcal{U})$ are compact ellipsoids for every $l \in \mathcal{M}$ and $k \in \{1,\dots,n\}$. Their intersection is, in general, not an ellipsoid but can be easily under-approximated by one. The operation $\bigcirc(\cdot)$ under-approximates this intersection by computing the maximum volume inscribed ellipsoid in $\mathcal{Y} \cap \mathcal{G}$. The result is an ellipsoid that, while aiming to minimize the accuracy loss, can be used directly as the target set for the reachability computation in the subsequent time step.

Let us rewrite the general ellipsoid as $\mathcal{E}(q,Q) = \{Hx + q \mid \|x\|_2 \leq 1\}$ with $H = Q^{\frac{1}{2}}$. Assume $\mathcal{Y} \cap \mathcal{G} \neq \emptyset$ and suppose $\mathcal{Y} = \mathcal{E}(q_1, Q_1)$ and $\mathcal{G} = \mathcal{E}(q_2, Q_2)$. Following [15], the computation of the maximum volume inscribed ellipsoid in $\mathcal{Y} \cap \mathcal{G}$ (a readily-available feature in ET) can be cast as a convex semidefinite program (SDP):

$$\underset{H \in \mathbb{R}^{n\times n},\, q \in \mathbb{R}^n,\, \lambda_i \in \mathbb{R}}{\text{minimize}} \quad \log\det H^{-1} \tag{5.19a}$$

$$\text{subject to} \quad \lambda_i > 0 \tag{5.19b}$$

$$\begin{bmatrix} 1-\lambda_i & 0 & (q-q_i)^T \\ 0 & \lambda_i I & H \\ q-q_i & H & Q_i \end{bmatrix} \succeq 0, \quad i = 1,2. \tag{5.19c}$$

Using the optimal values for H and q, we will have

$$\bigcirc(\mathcal{Y} \cap \mathcal{G}) = \mathcal{E}(q, H^T H). \tag{5.20}$$

Loss of Accuracy

A set generated by Algorithms 5.1 or 5.2 could be an inaccurate approximation of $Viab_{[0,\tau]}(\mathcal{K}, \mathcal{U})$, especially for large time horizons. The loss of accuracy is mainly attributed to the function $\bigcirc(\cdot)$, the under-approximation of the intersection at every iteration with its maximum volume inscribed ellip-

soid. This approximation error propagates through the algorithms making them subject to the "wrapping effect" (cf. [66]).

In the continuous-time case, the quality of approximation is also affected by the choice of time interval partition (Proposition 5.2). Choosing a finer partition increases the quality of approximation. However, doing so would also require a larger number of intersections to be performed in the intermediate steps of the recursion. As such, one would expect that the error generated by $\bigcirc(\cdot)$ would be amplified. Luckily, since with a finer partition the reachable sets change very little from one time step to the next, the intersection error at every iteration becomes smaller. The end result is a smaller accumulated error and therefore a better approximation, at least experimentally.

We show this using a trivial example: Consider the double integrator

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \tag{5.21}$$

subject to ellipsoidal constraints $u(t) \in \mathcal{U} := [-0.25, 0.25]$ and $x(t) \in \mathcal{K} := \mathcal{E}(\mathbf{0}, \left[\begin{smallmatrix} 0.25 & 0 \\ 0 & 0.25 \end{smallmatrix}\right])$, $\forall t \in [0, 1]$. We employ eight different partitions P of the time interval such that we have $|P| = 13, 21, 34, 55, 89, 144, 233, 377$ and $\|P\| = 1/|P|$. The linear vector field is bounded on $\mathcal{K}$ in the infinity norm by $M = \|\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]\| \sup_{x \in \mathcal{K}} \|x\| + \|\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]\| \sup_{u \in \mathcal{U}} \|u\| = 0.75$. Thus, in Algorithm 5.1, $K_{|P|}(P) = \mathcal{K} \ominus \mathcal{B}(0.75 \times \|P\|)$. A piecewise ellipsoidal under-approximation of $Viab_{[0,1]}(\mathcal{K}, \mathcal{U})$ for every partition P (with $|\mathcal{M}| = 10$ randomly chosen initial directions) is shown in Figure 5.3. Notice that as $|P|$ increases, the fidelity of approximation improves. A plot of the error in the under-approximation as a function of $|P|$ is provided in Figure 5.4.

Forming the Under-Approximation $\mathcal{K}_{\downarrow}(P)$ in the Continuous-Time Case

While a straightforward method to construct the under-approximation $\mathcal{K}_{\downarrow}(P)$ of the set $\mathcal{K}$ in Algorithm 5.1 for a fixed $P \in \mathscr{P}([0, \tau])$ is to erode $\mathcal{K}$ by a ball of radius $M\|P\|$ (for a given uniform bound M on $f(x, u) = Ax + Bu$),

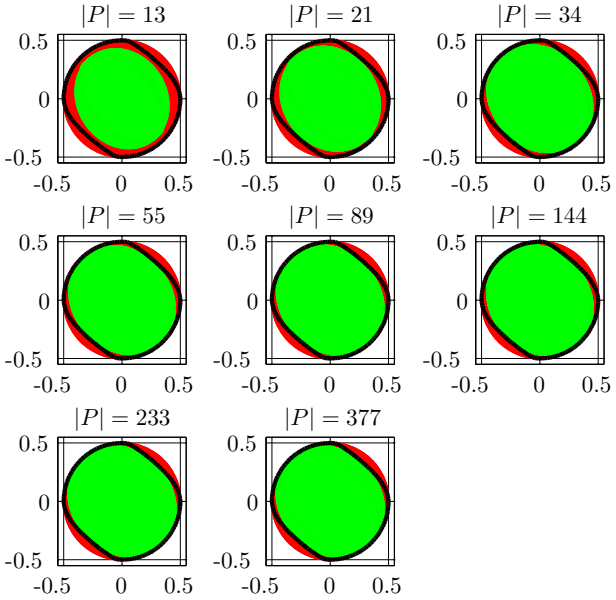


Figure 5.3: For the set $\mathcal{K}$ (red), $K_0(P)$ (green/light) under-approximates $Viab_{[0,1]}(\mathcal{K},\mathcal{U})$ (outlined in thick black lines via [87]) using Algorithm 5.1 under the double integrator dynamics. A finer time interval partition results in better approximation.

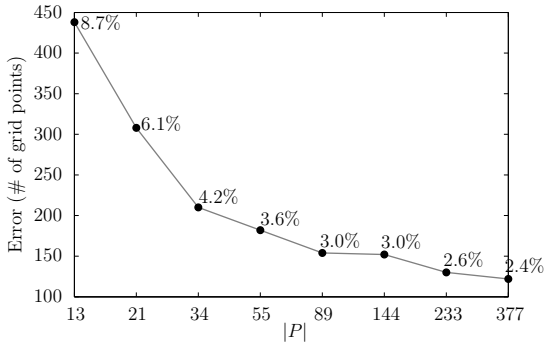


Figure 5.4: Convergence plot of the error as a function of $|P|$ for the double-integrator example. Error is quantified as the fraction of grid points (total of 71×71) contained in the set difference between the level set approximation of the viability kernel and its piecewise ellipsoidal under-approximation.

this under-approximation may be too conservative—particularly when the constraints $\mathcal{K}$ and $\mathcal{U}$ are nonsymmetric with respect to the origin. In such a case we would typically have to reevaluate our time interval partitioning and choose a P with extremely short sub-intervals so that it offsets the potential conservatism. This is clearly not an ideal approach.

Instead consider the respective linear differential inclusion

$$\dot{x}(t) \in A\mathcal{K} \oplus B\mathcal{U}. \tag{5.22}$$

Notice that the set $A\mathcal{K} \oplus B\mathcal{U}$ is a compact (closed and bounded) subset of $\mathcal{X}$ since both $\mathcal{K}$ and $\mathcal{U}$ are compact. With this in mind, a possibly less conservative approach to construct $\mathcal{K}_{\downarrow}(P)$ may be as follows.

1. Over-approximate the set $A\mathcal{K} \oplus B\mathcal{U}$ by computing tight enclosing ellipsoids over a finite subset of directions $\mathcal{M} \subset \{l \in \mathbb{R}^n \mid \langle l, l \rangle = 1\}$ and choose the one with the smallest volume:

$$\mathcal{E}(\sigma, \Sigma) := \min_{l \in \mathcal{M}} \operatorname{vol} \left(A\mathcal{K} \overset{\mathcal{E}_l^+}{\oplus} B\mathcal{U} \right). \tag{5.23}$$

   Here we have used $\overset{\mathcal{E}_l^+}{\oplus}$ to denote the tight *enclosing* ellipsoidal approximation of the Minkowski sum in the direction l.

2. Multiply this set by a factor of $\|P\|$:

$$\|P\| \mathcal{E}(\sigma, \Sigma) = \mathcal{E}(\|P\|\sigma, \|P\|^2 \Sigma). \tag{5.24}$$

3. Under-approximate the set $\mathcal{K} \ominus \mathcal{E}(\|P\|\sigma, \|P\|^2\Sigma)$ by computing tight enclosed ellipsoids over directions in $\mathcal{M}$ and choose the one with the largest volume to obtain $\mathcal{K}_{\downarrow}(P)$:

$$\mathcal{K}_{\downarrow}(P) = \max_{l \in \mathcal{M}} \operatorname{vol} \left(\mathcal{K} \overset{\mathcal{E}_l^-}{\ominus} \mathcal{E}(\|P\|\sigma, \|P\|^2\Sigma) \right). \tag{5.25}$$

   Here the operator $\overset{\mathcal{E}_l^-}{\ominus}$ returns the tight *enclosed* ellipsoidal approxi-

mation of the Pontryagin difference (or the erosion operator) between its operands in the direction l.

All of the above steps can be easily and efficiently performed in ET.

5.3 Practical Examples

All computations are performed on a dual core Intel-based computer with 2.8 GHz CPU, 6 MB of L2 cache and 3 GB of RAM running single-threaded 32-bit MATLAB 7.5.

5.3.1 Flight Envelope Protection (Continuous-Time)

Consider the linearized longitudinal aircraft dynamics $\dot{x}(t) = Ax(t) + B\delta_e(t)$,

$$A = \begin{bmatrix} -0.003 & 0.039 & 0 & -0.322 \\ -0.065 & -0.319 & 7.740 & 0 \\ 0.020 & -0.101 & -0.429 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 0.010 \\ -0.180 \\ -1.160 \\ 0 \end{bmatrix}$$

with state $x = [u, v, \dot{\theta}, \theta]^T \in \mathbb{R}^4$ comprised of deviations in aircraft velocity [ft/s] along and perpendicular to body axis, pitch-rate [crad/s],[3] and pitch angle [crad] respectively, and with input $\delta_e \in [-13.3°, 13.3°] \subseteq \mathbb{R}$ the elevator deflection. These matrices represent stability derivatives of a Boeing 747 cruising at an altitude of 40 kft with speed 774 ft/s [16]. The state constraint set

$$\mathcal{K} = \mathcal{E}\left(\begin{bmatrix} 0 \\ 0 \\ 2.18 \\ 0 \end{bmatrix}, \begin{bmatrix} 1075.84 & 0 & 0 & 0 \\ 0 & 67.24 & 0 & 0 \\ 0 & 0 & 42.7716 & 0 \\ 0 & 0 & 0 & 76.0384 \end{bmatrix} \right)$$

represents an ellipsoidal (under-)approximation of the flight envelope. We require $x(t) \in \mathcal{K} \; \forall t \in [0, 2]$.

[3] crad $= 0.01$ rad $\approx 0.57°$.

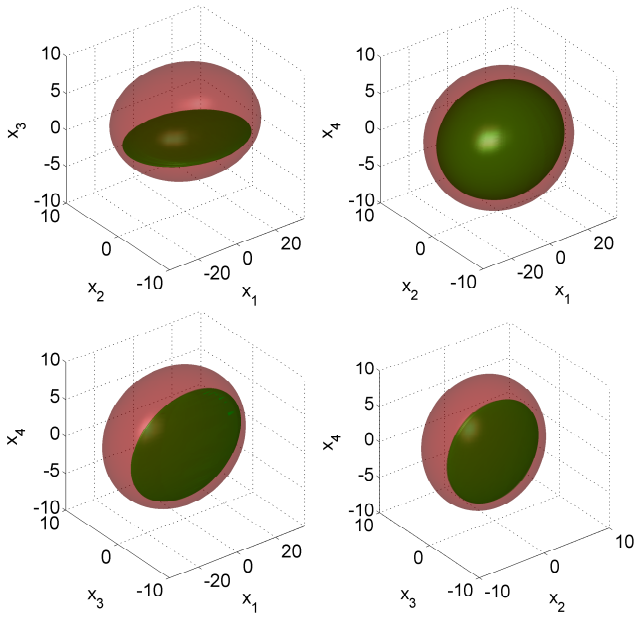


Figure 5.5: 3D projections of the under-approximation of $Viab_{[0,2]}(\mathcal{K},\mathcal{U})$ for Example 5.3.1. The flight envelope $\mathcal{K}$ is the red/light transparent region. The green/dark piecewise ellipsoidal sets under-approximate the viability kernel.

A uniform partition P is chosen such that $|P| = 400$. Algorithm 5.1 (with $|\mathcal{M}| = 8$ consisting of the standard basis vectors in $\mathbb{R}^4$ and four additional randomly generated vectors) computes via ET a piecewise ellipsoidal under-approximation of the viability kernel $Viab_{[0,2]}(\mathcal{K},\mathcal{U})$ as shown in Figures 5.5 and 5.6. Note that for any state belonging to this set, there exists an input that can protect the flight envelope over the specified time horizon. The overall computation time was roughly 10 mins. In comparison, the level set approximation of the viability kernel (also shown in Figure 5.6) is computed in 5.4 hrs with significantly larger memory footprint over a grid with 45 nodes in each dimension (still a rather coarse grid) using the Level Set Toolbox [87]. Since the computed sets are 4D, we plot a series of 3D and 2D projections of these 4D objects.

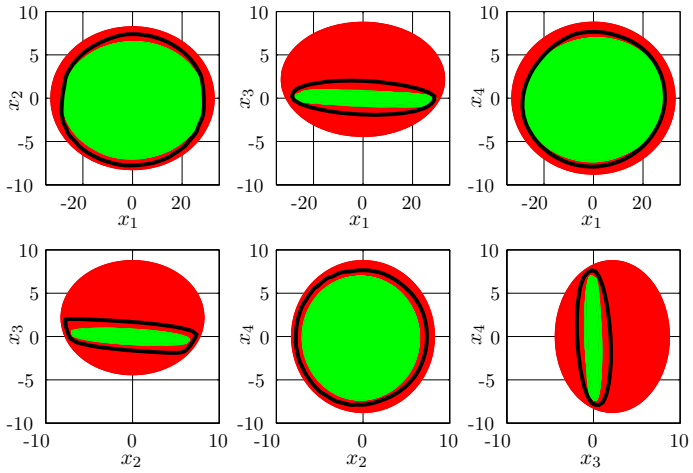


Figure 5.6: 2D projections of the under-approximation of $Viab_{[0,2]}(\mathcal{K}, \mathcal{U})$ for Example 5.3.1. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal under-approximation of the viability kernel (green/light) are shown. The level set approximation of the viability kernel, computed via [87], is outlined in thick black lines.

5.3.2 Safety in Anesthesia Automation (Discrete-Time)

To improve patient recovery, lessen anesthetic drug usage, and reduce time spent at drug saturation levels, a variety of approaches to controlling depth of anesthesia have been proposed e.g. in [10, 26, 52, 85, 97, 109, 115].

Over the past few years, an interdisciplinary team of researchers at the University of British Columbia has been developing an automated drug delivery system for anesthesia. As part of this effort, an open-loop bolus-based neuromuscular blockade system was developed and clinically validated in [34]. Discrete-time Laguerre-based LTI models of the dynamic response to rocuronium were identified using data collected from more than 80 patients via clinical trials.[4] To obtain regulatory certificates to fully close the loop while employing an infusion-based administration of the drugs, mathematical guarantees of safety of the system are likely to be required. The viability kernel can provide such guarantees. Let us consider the problem of computing the viability kernel for a constrained dynamical system that represents the pharmacological response of a patient under anesthesia subject to therapeutic bounds.

Patient Model and Constraints

Consider the following discrete-time LTI system describing the Laguerre dynamics of a patient [34, 35]:

$$x(t+1) = Ax(t) + Bu(t), \tag{5.26}$$

$$y(t) = Cx(t) \tag{5.27}$$

with time step $t \in \mathbb{Z}^+$, state vector $x(t) \in \mathbb{R}^6$, input (rocuronium infusion rate [mg/kg/min]) $u(t) \in \mathbb{R}$, and output (pseudo-occupancy, a metric related to the patient's plasma concentration of anesthetic e.g. rocuronium) $y(t) \in$

[4] The states of such a model correspond to Laguerre polynomials whose weighted sum attempts to closely rebuild the shape of the actual impulse (bolus) response of the patient; See [122] for more details on the Laguerre modeling framework.

$\mathbb{R}$. The sampling interval is 20 s and the system matrices are:

$$A = \begin{bmatrix} 0.9960 & 0 & 0 & 0 & 0 & 0 \\ 0.0080 & 0.9960 & 0 & 0 & 0 & 0 \\ -0.0080 & 0.0080 & 0.9960 & 0 & 0 & 0 \\ 0.0079 & -0.0080 & 0.0080 & 0.9960 & 0 & 0 \\ -0.0079 & 0.0079 & -0.0080 & 0.0080 & 0.9960 & 0 \\ 0.0079 & -0.0079 & 0.0079 & -0.0080 & 0.0080 & 0.9960 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.0894 & -0.0890 & 0.0886 & -0.0883 & 0.0879 & -0.0876 \end{bmatrix}^T,$$

$$C = \begin{bmatrix} 18.5000 & 8.2300 & 3.5300 & 4.3400 & 3.7000 & 3.0700 \end{bmatrix}.$$

The constraint set (desired clinical effect) is specified in terms of the pseudo-occupancy level and the input is bounded above and below by hard physical constraints:

$$\begin{cases} y(t) \in \mathcal{K}_0 := [0.1, 1], \\ u(t) \in \mathcal{U}_0 := [0, 0.8]. \end{cases} \tag{5.28}$$

Reformulating the Problem

Note that the therapeutic constraint is specified in the output space (as opposed to the state space) and the output signal y should track a reference setpoint that lies within $\mathcal{K}_0$. To perform our desired analysis on this system, we reformulate the problem by projecting the output bounds onto the state space while making the control action regulatory. For brevity we drop the time argument from the state, input, and output notations.

Projection of Bounds onto the State Space Consider the (nonsingular) linear transformation

$$\begin{bmatrix} C \\ \mathbf{0}_{5\times 1} \quad I_5 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_6 \end{bmatrix}^T =: \begin{bmatrix} w_1 & w_2 & \cdots & w_6 \end{bmatrix}^T.$$

The states $w_2, \ldots, w_6$ are the Laguerre states $x_2, \ldots, x_6$. In the new coordinate space, the bounds are state space constraints on the first state

$w_1 := Cx = y.$

Tracking vs. Regulating We perform an affine change of coordinates and shift the equilibrium point to the origin. This is done by augmenting the state vector in the w-space with the reference output signal y^* and applying a basis translation so that in the new coordinates the first state $w_1 := y$ becomes $z_1 := y - y^*$:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ w_2 \\ \vdots \\ w_6 \\ y^* \end{bmatrix} = \begin{bmatrix} y - y^* \\ w_2 \\ \vdots \\ w_6 \\ y^* \end{bmatrix} =: \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_6 \\ z_7 \end{bmatrix}. \tag{5.29}$$

Let $u(t) = u_{ss}$ be the steady state control input needed for tracking a constant setpoint $y^*(t) = y^* = 0.9$. This value can be easily calculated using a standard state-feedback procedure from

$$\begin{bmatrix} x_{ss} \\ u_{ss} \end{bmatrix} = \begin{bmatrix} A - I & B \\ C & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ y^* \end{bmatrix}, \tag{5.30}$$

where x_{ss} denotes the steady state equilibrium. To complete the reformulation, we deduct u_{ss} from the control set of the transformed system.

The new constraints for the transformed, extended system $z(t+1) = \widetilde{A}z(t) + \widetilde{B}u(t)$, $y(t) = \widetilde{C}z(t)$, with $\widetilde{A} \in \mathbb{R}^{7\times 7}$, $\widetilde{B} \in \mathbb{R}^{7\times 1}$, $\widetilde{C} \in \mathbb{R}^{1\times 7}$, are as follows:

$$\begin{cases} z(t) \in \mathcal{K} := (\mathcal{K}_0 - y^*) \times \mathbb{R}^6, \\ u(t) \in \mathcal{U} := \mathcal{U}_0 - u_{ss}. \end{cases} \tag{5.31}$$

In this new seven-dimensional state space the first state z_1 represents the drug pseudo-occupancy minus its setpoint value of 0.9 units, the next five states are the second to sixth Laguerre states transformed from the original coordinates, and the last state z_7 is a constant corresponding to the pseudo-occupancy setpoint. The states are assumed to be constrained

by a slab in $\mathbb{R}^7$ that is only bounded in the z_1 direction. Note that with this formulation, the last state z_7 is allowed to take on values that are not needed; of actual interest is the behavior of the remaining states when z_7 equals the pseudo-occupancy setpoint y^*.

The input constraint set is a one-dimensional ellipsoid. To under-approximate the state constraint with a non-degenerate ellipsoid we use *a priori* knowledge about the typical values of the (Laguerre) states $z_2, \ldots, z_6$ and bound them by an ellipsoid with a large spectral radius of $\lambda_{\max} = 30$ in those directions. (This imposed constraint can be further relaxed if necessary.) Guaranteeing that this ellipsoidal target set $\mathcal{K}$, which is our desired clinical effect, is not violated during the surgery provides a certificate of safety of the closed-loop system. Therefore, for a 30 min surgery for instance, we require $z(t) \in \mathcal{K}\ \forall t \in [0, 90]$ despite bounded input authority. Using appropriately synthesized infusion policies, the states belonging to the viability kernel of $\mathcal{K}$ under the extended system will never leave the desired clinical effect for the duration of the surgery.

We under-approximate $Viab_{[0,90]}(\mathcal{K}, \mathcal{U})$ in 986 s using Algorithm 5.2 with $|\mathcal{M}| = 30$. Of the 30 randomly chosen initial directions used in the ellipsoidal computations, 15 resulted in nonempty ellipsoids that make up the piecewise ellipsoidal under-approximation of the viability kernel (Figure 5.7). Note that no similar computations are currently possible in such high dimensions using Eulerian methods directly.

5.4 Summary and Further Discussions

In this chapter we presented a new connection between the viability kernel (and by duality, the minimal reachable tube) and the maximal reachable sets of possibly nonlinear systems. Owing to this connection, the efficient and scalable Lagrangian techniques that were traditionally developed for the approximation of the maximal reachability constructs can now be used to approximate the viability kernel. Motivated by a high-dimensional problem of guaranteed safety in control of anesthesia, we proposed a scalable algorithm that computes a piecewise ellipsoidal under-approximation of the viability

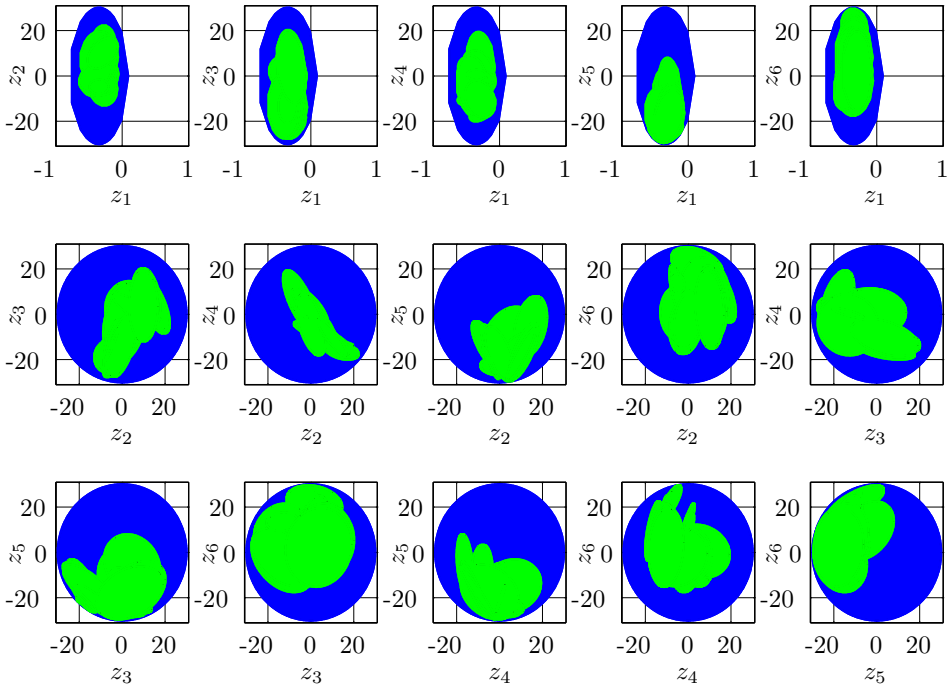


Figure 5.7: 2D projections of the under-approximation of $Viab_{[0,90]}(\mathcal{K}, \mathcal{U})$ for Example 5.3.2 for the first six states when z_7 equals the setpoint value. The constraint set $\mathcal{K}$ (blue/dark) and a piecewise ellipsoidal under-approximation of the provably safe regions (green/light) are shown.

kernel for LTI systems based on ellipsoidal techniques for reachability.

Empirically quantifying the computational complexity of the piecewise ellipsoidal algorithm is a work under way for which we expect a polynomial complexity in the order of $|\mathcal{M}||P|\left(\mathcal{O}(\mathfrak{R}_\delta) + \mathcal{O}(\mathfrak{S})\right)$ where $\mathcal{O}(\mathfrak{R}_\delta)$ is the complexity of computing the maximal reachable set along a given direction over the time interval δ and $\mathcal{O}(\mathfrak{S})$ is the complexity of solving the SDP (5.19). Our preliminary tests for a chain of double integrators using a single direction (yielding one ellipsoid in $\mathbb{R}^n$) show that the technique scales relatively well up to about 35 dimensions, which is computed in less than 200 s.

The presented connection between the viability kernel and the maximal reachable sets paves the way to synthesizing safety-preserving optimal control laws in a more efficient and scalable manner. This scalable synthesis methodology as well as an extension of the results of this chapter to the differential games setting and the approximation of the discriminating kernel are presented in the following chapter.

Chapter 6

Robust Approximation and Scalable Safety-Preserving Control Synthesis

In this chapter we will first extend the results presented in the previous chapter to the differential games setting—the case in which the system is perturbed by an unknown but bounded disturbance input or uncertainty in the model. We will then propose a safety-preserving control strategy based on the piecewise ellipsoidal algorithm discussed in Section 5.2. Although the synthesis of safety-preserving controllers is by no means a new research direction (cf. e.g. [5, 11, 81, 116]), the results presented here provide a *scalable* synthesis of such controllers for a class of constrained systems.

We will make use of the definitions and preliminaries in Sections 2.5 and 2.6. We focus only on the continuous-time case. In the discrete-time case the Isaac's condition does not in general hold and consequently the discussion is more involved.

6.1 The Discriminating Kernel in Terms of Robust Maximal Reachable Sets

For the constrained system (2.29), i.e.

$$\dot{x}(t) = f(x(t), u(t), v(t)), \quad x(0) = x_0 \tag{6.1}$$

with control input $u(t) \in \mathcal{U}$ and disturbance input $v(t) \in \mathcal{V}$, we show that we can approximate $Disc_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V})$ by considering a nested sequence of sets that are robustly reachable in small sub-intervals of $[0, \tau]$.

Similar to before, we say that the vector field f is bounded on $\mathcal{K}$ if there exists a norm $\|\cdot\|\colon \mathcal{X} \to \mathbb{R}^+$ and a real number $M > 0$ such that for all $x \in \mathcal{K}$, $u \in \mathcal{U}$, and $v \in \mathcal{V}$ we have $\|f(x, u, v)\| \leq M$. If $\mathcal{K}$ is compact, every continuous vector field f is bounded on $\mathcal{K}$.

6.1.1 Under-Approximating the Discriminating Kernel

The discriminating kernel can be under-approximated via the recursion

$$K_{|P|}(P) = \mathcal{K}_{\downarrow}(P), \tag{6.2a}$$

$$K_{k-1}(P) = \mathcal{K}_{\downarrow}(P) \cap Reach^{\sharp}_{t_k - t_{k-1}}(K_k(P), \mathcal{U}, \mathcal{V}) \quad \text{for} \quad k \in \{1, \ldots, |P|\} \tag{6.2b}$$

where $\mathcal{K}_{\downarrow}(P) := \{x \in \mathcal{K} \mid \operatorname{dist}(x, \mathcal{K}^c) \geq M\|P\|\}$ is a subset of $\mathcal{K}$ deliberately chosen at a distance of $M\|P\|$ from its boundary.

Proposition 6.1. *For any partition $P \in \mathscr{P}([0, \tau])$ the final set $K_0(P)$ defined by the recurrence relation* (6.2) *satisfies*

$$K_0(P) \subseteq Disc_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V}). \tag{6.3}$$

Proof. The proof is a straightforward extension of the proof of Proposition 5.1: Since f is bounded on $\mathcal{K}$, there exists a norm $\|\cdot\|$ and a real number $M > 0$ such that $\|f(x, u, v)\| \leq M \ \forall (x, u, v) \in \mathcal{K} \times \mathcal{U} \times \mathcal{V}$. Now, fix a partition P of $[0, \tau]$ and take a point $x_0 \in K_0(P)$. By the construction of $K_0(P)$, this means that for each $k = 1, \ldots, |P|$ there is some point $x_k \in K_k(P)$ and a control $u_k\colon [0, t_k - t_{k-1}] \to \mathcal{U}$ for any (non-anticipative) disturbance $v_k = \rho[u_k]\colon [0, t_k - t_{k-1}] \to \mathcal{V}$ such that x_k can be reached from x_{k-1} at time $t_k - t_{k-1}$ using u_k. Thus, taking the concatenation of the inputs u_k and v_k, we get a control input $u\colon [0, \tau] \to \mathcal{U}$ for every disturbance input $v = \rho[u]\colon [0, \tau] \to \mathcal{V}$ such that the solution

$x\colon [0,\tau] \to \mathcal{X}$ to the initial value problem $\dot{x} = f(x,u,v)$, $x(0) = x_0$, satisfies $x(t_k) = x_k \in K_k(P) \subseteq \{x \in \mathcal{K} \mid \operatorname{dist}(x, \mathcal{K}^c) \geq M\|P\|\}$. We claim that this guarantees that $x(t) \in \mathcal{K}\ \forall t \in [0,\tau]$. Indeed, any $t \in [0,\tau)$ lies in some interval $[t_k, t_{k+1})$. Since f is bounded by M, we have

$$\begin{aligned} \|x(t) - x(t_k)\| &\leq \int_{t_k}^{t} \|\dot{x}(s)\| ds \leq M(t - t_k) \\ &< M(t_{k+1} - t_k) \leq M\|P\|. \end{aligned} \tag{6.4}$$

Further, $x(t_k) \in K_k(P)$ implies $\operatorname{dist}(x(t_k), \mathcal{K}^c) \geq M\|P\|$. Combining these, we see that

$$\begin{aligned} \operatorname{dist}(x(t), \mathcal{K}^c) &\geq \operatorname{dist}(x(t_k), \mathcal{K}^c) - \|x(t) - x(t_k)\| \\ &> M\|P\| - M\|P\| = 0 \end{aligned} \tag{6.5}$$

and hence $x(t) \in \mathcal{K}$. Thus, $x_0 \in Disc_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V})$. □

Corollary 6.1 (Intermediate Discriminating Kernels)**.** *For any partition $P \in \mathscr{P}([0,\tau])$ and every $k \in \{1,\ldots,|P|\}$ the sets $K_{k-1}(P)$ defined by the recurrence relation* (6.2) *satisfy*

$$K_{k-1}(P) \subseteq Disc_{[0,\tau - t_{k-1}]}(\mathcal{K}, \mathcal{U}, \mathcal{V}) \tag{6.6}$$

with $K_{|P|}(P) \subseteq Disc_{[0,0]}(\mathcal{K}, \mathcal{U}, \mathcal{V}) = \mathcal{K}$.

Similar to the viability kernel case, the approximation here can be made to be arbitrarily precise by choosing a sufficiently fine partition.

Proposition 6.2. *Suppose that the vector field $f\colon \mathcal{X} \times \mathcal{U} \times \mathcal{V} \to \mathcal{X}$ is bounded on a set $\mathcal{K} \subseteq \mathcal{X}$. Then we have*

$$Disc_{[0,\tau]}(\mathring{\mathcal{K}}, \mathcal{U}, \mathcal{V}) \subseteq \bigcup_{P \in \mathscr{P}([0,\tau])} K_0(P) \subseteq Disc_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V}). \tag{6.7}$$

In particular, when $\mathcal{K}$ is open,

$$\bigcup_{P\in\mathscr{P}([0,\tau])} K_0(P) = Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V}). \tag{6.8}$$

Proof. The proof is a straightforward extension of the proof of Proposition 5.2: The second inclusion in (6.7) follows directly from Proposition 6.1. To prove the first inclusion, take a state $x_0 \in Disc_{[0,\tau]}(\mathring{\mathcal{K}},\mathcal{U},\mathcal{V})$. There exists a control $u\colon [0,\tau] \to \mathcal{U}$ for every disturbance $v = \rho[u]\colon [0,\tau] \to \mathcal{V}$ such that the trajectory $x(\cdot)$ satisfies $x(t) \in \mathring{\mathcal{K}}\ \forall t \in [0,\tau]$. Since $\mathring{\mathcal{K}}$ is open, $\forall x \in \mathring{\mathcal{K}}\ \ \mathrm{dist}(x,\mathcal{K}^c) > 0$. Further, $x\colon [0,\tau] \to \mathcal{X}$ is continuous so the function $t \mapsto \mathrm{dist}(x(t),\mathcal{K}^c)$ is continuous on the compact set $[0,\tau]$. Thus, we can define $d > 0$ to be its minimum value. Take a partition P of $[0,\tau]$ such that $\|P\| < d/M$. We need to show that $x_0 \in K_0(P)$.

First note that our partition P is chosen such that $\mathrm{dist}(x(t),\mathcal{K}^c) > M\|P\|$ $\forall t \in [0,\tau]$. Hence $x(t_k) \in K_{|P|}(P)$ for all $k = 0,\ldots,|P|$. To show that $x(t_{k-1}) \in Reach^{\sharp}_{t_k - t_{k-1}}(K_k(P),\mathcal{U},\mathcal{V})$ for all $k = 1,\ldots,|P|$, consider the tokenizations $\{u_k\}_k$ and $\{v_k\}_k$ of the control u and the disturbance v, respectively, corresponding to P. It is easy to verify that for all k we can reach $x(t_k)$ from $x(t_{k-1})$ at time $t_k - t_{k-1}$ using the control input u_k for every $v_k = \rho[u_k]$. Thus, in particular, we have $x_0 = x(t_0) \in Reach^{\sharp}_{t_1-t_0}(K_1(P),\mathcal{U},\mathcal{V})$. So $x_0 \in K_0(P)$. Hence $Disc_{[0,\tau]}(\mathring{\mathcal{K}},\mathcal{U},\mathcal{V}) \subseteq \bigcup_{P\in\mathscr{P}([0,\tau])} K_0(P)$. □

By approximating $Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V})$ using the sub-interval maximal reachable sets via recursion (6.2) we enable the use of scalable Lagrangian techniques for the computation of the discriminating kernel for high-dimensional systems. As we shall see, this will also allow us to compute the safety-preserving control laws more efficiently.

6.2 Computational Algorithm & Safety-Preserving Control Strategy

The following corollary forms the basis of our safety-preserving feedback strategy. It follows directly from Proposition 6.1, its proof, and the recur-

sion (6.2).

Corollary 6.2. *For a fixed time partition $P \in \mathscr{P}([0,\tau])$ suppose $K_0(P) \neq \emptyset$. For any initial condition $x_0 \in K_0(P)$ the concatenation u of sub-interval control inputs $u_k \colon [0, t_k - t_{k-1}] \to \mathcal{U}$ corresponding to robust maximal reachable sets $Reach^{\sharp}_{t_k - t_{k-1}}(K_k(P), \mathcal{U}, \mathcal{V})$ for $k = 1, \ldots, |P|$ is a safety-preserving control law that keeps the trajectory $x(\cdot)$ of the system (6.1) with $x(0) = x_0$ contained in $\mathcal{K}$ for every $v(t) = \rho[u](t) \in \mathcal{V}$ for all time $t \in [0,\tau]$.*

Notice that via Corollary 6.2 any Lagrangian method that facilitates the synthesis of maximal reachability controllers can be employed to form a safety-preserving policy. One such Lagrangian method is the ellipsoidal techniques [70] implemented in Ellipsoidal Toolbox (ET) [73].

Consider the case in which (2.1) is a linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t) + Gv(t) \tag{6.9}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m_u}$, and $G \in \mathbb{R}^{n\times m_v}$. We will further assume that the constraints $\mathcal{K}$, $\mathcal{U}$, and $\mathcal{V}$ are (or can be closely under-approximated by) nonempty compact ellipsoids.

In Chapter 5 we introduced a scalable piecewise ellipsoidal algorithm (Algorithm 5.1) based on the ellipsoidal techniques for under-approximating the viability kernel $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$. That result can easily be extended so that the generated set approximates the discriminating kernel $Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V})$ instead by simply computing the intermediate maximal reachable sets for adversarial inputs in line with the analysis in Proposition 6.1. Before we describe the corresponding safety-preserving feedback strategy, let us summarize the aforementioned algorithm used to under-approximate the discriminating kernel.

6.2.1 Piecewise Ellipsoidal Approximation of the Discriminating Kernel

Similar to the discussions of Section 5.2.1, the robust maximal reachable set computed using the ellipsoidal techniques for a single direction is an

ellipsoidal subset of the actual robust maximal reachable set. With this in mind, and recalling that the function $\bigcirc(\cdot)$ maps a set to its maximum volume inscribed ellipsoid, we can state the following.

Proposition 6.3. *For a fixed partition $P \in \mathscr{P}([0,\tau])$ and a terminal direction $\ell_\tau \in \mathcal{M}$, the recursion*

$$K_{k-1}^{*[\ell_\tau]} = \bigcirc(K_{|P|}(P) \cap Reach_{t_k - t_{k-1}}^{\sharp[\ell_\tau]}(K_k^{*[\ell_\tau]}(P), \mathcal{U}, \mathcal{V})) \quad \text{for} \quad k \in \{1, \dots, |P|\} \tag{6.10}$$

with $K_{|P|}^{*[\ell_\tau]}(P) = K_{|P|}(P)$ defined as in (6.2a) *generates an ellipsoidal set $K_0^{*[\ell_\tau]}(P)$ such that*

$$\bigcup_{\ell_\tau \in \mathcal{M}} K_0^{*[\ell_\tau]}(P) := K_0^*(P) \subseteq Disc_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V}). \tag{6.11}$$

Proof. The proof is similar to that of Proposition 5.4 which discusses the case in which $\mathcal{V} = \{0\}$. □

The set $K_0^*(P)$ is therefore a piecewise ellipsoidal under-approximation of the discriminating kernel. Notice however that the final $\bigcirc(\cdot)$ operation when $k = 1$ is not necessary if a closed-form piecewise ellipsoidal expression is not needed. Indeed, one can easily verify that

$$\begin{aligned} K_0^*(P) &\subseteq K_{|P|}(P) \cap \bigcup_{\ell_\tau \in \mathcal{M}} Reach_{t_1}^{\sharp[\ell_\tau]}(K_1^{*[\ell_\tau]}(P), \mathcal{U}, \mathcal{V}) \\ &\subseteq Disc_{[0,\tau]}(\mathcal{K}, \mathcal{U}, \mathcal{V}). \end{aligned} \tag{6.12}$$

Finally, note that similar to (6.6) the intermediate discriminating kernels are under-approximated via

$$\bigcup_{\ell_\tau \in \mathcal{M}} K_{k-1}^{*[\ell_\tau]}(P) := K_{k-1}^*(P) \subseteq Disc_{[0,\tau - t_{k-1}]}(\mathcal{K}, \mathcal{U}, \mathcal{V}). \tag{6.13}$$

Remark 6.1. *The under-approximation $\mathcal{K}_{|P|}(P) = \mathcal{K}_\downarrow(P)$ may be constructed by either eroding the set $\mathcal{K}$ by a ball of radius $M\|P\|$ (for a given*

uniform bound M), or alternatively, by using the method described on page 94 with the difference that the right-hand side of (5.23) *is now replaced by*

$$\underset{l\in\mathcal{M}}{\text{minvol}}\left(\bigoplus^{\mathcal{E}_l^+}\{A\mathcal{K},\, B\mathcal{U},\, G\mathcal{V}\}\right). \tag{6.14}$$

6.2.2 Safety-Preserving Feedback Policy

Suppose

$$\mathcal{U} := \mathcal{E}(\mu, U) \tag{6.15}$$

with center $\mu \in \mathbb{R}^{m_u}$ and shape matrix $U \in \mathbb{R}^{m_u\times m_u}$. Based on the piecewise ellipsoidal algorithm described above, we can form a control policy taking values pointwise on $\mathcal{U}$ that keeps the trajectory of the system in $\mathcal{K}$ over the entire time horizon (despite the actions of the disturbance) by concatenating the sub-interval robust maximal reachability control laws according to Corollary 6.2.

Suppose that in computing the reachable set $Reach^{\sharp[\ell_\tau]}_{t_k-t_{k-1}}(K_k^{*[\ell_\tau]}(P),\mathcal{U},\mathcal{V})$ we can also compute

$$Reach^{\sharp[\ell_\tau]}_{t_k-t}(K_k^{*[\ell_\tau]}(P),\mathcal{U},\mathcal{V}) \quad \forall t \in (t_{k-1}, t_k). \tag{6.16}$$

That is, we compute the entire reachable *tube* over the kth sub-interval. For fixed $\ell_\tau \in \mathcal{M}$ and $k \in \{1,\dots,|P|\}$ let $x_k^{c[\ell_\tau]}(t-t_{k-1})$ and $X_{\ell,k}^{-[\ell_\tau]}(t-t_{k-1})$ denote the center and the shape matrix of the ellipsoid $Reach^{\sharp[\ell_\tau]}_{t_k-t}(K_k^{*[\ell_\tau]}(P),\mathcal{U},\mathcal{V})$ at time $t \in [t_{k-1}, t_k]$ with $K_k^{*[\ell_\tau]}(P) = \mathcal{E}(x_k^{c[\ell_\tau]}(t_k - t_{k-1}), X_{\ell,k}^{-[\ell_\tau]}(t_k - t_{k-1}))$ and the terminal direction ℓ_τ. Define a shorthand notation

$$\begin{aligned}\mathcal{R}(t,k) &:= \bigcup_{\ell_\tau\in\mathcal{M}} \mathcal{R}^{[\ell_\tau]}(t,k)\\ &:= \bigcup_{\ell_\tau\in\mathcal{M}} Reach^{\sharp[\ell_\tau]}_{t_k-t}(K_k^{*[\ell_\tau]}(P),\mathcal{U},\mathcal{V})\\ &= \bigcup_{\ell_\tau\in\mathcal{M}} \mathcal{E}(x_k^{c[\ell_\tau]}(t-t_{k-1}), X_{\ell,k}^{-[\ell_\tau]}(t-t_{k-1})),\end{aligned} \tag{6.17}$$

and suppose $\mathcal{R}(0,1) \neq \emptyset$.

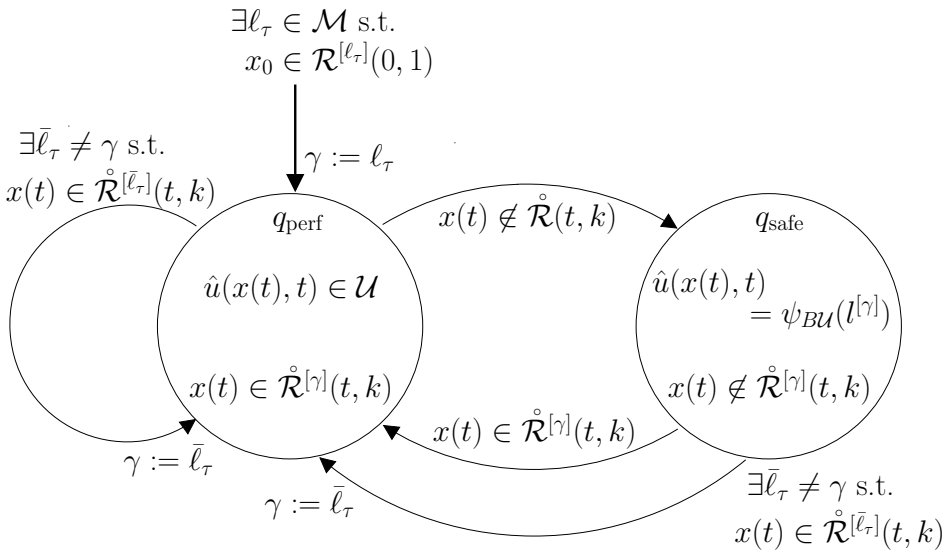


Figure 6.1: The graph of the hybrid automaton H representing the safety-preserving controller.

Now, consider a controller described by the hybrid automaton

$$H = (\mathcal{Q}, \Sigma_e, \Sigma_i, Init, Dom, E, G, R, \mathcal{U}_{\text{fb}}) \tag{6.18}$$

where $\mathcal{Q} = \{q_{\text{perf}}, q_{\text{safe}}\}$ is the set of discrete states with q_{perf} representing the case in which the controller is free to choose any value in $\mathcal{U}$ ("performance" mode) and with q_{safe} representing the case in which the controller is required to return the optimal safety-preserving law ("safety" mode); See Figure 6.1. The inputs to the controller are drawn from the sets $\Sigma_e \subseteq \mathcal{X} \times [0,\tau]$ (*external* input) and $\Sigma_i \subseteq \mathcal{M}$ (*internal* input). The external input is the pair $(x(t),t) \in \Sigma_e$ and the internal input is the direction vector $\gamma \in \Sigma_i$. The initial state of the automaton $Init \subseteq \mathcal{Q} \times \Sigma_e \times \Sigma_i$ is assumed to be $Init = (q_{\text{perf}}, x_0, 0, \{\gamma \mid \exists \ell_\tau \in \mathcal{M},\, x_0 \in \mathcal{R}^{[\ell_\tau]}(0,1),\, \gamma = \ell_\tau\})$. The domains $Dom(\cdot,\gamma,t)\colon \mathcal{Q} \to 2^{\mathcal{X}}$ of the automaton for every $\gamma \in \Sigma_i$ and $t \in [0,\tau]$ are $Dom(q_{\text{perf}},\gamma,t) = \mathring{\mathcal{R}}^{[\gamma]}(t,k)$ and $Dom(q_{\text{safe}},\gamma,t) = (Dom(q_{\text{perf}},\gamma,t))^c$. The domains specify (γ,t)-varying invariants for every $(x(t),t) \in \Sigma_e$ that must be satisfied in each mode. The edges $E \subseteq \mathcal{Q} \times \mathcal{Q}$ are

$E = \{(q_{\text{perf}}, q_{\text{perf}}), (q_{\text{perf}}, q_{\text{safe}}), (q_{\text{safe}}, q_{\text{perf}})_1, (q_{\text{safe}}, q_{\text{perf}})_2\}$ where subscripts are used when necessary in order to distinguish between the two edges of the automaton that enable transitions from q_{safe} to q_{perf} that possess different properties (e.g. guards and reset rules). The guards $G(\cdot, \gamma, t)\colon E \to 2^{\mathcal{X}}$ for every $\gamma \in \Sigma_i$ and $t \in [0, \tau]$ are conditions on $(x(t), t) \in \Sigma_e$ defined as: $G(q_{\text{perf}}, q_{\text{safe}}, \gamma, t) = (\mathring{\mathcal{R}}(t,k))^c$, $G((q_{\text{safe}}, q_{\text{perf}})_1, \gamma, t) = \mathring{\mathcal{R}}^{[\gamma]}(t,k)$, and $G((q_{\text{safe}}, q_{\text{perf}})_2, \gamma, t) = G(q_{\text{perf}}, q_{\text{perf}}, \gamma, t) = \{\mathring{\mathcal{R}}^{[\bar{\ell}_\tau]}(t,k)$ for some $\bar{\ell}_\tau \in \Sigma_i$, $\bar{\ell}_\tau \neq \gamma\}$. The domains and the guards are chosen in terms of the *interior* of the set $\mathcal{R}^{[\gamma]}(t,k)$ to ensure that the automaton H is non-blocking and that transitions over E can take place when necessary. A transition corresponding to an edge is enabled for every t if $x(t)$ satisfies its guard. For example, the automaton can make a transition from q_{safe} to q_{perf} over the edge $(q_{\text{safe}}, q_{\text{perf}})_2 \in E$ for a fixed $(\gamma, x(t), t) \in \Sigma_i \times \Sigma_e$ if $\exists \bar{\ell}_\tau \neq \gamma$ s.t. $x(t) \in \mathcal{R}^{[\bar{\ell}_\tau]}(t,k)$. The map $R\colon E \times \Sigma_i \to \Sigma_i$ resets the internal input via $R((q_{\text{safe}}, q_{\text{perf}})_2, \gamma) = \bar{\ell}_\tau$, $R(q_{\text{perf}}, q_{\text{perf}}, \gamma) = \bar{\ell}_\tau$, and $R((q_{\text{safe}}, q_{\text{perf}})_1, \gamma) = R(q_{\text{perf}}, q_{\text{safe}}, \gamma) = \gamma$. Finally, the output of H is a set-valued map $\mathcal{U}_{\text{fb}}\colon \mathcal{Q} \times \Sigma_i \times \Sigma_e \to 2^{\mathcal{U}}$ given by

$$\begin{cases} \mathcal{U}_{\text{fb}}(q_{\text{perf}}, \gamma, x(t), t) = \mathcal{U}; \\ \mathcal{U}_{\text{fb}}(q_{\text{safe}}, \gamma, x(t), t) = \psi_{B\mathcal{U}}(l^{[\gamma]}), \end{cases} \tag{6.19}$$

where

$$\begin{aligned} &\psi_{B\mathcal{U}}\colon \mathbb{R}^n \to \mathcal{U}, \\ &\qquad l^{[\gamma]} \mapsto \mu - UB^T l^{[\gamma]} \big\langle l^{[\gamma]}, BUB^T l^{[\gamma]} \big\rangle^{-1/2} \end{aligned} \tag{6.20}$$

is chosen so that $B\psi_{B\mathcal{U}}(l^{[\gamma]})$ is the support vector of the set $B\mathcal{U} = \mathcal{E}(B\mu, BUB^T) \subseteq \mathcal{X}$ in the direction $-l^{[\gamma]} \in \mathbb{R}^n$ with

$$\begin{aligned} l^{[\gamma]} &= l^{[\gamma]}(x(t), t) \\ &= (X_{\ell,k}^{-[\gamma]}(t - t_{k-1}))^{-1}(x(t) - x_k^{c[\gamma]}(t - t_{k-1})). \end{aligned} \tag{6.21}$$

(This strategy is based on the optimal control design presented in [69] and

[75].)

Notice that we allow non-determinism in the mode transitions of the hybrid automaton to formulate a non-restrictive policy (in the sense of [116]); $q = q_{\text{safe}}$ only when safety is at stake, and $q = q_{\text{perf}}$ otherwise. As we shall see, the primary objective of the controller, i.e. preserving safety, is achievable regardless of this behavior.

Theorem 6.1 (Safety-Preserving Controller)**.** *For a given partition $P \in \mathscr{P}([0,\tau])$ for any $x_0 \in K_0^*(P)$ where $K_0^*(P)$ is the piecewise ellipsoidal set obtained through* (6.10)–(6.11)*, the feedback policy*

$$u(t) = \hat{u}(x(t), t) \in \mathcal{U}_{\text{fb}} \tag{6.22}$$

generated by the hybrid automaton H keeps the trajectory x of the system (6.9) *with initial condition $x(0) = x_0$ contained in $\mathcal{K}$ for any disturbance $v(t) = \rho[u](t) \in \mathcal{V}$ for all time $t \in [0,\tau]$.*

Proof (Adapted from [69]). Let $k \in \{1,\dots,|P|\}$ be the unique integer such that $t \in [t_{k-1}, t_k) =: \mathbb{T}_k$. We prove that safety is preserved in each mode for any given $(\gamma, x(t), t) \in \mathcal{M} \times \mathcal{X} \times \mathbb{T}_k$ for all $v(t) \in \mathcal{V}$.

First, note that if for every k, $x(t_{k-1}) \in K_{k-1}^*(P)$ (which, as we shall see, is indeed the case) then $x(t_{k-1}) \in \mathcal{R}(t_{k-1}, k)$ since $K_{k-1}^*(P) \subseteq \mathcal{R}(t_{k-1}, k)$. Fix $\gamma = \ell_\tau \in \mathcal{M}$ and k and let $x(t_{k-1}) \in \mathcal{R}^{[\gamma]}(t_{k-1}, k)$. Define a continuously differentiable value function $V_k^{[\gamma]} \colon \mathcal{X} \times \mathbb{T}_k \to \mathbb{R}$ such that

$$V_k^{[\gamma]}(x(t), t) = \operatorname{dist}^2(x(t), \mathcal{R}^{[\gamma]}(t, k)). \tag{6.23}$$

Notice that $V_k^{[\gamma]}(x(t), t) = 0$ for $x(t) \in \mathring{\mathcal{R}}^{[\gamma]}(t, k)$ and $V_k^{[\gamma]}(x(t), t) \geq 0$ for $x(t) \notin \mathring{\mathcal{R}}^{[\gamma]}(t, k)$. We use the convention

$$\mathcal{R}^{[\gamma]}(t_{k-1}, k) \equiv \big\{ x \in \mathcal{X} \mid V_k^{[\gamma]}(x, t_{k-1}) \leq 0 \big\}. \tag{6.24}$$

and assume without loss of generality that the disturbance plays a worst-case game against the control.

Consider the case in which $x(t) \notin \mathring{\mathcal{R}}^{[\gamma]}(t, k)$ and the active mode of

the automaton is q_{safe}. Clearly, we have that $\operatorname{dist}(x(t), \mathcal{R}^{[\gamma]}(t,k)) \geq 0$. As shown in [69], computing the Lie derivative of $V_k^{[\gamma]}(x,t)$ along the system (6.9) yields

$$\begin{aligned}\frac{d}{dt}V_k^{[\gamma]}(x(t),t) &= 2\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) \\ &\quad \times \frac{d}{dt}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) \qquad (6.25)\end{aligned}$$

with

$$\begin{aligned}\frac{d}{dt}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) &= \frac{\partial}{\partial t}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) \\ &\quad + \Big\langle \frac{\partial}{\partial x}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)),\, \dot{x}(t)\Big\rangle \qquad (6.26) \\ &= \big\langle l^{[\gamma]}, Bu(t)+Gv(t)\big\rangle + \rho_{B\mathcal{U}}(-l^{[\gamma]}) \\ &\quad - \big\langle l^{[\gamma]}, Gv(t)\big\rangle \qquad (6.27)\end{aligned}$$

where $\rho_{B\mathcal{U}}(l) := \sup_{z\in B\mathcal{U}}\langle l, z\rangle$ is the support function of the set $B\mathcal{U}$ in the direction $l \in \mathbb{R}^n$ and $l^{[\gamma]}$ is the unique direction vector given by (6.21) [75]. We refer the reader to [69, Sections 1.4 and 1.8] for additional details on how (6.27) is obtained from (6.26). Notice that the right-hand side simplifies to

$$\begin{aligned}&\big\langle l^{[\gamma]}, Bu(t)\big\rangle + \rho_{B\mathcal{U}}(-l^{[\gamma]}) \\ &\qquad = \big\langle l^{[\gamma]}, Bu(t)\big\rangle + \sup_{z\in B\mathcal{U}}\big\langle -l^{[\gamma]}, z\big\rangle \\ &\qquad = \big\langle l^{[\gamma]}, Bu(t)\big\rangle - \inf_{z\in B\mathcal{U}}\big\langle l^{[\gamma]}, z\big\rangle. \qquad (6.28)\end{aligned}$$

Recalling the fact that the set $B\mathcal{U}$ is a compact ellipsoid and invoking (6.20), one can verify that

$$\frac{d}{dt}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) = 0 \qquad \text{for } u(t) = \psi_{B\mathcal{U}}(l^{[\gamma]}); \qquad (6.29)$$

$$\frac{d}{dt}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) \geq 0 \qquad \text{for any } u(t) \in \mathcal{U}. \qquad (6.30)$$

When the disturbance does not play a worst-case game, (6.27) is replaced by $\frac{d}{dt}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) \leq \langle l^{[\gamma]}, Bu(t)+Gv(t)\rangle + \rho_{B\mathcal{U}}(-l^{[\gamma]}) - \rho_{G\mathcal{V}}(l^{[\gamma]})$

which would imply $\frac{d}{dt}\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) \leq 0$ for $u(t) = \psi_{B\mathcal{U}}(l^{[\gamma]})$ and $\forall v(t) = \rho[u](t) \in \mathcal{V}$.

On the other hand, when $x(t) \in \mathring{\mathcal{R}}^{[\gamma]}(t,k)$ and the active mode of the controller is q_{perf} we have that $\operatorname{dist}(x(t),\mathcal{R}^{[\gamma]}(t,k)) = 0$. Thus the derivative of the distance function need not be examined as the Lie derivative of the value function (6.25) is automatically zero regardless of the value of $u(t)$.

Combining the above we see that with $u(t) = \hat{u}(x(t),t) \in \mathcal{U}_{\text{fb}}(q,\gamma,x(t),t)$ we have

$$\frac{d}{dt}V_k^{[\gamma]}(x,t)\Big|_{(6.9)} \leq 0 \quad \forall (q,\gamma,x,t,v) \in \mathcal{Q}\times\mathcal{M}\times\mathcal{X}\times\mathbb{T}_k\times\mathcal{V}. \tag{6.31}$$

This yields

$$\begin{aligned}\int_{t_{k-1}}^{t} \frac{d}{ds}V_k^{[\gamma]}(x(s),s)ds \\ = V_k^{[\gamma]}(x(t),t) - V_k^{[\gamma]}(x(t_{k-1}),t_{k-1}) \leq 0\end{aligned} \tag{6.32}$$

which in turn implies (via (6.24))

$$V_k^{[\gamma]}(x(t),t) \leq V_k^{[\gamma]}(x(t_{k-1}),t_{k-1}) \leq 0 \quad \forall t \in \mathbb{T}_k. \tag{6.33}$$

Therefore, for every solution x of the differential inclusion $\dot{x}(t) \in Ax(t) \oplus B\mathcal{U}_{\text{fb}}(q,\gamma,x(t),t) \oplus G\mathcal{V}$ we will have

$$x(t) \in \mathcal{R}^{[\gamma]}(t,k) = Reach_{t_k-t}^{\sharp[\gamma]}(K_k^{*[\gamma]}(P),\mathcal{U},\mathcal{V}) \quad \forall t \in \mathbb{T}_k, \tag{6.34}$$

$$x(t_k) \in K_k^{*[\gamma]}(P). \tag{6.35}$$

Notice that according to (6.13)

$$\begin{aligned}K_k^{*[\gamma]}(P) \subseteq \bigcup_{\gamma\in\mathcal{M}} K_k^{*[\gamma]}(P) := K_k^*(P) \\ \subseteq Disc_{[0,\tau-t_k]}(\mathcal{K},\mathcal{U},\mathcal{V}) \subseteq \mathcal{K}.\end{aligned} \tag{6.36}$$

Therefore, $x(t_k) \in \mathcal{K}$, for every $k \in \{1,\ldots,|P|\}$ granted that the feedback control law (6.22) is applied.

It remains to show that $x(t) \in \mathcal{K}$ for all $t \in (t_{k-1}, t_k)$. Indeed, by construction of the recursion (6.10) via Proposition 6.1 we have that for all $t \in \mathbb{T}_k$ and $k \in \{1, \ldots, |P|\}$

$$Reach^{\sharp[\gamma]}_{t_k - t}(K_k^{*[\gamma]}(P), \mathcal{U}, \mathcal{V}) \subseteq \{x \in \mathcal{K} \mid \mathrm{dist}(x, \mathcal{K}^c) \geq M\|P\|\}. \tag{6.37}$$

Therefore combining all this we conclude that with $u(t) \in \mathcal{U}_{\mathrm{fb}}$ and $x_0 \in \mathcal{R}(0,1)$ we have $x_{x_0}^{u,\rho[u]}(t) \in \mathcal{K}\ \forall t \in [0, \tau]$ for every $v(t) = \rho[u](t) \in \mathcal{V}$. □

6.2.3 Remarks and Practical Considerations

Respecting the Intermediate Kernels

The strategy (6.22) will ensure that the trajectories of the closed-loop system evolve in the interior or else on the boundary of $\mathcal{R}(t,k)$ for all $t \in [0,\tau]$ and $k \in \{1, \ldots, |P|\}$. Thus as we have seen in (6.35) and (6.36), for a given $P \in \mathscr{P}([0,\tau])$ even if the disturbance always plays a worst-case game the trajectories satisfy $x(t_k) \in K_k^*(P) \subseteq Disc_{[0,\tau - t_k]}(\mathcal{K}, \mathcal{U}, \mathcal{V})$ for all k.

Shared Boundary Points

In practice for common points that are on the boundaries of two or more ellipsoids $Reach^{\sharp[\ell_\tau]}_{t_k - t}(K_k^{*[\ell_\tau]}(P), \mathcal{U}, \mathcal{V})$ at every time instant t, any one of these ellipsoids can be used for the computation of $l^{[\ell_\tau]}(x(t), t)$ in (6.21) and its corresponding optimal control law $\psi_{B\mathcal{U}}(l^{[\ell_\tau]}(x(t), t))$ in mode q_{safe}. Any such control will ensure that the trajectory remains within the corresponding robust maximal reachable tube while being steered towards $K_k^{*[\ell_\tau]}(P)$ (which, as we have seen, will ultimately ensure constraint satisfaction and safety).

To allow a more permissive control policy, the automaton H could be modified by adding a self-loop in mode q_{safe} (i.e. an edge $(q_{\mathrm{safe}}, q_{\mathrm{safe}})$) so that among all possible safe control laws corresponding to those ellipsoids that share the common boundary point, one that simultaneously better satisfies a given performance criterion is selected.

Conservatism of $K_0^*(P)$: Synthesis for $x_0 \notin K_0^*(P)$

When the initial condition x_0 lies inside of the true discriminating kernel but outside of its piecewise ellipsoidal under-approximation, i.e. $x_0 \in Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V})\backslash K_0^*(P)$ for a given $P \in \mathscr{P}([0,\tau])$, the control policy (6.22) will not in general guarantee that the constraint $\mathcal{K}$ is respected for all $t \in [0,\tau]$. However, with the following modification the feedback law will attempt to keep the trajectory within $\mathcal{K}$: For every $t \in [0,s]$, $s \in (0,\tau]$, with s being the first instant such that $x(s) \in \mathcal{R}(s,k')$ for some $k' \in \{1,\ldots,|P|\}$, for every $k \leq k'$ the direction vector $l^{[\ell_\tau]}(x(t),t)$ in (6.21) is modified to be the direction that corresponds to any $\ell_\tau \in \mathcal{M}$ for which $\mathrm{dist}(x(t), Reach_{t_k-t}^{\sharp[\ell_\tau]}(K_k^{*[\ell_\tau]}(P),\mathcal{U},\mathcal{V})) = \min$. We will demonstrate this using an example in Section 6.3.

Finally, note that for $x_0 \notin Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V})$ the modified control law described above *may* still be able to keep x in $\mathcal{K}$ over $[0,\tau]$, only if the disturbance does not play optimally against the control. (In formulation of $Disc_{[0,\tau]}(\mathcal{K},\mathcal{U},\mathcal{V})$ we always assume that the disturbance plays its worst-case game.) On the other hand, in the deterministic case ($\mathcal{V} = \{0\}$), for $x_0 \notin Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ there does not exist a control law that can keep the trajectory within $\mathcal{K}$ over $[0,\tau]$. Therefore any safety-preserving strategy ultimately fails.

Smooth Transition Between Safety and Performance

When the mode q_{perf} of the controller is active, the strategy $\mathcal{U}_{\text{fb}}(q_{\text{perf}},\gamma,x(t),t) = \mathcal{U}$ for some $(\gamma,x(t),t) \in \Sigma_i \times \Sigma_e$ is returned and any performance-satisfying control input in $\mathcal{U}$ can be chosen and applied to the system; denote this input as $u_{\text{perf}}(t)$. On the other hand, when the mode q_{safe} is active, the controller returns $\mathcal{U}_{\text{fb}}(q_{\text{safe}},\gamma,x(t),t) = \psi_{B\mathcal{U}}(l^{[\gamma]})$ and only this specific safety-preserving control law must be applied to the system to ensure safety; we denote this input as $u_{\text{safe}}(t)$.

Choosing u_{perf} arbitrarily without considering the main objective of the closed-loop system (preserving safety) may result in excessive switching between the two modes of the controller (whose main priority is to preserve

safety). Thus the resulting control policy could have a high switching frequency and the controller could end up spending a significant amount of time at the extremum points of the input constraint set. Such a policy, among other shortcomings, may be hard on actuators in a practical setting.

An ideal control policy in mode q_{perf} should be a combination of both u_{perf} and u_{safe} (even though technically the safety component u_{safe} is not needed when the controller is in mode q_{perf}). One solution is a simple convex combination of these two inputs. That is, for every $(\gamma, x(t), t) \in \Sigma_i \times \Sigma_e$ and a given domain $\mathring{\mathcal{R}}^{[\gamma]}(t,k) = \mathring{\mathcal{E}}(x_k^{c[\gamma]}(t-t_{k-1}), X_{\ell,k}^{-[\gamma]}(t-t_{k-1}))$ in q_{perf} we shall choose an input such that

$$u(t) = (1 - \beta_\alpha[\phi^{[\gamma]}](x(t),t))u_{\text{perf}}(t) + \beta_\alpha[\phi^{[\gamma]}](x(t),t)u_{\text{safe}}(t), \tag{6.38}$$

with

$$\beta_\alpha[\phi^{[\gamma]}](x(t),t) := \begin{cases} 1 & \text{if} \quad \phi^{[\gamma]}(x(t),t) \geq 1; \\ \frac{1}{1-\alpha}(\phi^{[\gamma]}(x(t),t) - \alpha) & \text{if} \quad \alpha \leq \phi^{[\gamma]}(x(t),t) \leq 1; \\ 0 & \text{if} \quad \phi^{[\gamma]}(x(t),t) \leq \alpha, \end{cases} \tag{6.39}$$

where $\alpha \in [0,1)$ is a design parameter (Figure 6.2) and

$$\phi^{[\gamma]} \colon \Sigma_e \to \mathbb{R}^+, \tag{6.40}$$

$$\begin{aligned}(x(t),t) \mapsto \big\langle &(x(t) - x_k^{c[\gamma]}(t-t_{k-1})), \\ &(X_{\ell,k}^{-[\gamma]}(t-t_{k-1}))^{-1}(x(t) - x_k^{c[\gamma]}(t-t_{k-1}))\big\rangle.\end{aligned} \tag{6.41}$$

Note that in q_{perf} for fixed γ and t we have that $\text{dom}(\phi^{[\gamma]}(\cdot,t)) = \mathring{\mathcal{R}}^{[\gamma]}(t,k)$. Therefore, $\text{range}(\phi^{[\gamma]}(\cdot,t)) = [0,1)$. This is true simply because the set $\mathcal{R}^{[\gamma]}(t,k)$ is an ellipsoid, and therefore by definition, the one sub-level sets of the function $\phi^{[\gamma]}(\cdot,t)$ in q_{perf} form the interior of $\mathcal{R}^{[\gamma]}(t,k)$. That is, $\mathring{\mathcal{R}}^{[\gamma]}(t,k) := \{x \in \mathcal{X} \mid \phi^{[\gamma]}(x,t) < 1\}$.

Notice that $\phi^{[\gamma]}(x(t),t)$ determines how "deep" inside the domain of q_{perf} the trajectory is at time t by evaluating the one sub-level sets of the ellipsoid

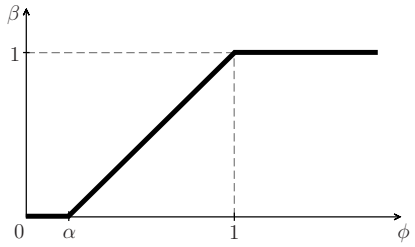


Figure 6.2: β_α as a function of ϕ for a given design parameter $\alpha \in [0, 1)$.

$\mathcal{R}^{[\gamma]}(t, k)$ (the closure of the domain of q_{perf}) at $x(t)$. The closer the trajectory is to the boundary of $\mathcal{R}^{[\gamma]}(t, k)$ the greater the value of $\beta_\alpha[\phi^{[\gamma]}](x(t), t)$ and therefore the more pronounced the safety component $u_{\text{safe}}(t)$ of the input will be. On the other hand, if the trajectory is deep inside the domain of q_{perf}, $\phi^{[\gamma]}(x(t), t)$ tends to α, and therefore $\beta_\alpha[\phi^{[\gamma]}](x(t), t)$ goes to zero. As a result, a greater emphasis is given to the performance component $u_{\text{perf}}(t)$ of the input. As such, the parameter α determines where within the domain of q_{perf} the safety component $u_{\text{safe}}(t)$ should kick in. Clearly, larger values of α imply more emphasis on performance, while smaller values yield smoother transition between q_{perf} and q_{safe}. Note that the limit of the control law $u(t)$ in (6.38) as $x(t) \to \partial\mathcal{R}^{[\gamma]}(t, k)$ is $u_{\text{safe}}(t)$. Therefore the control is continuous across the automaton's transition to q_{safe}.

Such a policy will ensure a gradual change of the effective component of the control law from one form to the other, resulting in less switching frequency between performance and safety. We will show this policy using a number of examples in Section 6.3.

Intermediate Maximal Reachable Sets vs. Tubes

Notice that while in the approximation of the discriminating/viability kernel via recursion (6.10) only the final intermediate maximal reachable sets $Reach^{\sharp[\ell_\tau]}_{t_k - t_{k-1}}(K^{*[\ell_\tau]}_k(P), \mathcal{U}, \mathcal{V})$ are used, in the control synthesis scheme described above (which is based on the ellipsoidal techniques [69]) to form a

proper state-feedback law all intermediate reachable *tubes*

$$\bigcup_{t\in\mathbb{T}_k} Reach_{t_k-t}^{\sharp[\ell_\tau]}(K_k^{*[\ell_\tau]}(P),\mathcal{U},\mathcal{V}) \tag{6.42}$$

are required and must be recorded.

Chattering and Zeno Executions

In theory, the continuous-time evolution of the dynamics in each mode of H and the particular formulation of the automaton may cause chattering (frequent switches between the two modes) or Zeno behavior (infinite switches in finite time). While such undesirable phenomena may be avoided by imposing certain conditions such as dwell-time [92] (requiring that the automaton remains in each mode for a non-zero amount of time) or other techniques as discussed in e.g. [2, 29, 53], in practice for a machine implementation of our control algorithm a finite number of sets can be used to form the intermediate maximal reachable tubes due to the discrete nature of numerical evaluations of the integration of the differential equations. This, in principle, imposes a constant dwell-time (which equals to the integration time-step) and thus ensures that the automaton is well-behaved. In addition, the previously presented scheme for smooth transition between safety and performance (page 119) can also prevent chattering to a great extent; See Section 6.3.1.

6.3 Numerical Examples

We start with a number of toy examples to demonstrate the results and then proceed with more realistic and practical examples.

6.3.1 2D System Without Uncertainty

Consider the system

$$\dot{x} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u$$

with $u(t) \in \mathcal{U} := [-0.15, 0.15]$ and a constraint set $\mathcal{K}$ defined to be a disc of radius 0.5 about the origin. We approximate the viability kernel $Viab_{[0,2]}(\mathcal{K},\mathcal{U})$ (using $|\mathcal{M}| = 2$ directions and uniform partition with $|P| = 80$) and synthesize safety-preserving control laws that ensure $x(t) \in \mathcal{K}$ $\forall t \in [0, 2]$. Since u in q_{perf} can be chosen arbitrarily in $\mathcal{U}$, we simply apply $u = 0$. The simulation results for a few initial conditions are given in Figures 6.3 and 6.4.

Smooth Transition Between Safety and Performance

Here we will apply the mildly varying control law described by (6.38)–(6.41) (with $\alpha = 0$) to obtain less switching frequency for the controller. Figure 6.5 shows the closed-loop trajectories generated by such a policy, while Figure 6.6 shows the corresponding feedback strategies.

6.3.2 2D System With Uncertainty

Consider the system

$$\dot{x} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v$$

subject to unknown but bounded disturbance $v(t) \in \mathcal{V} := [-0.1, 0.1]$, bounded control input $u(t) \in \mathcal{U} := [-0.15, 0.15]$, and state constraint $x(t) \in \mathcal{K} := \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 0.5\}$ for all $t \in [0, 2]$. We approximate the (finite-horizon) discriminating kernel $Disc_{[0,2]}(\mathcal{K},\mathcal{U},\mathcal{V})$ (using $|\mathcal{M}| = 4$ directions and uniform partition with $|P| = 80$) and synthesize safety-preserving control laws that ensure $x(t) \in \mathcal{K}$ $\forall t \in [0, 2]$ despite the action of the unknown disturbance. As before, since u in q_{perf} can be chosen arbitrarily in $\mathcal{U}$, we simply apply $u = 0$. The simulation results for a few initial conditions are given in Figures 6.7, 6.8, and 6.9, where we have assumed that for each initial condition the disturbance is a random signal with uniform distribution on $\mathcal{V}$, i.e. $v(t) \in U([-0.1, 0.1])$. Safety is preserved despite this (unknown) disturbance.

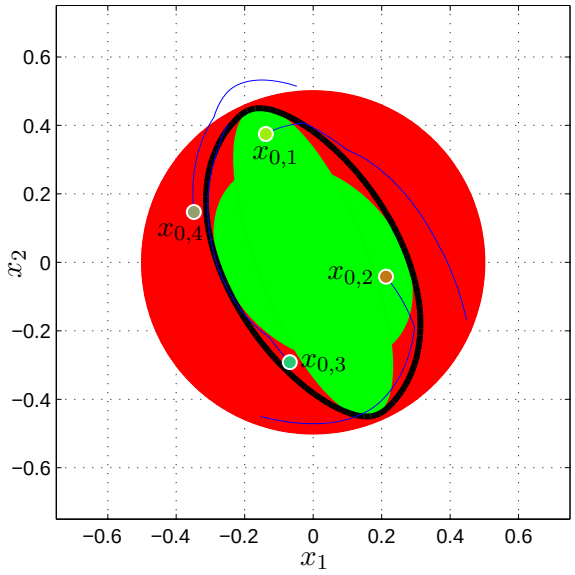


Figure 6.3: Closed-loop trajectories in the phase-plane for four sample initial conditions $x_{0,i}$ for Example 6.3.1. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal under-approximation of the viability kernel (green/light) are shown. The level set approximation of the viability kernel is outlined in thick black lines. The feedback law is chosen such that $u = 0$ when $q = q_{\text{perf}}$ and $u = u_{\text{safe}}$ when $q = q_{\text{safe}}$. Safety is preserved over the finite horizon for $x_{0,i} \in Viab_{[0,2]}(\mathcal{K},\mathcal{U})$. While $x_{0,3}$ is not in the under-approximation of the viability kernel, the designed control policy still maintains safety (see page 119). Note that the trajectories can leave the kernel since it is finite-horizon and hence not invariant; however, they do not violate the constraints. For $x_{0,4} \notin Viab_{[0,2]}(\mathcal{K},\mathcal{U})$, even though the control is always at its extremum points (see the bottom plot in Figure 6.4), the trajectory eventually leaves $\mathcal{K}$.

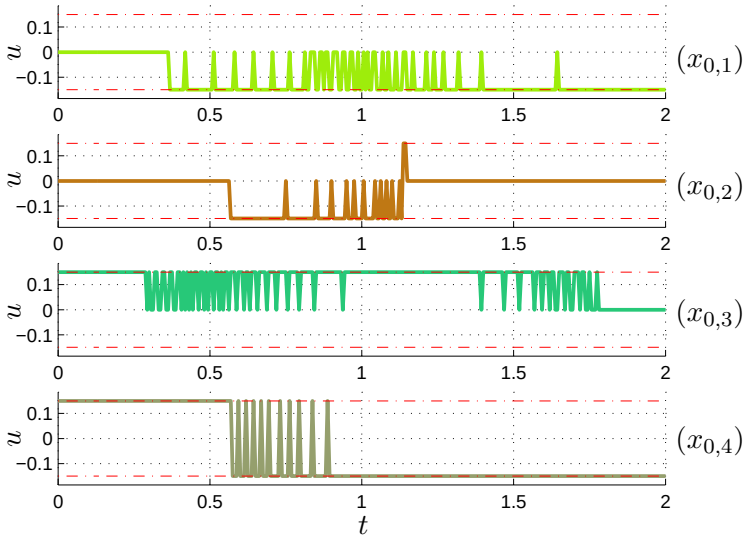


Figure 6.4: The corresponding safety-preserving feedback policy for Example 6.3.1 for each initial condition $x_{0,i}$ (in Figure 6.3). The dashed lines (red) indicate the hard bounds on the input.

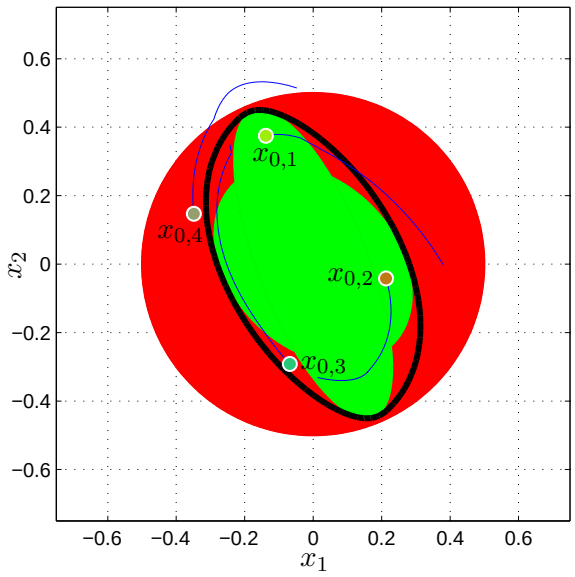


Figure 6.5: Closed-loop trajectories with smooth transition (via (6.38)–(6.41) with $\alpha = 0$) between q_{perf} and q_{safe} for Example 6.3.1. Safety is maintained.

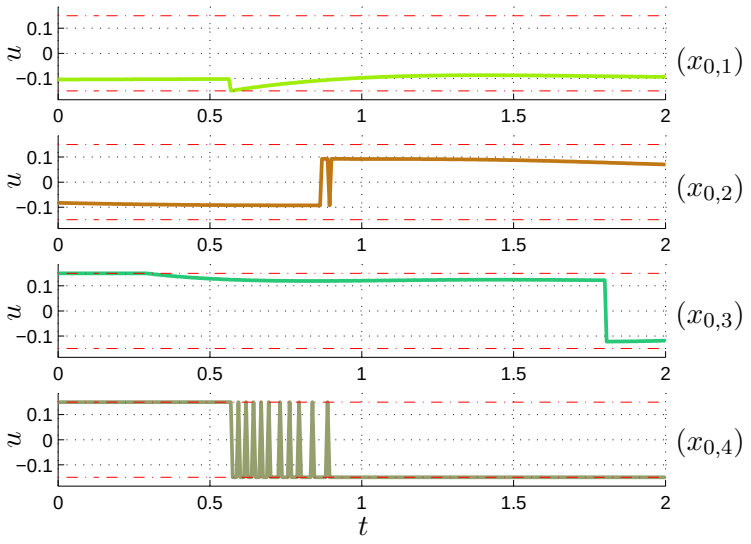


Figure 6.6: The corresponding safety-preserving feedback policy with less switching frequency for Example 6.3.1 for $x_{0,i}$, $i = 1, \ldots, 3$ (in Figure 6.5). Note that for $x_{0,4}$ the controller is never in q_{perf}.

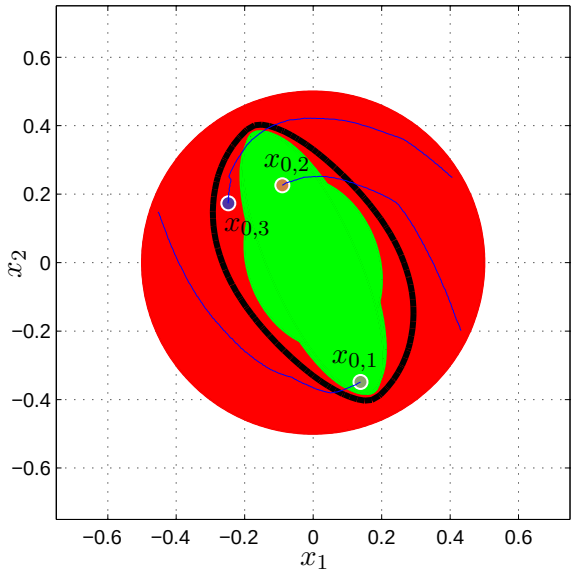


Figure 6.7: Closed-loop trajectories in the phase-plane for three sample initial conditions $x_{0,i}$ for Example 6.3.2. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal under-approximation of the discriminating kernel (green/light) are shown. The level set approximation of the kernel is outlined in thick black lines. The feedback law is chosen such that $u = 0$ when $q = q_{\text{perf}}$ and $u = u_{\text{safe}}$ when $q = q_{\text{safe}}$. Safety is preserved despite the disturbance.

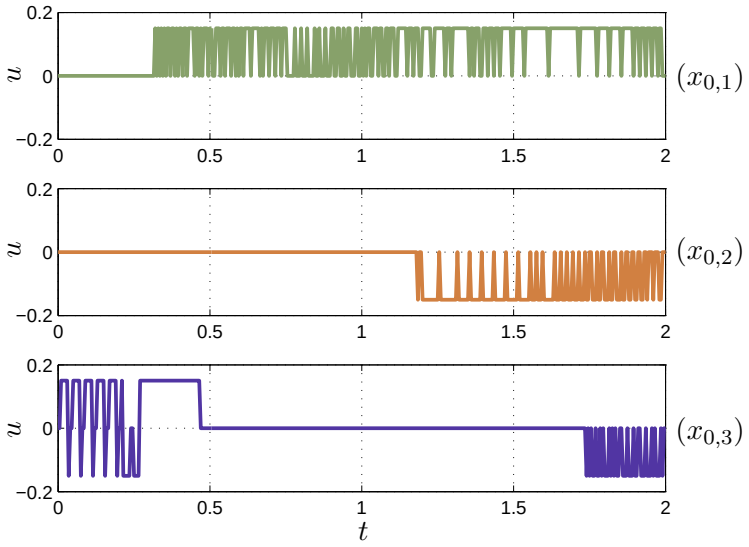


Figure 6.8: The corresponding safety-preserving feedback policy for Example 6.3.2 for each initial condition $x_{0,i}$.

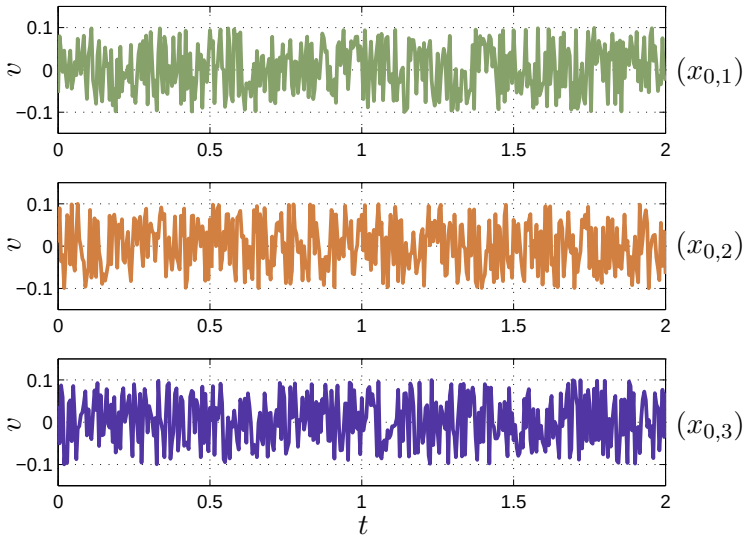


Figure 6.9: The simulated disturbance signal for Example 6.3.2 for each initial condition $x_{0,i}$.

k_{10}	k_{12}	k_{13}	k_{21}	k_{31}	V_1
0.4436	0.1140	0.0419	0.0550	0.0033	16.044

Table 6.1: Model parameters for the patient (11 years old, 35 kg); cf. [1]

6.3.3 3D Control of Anesthesia

Consider the problem of safety in control of anesthesia as described in Section 5.3.2. Instead of the discrete-time Laguerre model of the patient, here we use a three-dimensional continuous-time pharmacokinetic/pharmacodynamic (PKPD) compartmental model whose state variables describe the concentration of propofol in each compartment of the body:

$$\dot{x} = \begin{bmatrix} -(k_{10}+k_{12}+k_{13}) & k_{12} & k_{13} \\ k_{21} & -k_{21} & 0 \\ k_{31} & 0 & -k_{31} \end{bmatrix} x + \begin{bmatrix} 1/V_1 \\ 0 \\ 0 \end{bmatrix} u.$$

The model parameters, taken from [1], are given in Table 6.1. Suppose that the patient is to undergo a 30 min surgery and that the input (propofol infusion rate [mg/min]) is hard bounded above and below such that $u(t) \in \mathcal{U} := [0, 20]$. To keep the plasma drug concentration within therapeutic range, we require

$$x(t) \in \mathcal{K} := \mathcal{E}\left(\begin{bmatrix} 3.5 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 6.25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 25 \end{bmatrix} \right) \quad \forall t \in [0, 30]. \tag{6.43}$$

We approximate the viability kernel $Viab_{[0,30]}(\mathcal{K}, \mathcal{U})$ via Algorithm 5.1 using $|\mathcal{M}| = 15$ directions and uniform partition with $|P| = 600$. Consider the initial condition $x_0 = \begin{bmatrix} 4 & 5 & 8 \end{bmatrix}^T$ and assume that no drug is being administered, i.e. $u = \min(\mathcal{U}) = 0$. When the safety-preserving controller is not engaged, the therapeutic constraint $\mathcal{K}$ is eventually violated (Figure 6.10) putting the patient at risk of intra- and post-operative complications. When the safety-preserving controller is engaged (using the same input $u = 0$ in q_{perf} and enforcing the optimal safety control law in q_{safe}) the constraint $\mathcal{K}$ is satisfied for the entire duration of surgery (Figure 6.11). To ensure

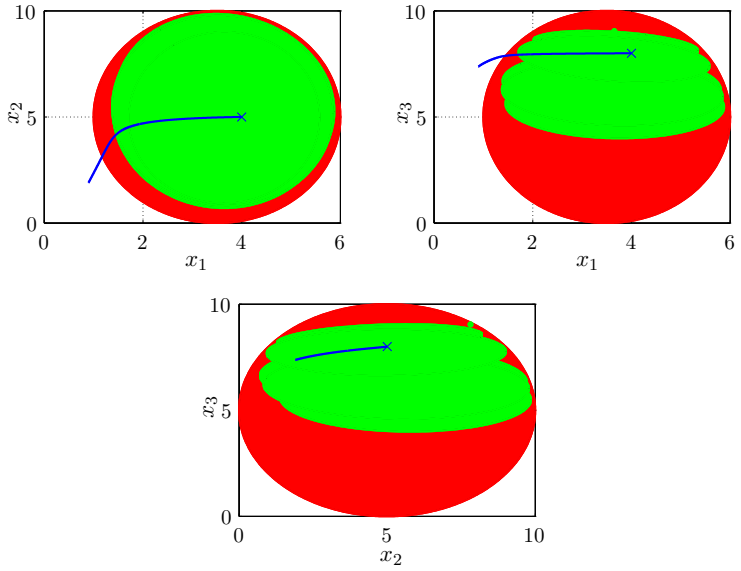


Figure 6.10: 2D projections of the closed-loop trajectories with $u = 0$ and no safety control for Example 6.3.3. The initial condition x_0 is marked by '×'. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal under-approximation of the viability kernel (green/light) are also shown. The constraint set $\mathcal{K}$ (safety of the patient) is violated.

a slowly varying safety-preserving infusion policy the modified strategy in (6.38)–(6.41) (with $\alpha = 0$) is employed: Figure 6.12 shows this feedback policy, while Figures 6.13–6.16 discuss various behaviors and characteristics of the hybrid controller using such a strategy. In comparison, a direct application of the safety-preserving controller without the modified strategy in (6.38)–(6.41), while still capable of maintaining safety, results in significant chattering and an aggressive infusion policy as shown in Figures 6.17–6.21.

6.3.4 6D Flight Envelope Protection

Consider the problem of aerodynamic flight envelope protection for NASA's Highly Maneuverable Aircraft Technology (HiMAT) [105]. The longitudinal dynamics of the HiMAT aircraft trimmed at 25 kft and 0.9 Mach are given

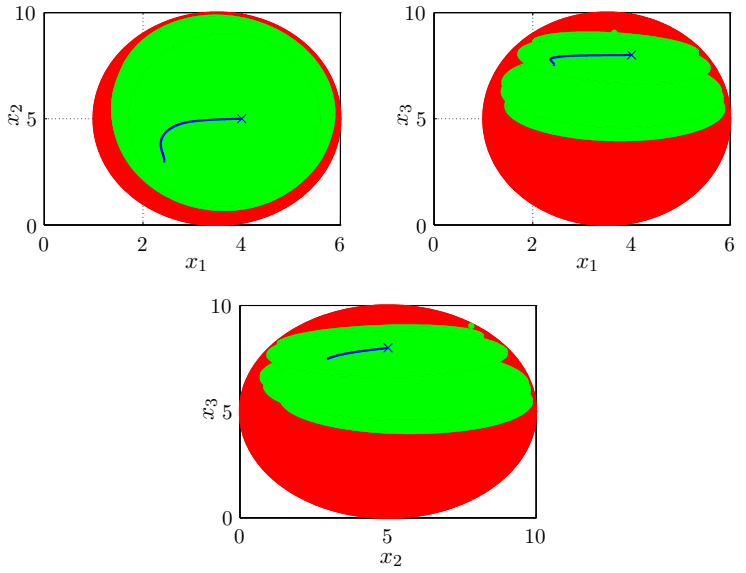


Figure 6.11: 2D projections of the closed-loop trajectories with $u = u_{\text{safe}}$ in q_{safe} and $u_{\text{perf}} = 0$ in q_{perf} using the modified policy (6.38)–(6.41) with $\alpha = 0$ for Example 6.3.3. The initial condition x_0 is marked by '$\times$'. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal under-approximation of the viability kernel (green/light) are also shown. Safety is preserved.

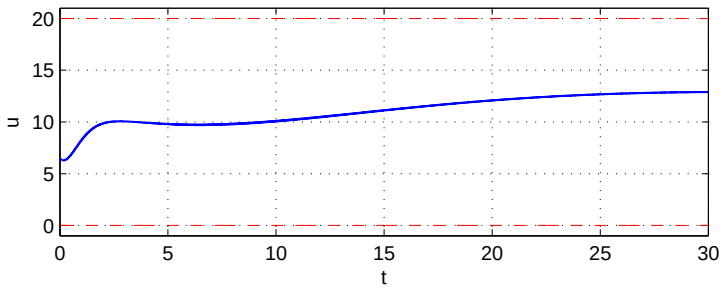


Figure 6.12: The safety-preserving feedback policy via (6.38)–(6.41) with $\alpha = 0$ for Figure 6.11 for Example 6.3.3. The bounds on the input are shown as dashed lines.

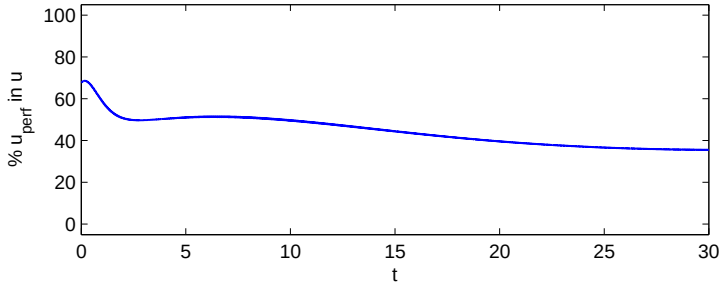


Figure 6.13: $u_{\text{perf}}(t)$ as a percentage of $u(t)$ in q_{perf} for the modified policy (6.38)–(6.41) with $\alpha = 0$ for Example 6.3.3. The safety component of the input becomes more dominant as the trajectory gets closer to the boundary of the ellipsoid that is the domain of q_{perf} as shown in Figure 6.16. A value of 0% would correspond to when the automaton has switched to q_{safe}, while a value of 100% would correspond to when $x(t)$ is at the center of the domain of q_{perf}. Increasing α would further emphasize the u_{perf} component, at the cost of a more aggressive control policy and possibly more frequent switchings between the two modes of the automaton.

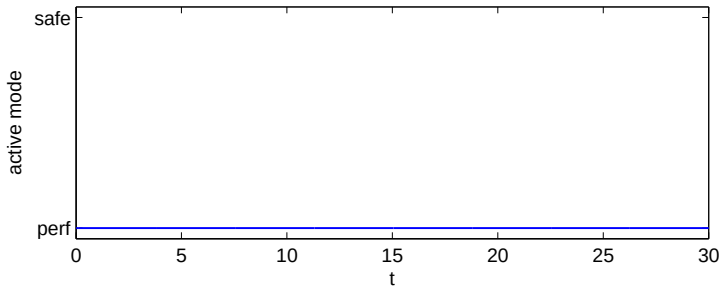


Figure 6.14: The active mode of the hybrid automaton H at time t using the modified policy (6.38)–(6.41) with $\alpha = 0$ for Example 6.3.3. The safety component u_{safe} of the input in q_{perf} seems to be sufficient to maintain safety in this case, and thus the controller does not switch to q_{safe}.

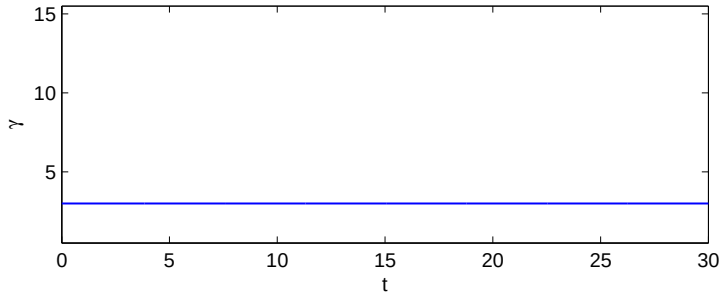


Figure 6.15: The index of the ellipsoid being used by the automaton (i.e. the internal input $\gamma \in \Sigma_i$ of H) at time t when using the modified policy (6.38)–(6.41) with $\alpha = 0$ for Example 6.3.3. The ellipsoid does not change.

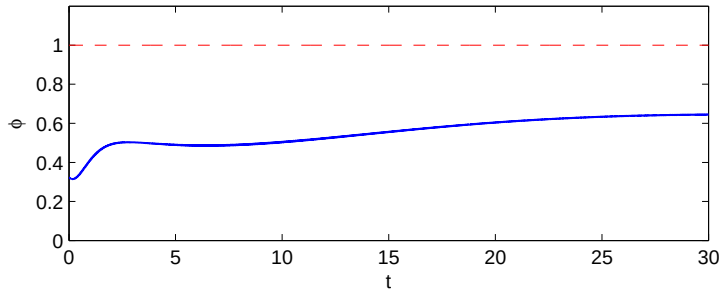


Figure 6.16: The location of $x(t)$ within the domain of q_{perf}, i.e. $\phi^{[\gamma]}(x(t), t)$, when using the modified policy (6.38)–(6.41) with $\alpha = 0$ for Example 6.3.3. The boundary of the corresponding ellipsoid is marked by the dashed line at 1. The center of the ellipsoid is at 0.

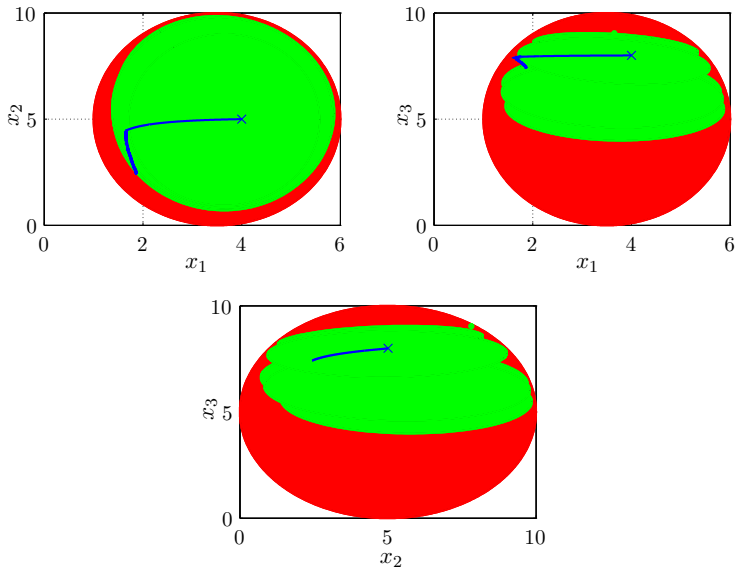


Figure 6.17: 2D projections of the closed-loop trajectories with $u = u_{\text{perf}} = 0$ in q_{perf} and $u = u_{\text{safe}}$ in q_{safe} for Example 6.3.3. While safety is still preserved, the corresponding feedback policy chatters significantly (Figure 6.18) as compared to the slowly varying policy in Figure 6.12.

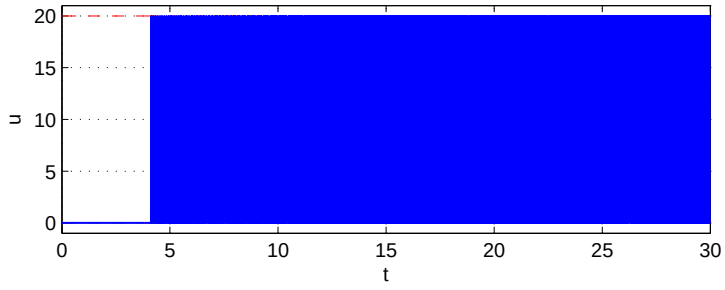


Figure 6.18: The safety-preserving feedback policy for Figure 6.17 for Example 6.3.3. The bounds on the input are shown as dashed lines. The controller chatters between q_{perf} and q_{safe} as shown in Figure 6.19.

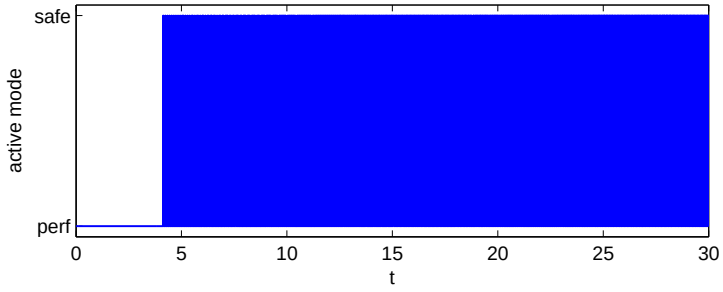


Figure 6.19: The active mode of the hybrid automaton H at time t for Example 6.3.3 when the modified policy (6.38)–(6.41) is *not* employed.

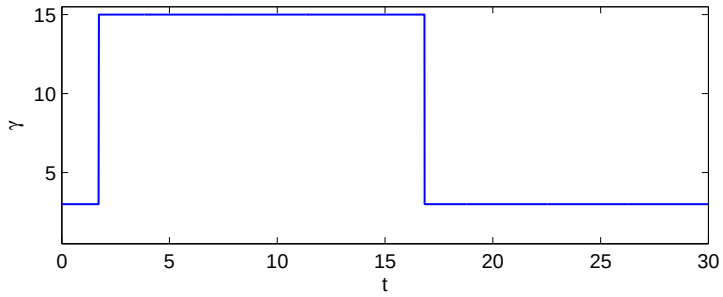


Figure 6.20: The index of the ellipsoid being used by the automaton (i.e. the internal input $\gamma \in \Sigma_i$ of H) at time t for Example 6.3.3. Note that the ability to switch ellipsoids in H is not the cause of the chatter (as evidenced by the fact that the ellipsoid appears to only switch twice).

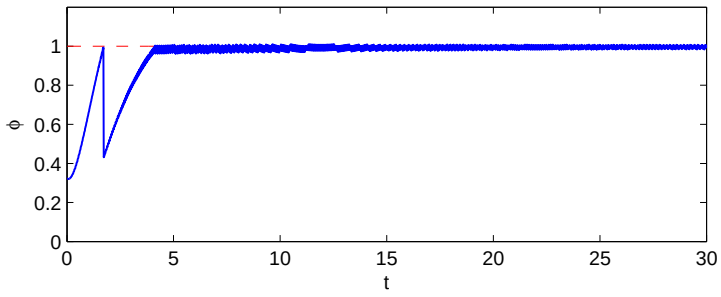


Figure 6.21: The location of $x(t)$ within the domains of q_{perf} and q_{safe}, i.e. $\phi^{[\gamma]}(x(t),t)$, for Example 6.3.3. For $\phi^{[\gamma]}(x(t),t) < 1$ the active mode is q_{perf} (whose domain is an ellipsoid whose index is shown in Figure 6.20). On the other hand, for $\phi^{[\gamma]}(x(t),t) \geq 1$ the active mode is q_{safe}. At time $t = 1.71$ the automaton takes a transition on $(q_{\text{perf}}, q_{\text{perf}}) \in E$ while resetting γ to correspond to a new direction vector (ellipsoid). This ensures that the active mode remains q_{perf}. The change of ellipsoid causes a change in the definition of ϕ and hence a discontinuous jump in the plot. At $t = 4.1$ the automaton is forced to switch to q_{safe} to maintain safety, but progresses by chattering between q_{safe} and q_{perf} as indicated by the values of $\phi^{[\gamma]}(x(t),t)$.

by $\dot{x} = Ax + Bu$ with

$$A = \begin{bmatrix} -0.0226 & -36.6170 & -18.8970 & -32.0900 & 3.2509 & -0.7626 \\ 0.0001 & -1.8997 & 0.9831 & -0.0007 & -0.1708 & -0.0050 \\ 0.0123 & 11.7200 & -2.6316 & 0.0009 & -31.6040 & 22.3960 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -30.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & -30.0000 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 0 & 0 & 30 \end{bmatrix}^T$$

and state vector $x = \begin{bmatrix} \dot{\alpha} & \alpha & \dot{\theta} & \theta & x_{\delta_e} & x_{\delta_c} \end{bmatrix}^T$ in which the first four states represent angle of attack and attitude angle and their rates of change, and the last two states represent elevon and canard control actuator dynamics. The control input $u = \begin{bmatrix} \delta_e & \delta_c \end{bmatrix}^T$ is comprised of elevon and canard actuators. For all $t \in [0, 1]$ we assume that $u(t) \in \mathcal{U}$, a disc of radius 0.5236 rad ($\approx 30^\circ$) about the origin in $\mathbb{R}^2$, and require $x(t) \in \mathcal{K} := \{x \in \mathbb{R}^6 \mid \|x\|_2 \leq 5\}$.[1]

Suppose that a pre-designed Linear Quadratic Regulator (LQR) controller is to be used that satisfies the performance functional

$$J(u) = \int_0^\infty (x^T Q x + u^T R u) dt \tag{6.44}$$

with $Q = I_6$ and $R = 10^2 \times I_2$ subject to the system dynamics. The corresponding state-feedback control law that minimizes $J(u)$ is $u_{\text{lqr}} := -Kx$ with

$$K = \begin{bmatrix} 0.0823 & -0.8209 & -0.3457 & -0.5247 & 0.3119 & -0.2107 \\ -0.0552 & 0.5619 & 0.2367 & 0.3588 & -0.2107 & 0.1497 \end{bmatrix}.$$

Since the control authority is constrained by the ellipsoid $\mathcal{U} =: \mathcal{E}(\omega, \Omega)$,

[1] These assumptions are purely for the algorithmic convenience of dealing with ellipsoids. The input constraint set, for instance, may be better described by a rectangle in $\mathbb{R}^2$ as the actuators are in fact dynamically independent, and the state constraint set is entirely fictitious.

applying the above LQR control law may result in saturation such that $u = \mathrm{sat}(u_{\mathrm{lqr}})$ is effectively applied to the system. Here the saturation function $\mathrm{sat}\colon \mathbb{R}^2 \to \mathcal{U}$ is determined by the support vector of the set $\mathcal{U}$ in the direction of $u_{\mathrm{lqr}}/\|u_{\mathrm{lqr}}\|$ and is defined as

$$\mathrm{sat}(u_{\mathrm{lqr}}) := \begin{cases} u_{\mathrm{lqr}} & \text{if} \quad u_{\mathrm{lqr}} \in \mathcal{U}; \\ \omega + \Omega u_{\mathrm{lqr}} \big\langle u_{\mathrm{lqr}},\, \Omega u_{\mathrm{lqr}} \big\rangle^{-1/2} & \text{if} \quad u_{\mathrm{lqr}} \notin \mathcal{U}. \end{cases} \tag{6.45}$$

We approximate the viability kernel $Viab_{[0,1]}(\mathcal{K},\mathcal{U})$ via Algorithm 5.1 using $|\mathcal{M}| = 15$ directions and uniform partition with $|P| = 300$. Consider the initial condition

$$x_0 = \begin{bmatrix} -1.7064 & 1.7769 & -1.8770 & -1.1272 & 1.5994 & 1.7680 \end{bmatrix}^T.$$

Applying the LQR controller without a safety-preserving strategy in place results in violation of the aerodynamic envelope $\mathcal{K}$ as shown in Figure 6.22 due to unaccounted actuator saturations. On the other hand, when the safety-preserving controller is employed (with $u = \mathrm{sat}(u_{\mathrm{lqr}})$ in mode q_{perf} and the optimal safety control law in q_{safe}) the flight envelope is protected over the horizon $[0, 1]$ despite the unaccounted actuator saturations in the LQR action (Figure 6.23).

6.4 Summary and Further Discussions

In this chapter we first extended the results of Chapter 5 to the differential games setting in which the system is subject to unknown but bounded disturbance/uncertainty. Consequently, the discriminating kernel is expressed in terms of robust maximal reachable sets. Owing to this new connection, scalable and efficient Lagrangian methods can now be used to correctly approximate the discriminating kernel. We also presented an extension of the piecewise ellipsoidal algorithm from Section 5.2.1 that facilitates the under-approximation of the discriminating kernel for high-dimensional LTI systems. Based on this algorithm (and its underlying ellipsoidal tech-

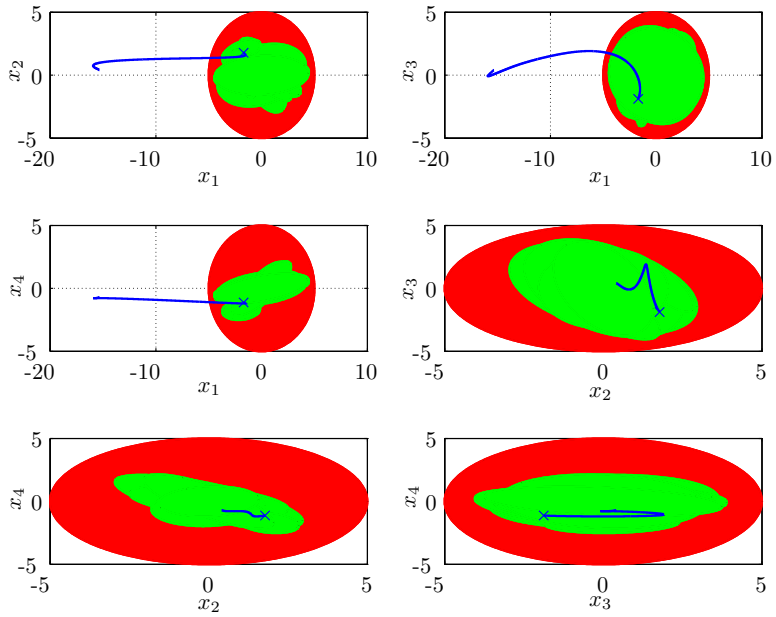


Figure 6.22: 2D projections of the closed-loop trajectories with saturated LQR and no safety control for the first four states for Example 6.3.4. The initial condition x_0 is marked by '$\times$'. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal underapproximation of the viability kernel (green/light) are also shown. The constraint set $\mathcal{K}$ is violated due to unaccounted actuator saturations.

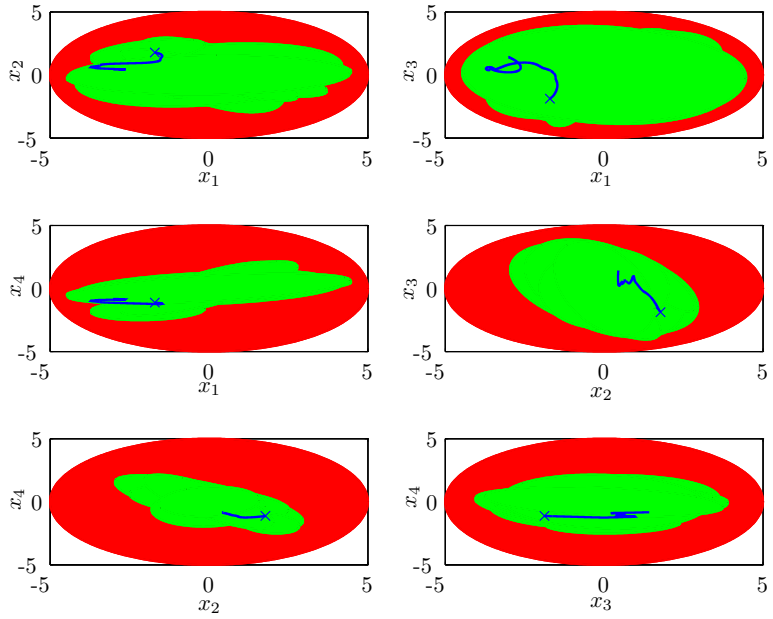


Figure 6.23: 2D projections of the closed-loop trajectories with safety-preserving controller for the first four states for Example 6.3.4. The initial condition x_0 is marked by '$\times$'. The feedback law is chosen such that $u = \mathrm{sat}(u_{\mathrm{lqr}})$ when $q = q_{\mathrm{perf}}$ and $u = u_{\mathrm{safe}}$ when $q = q_{\mathrm{safe}}$. The constraint set $\mathcal{K}$ (red/dark) and a piecewise ellipsoidal under-approximation of the viability kernel (green/light) are shown. Safety is preserved in spite of the actuator saturations in the LQR controller.

niques for reachability [72]), we then proposed a scalable, non-restrictive, safety-preserving, hybrid state-feedback control strategy for continuous-time LTI systems that ensures that the state constraint is not violated despite bounded control authority and, if present, unknown disturbances. We demonstrated the results on several examples including a realistic problem of safety in anesthesia and a 6D problem of aerodynamic flight envelope protection.

Chapter 7

Conclusions and Future Work

We considered the problem of guaranteed safety and constraint satisfaction in high-dimensional, safety-critical, controlled dynamic systems. Reachability analysis and viability theory provide an appropriate framework for set-valued analysis of constrained dynamical systems. To guarantee safety of such systems and to synthesize controllers that are capable of preserving this safety despite bounded control authority (and possibly disturbances or uncertainties), the computation of the minimal reachable tube or by duality, the viability kernel is required. Historically, the algorithms that approximate these sets—known as Eulerian methods—are based on gridding the state space. While powerful and versatile, their computational complexity increases exponentially with the dimension of the state. This renders them impractical for systems of dimensions higher than three or four.

We presented two separate approaches for reduction of complexity in computing the minimal reachable tube or the viability kernel for higher-dimensional systems. The first approach, based on structure decomposition, aims to facilitate the use of Eulerian methods on higher-dimensional, continuous-time, continuously-valued LTI systems (and by extension, hybrid systems with LTI continuous dynamics). It does so by constructing an appropriate similarity transformation that not only results in decoupling (or weak unidirectional coupling) of the dynamics, but also yields disjoint input. This imposed structure is then exploited for decomposition of the system for the purpose of computing the minimal reachable tube and the viability kernel in lower dimensions. A number of algorithms are then presented that

enable a sound approximation of these constructs in lower-dimensional subspaces. It is shown that the reverse transformation of the intersection of the back-projection of these resulting sets correctly approximates the actual constructs. Within the framework of structure decomposition, we proposed two techniques: the Schur-based decomposition and the Riccati-based decomposition, each with its own merits. While the Schur-based decomposition is quite generic and applicable to most systems, the Riccati-based decomposition may yield less conservative reachability computations for two-time-scale or ill-conditioned systems as was shown with an example in Chapter 4.

The second complexity reduction approach, based on set-theoretic methods, draws a connection between the viability kernel and the maximal reachable sets for continuous- and discrete-time systems. Since the maximal reachable sets can be computed using efficient and scalable techniques—known as Lagrangian methods—that employ compact set representations and follow the flow of the dynamics, the viability kernel can now be underapproximated with polynomial complexity. Using the well-established ellipsoidal techniques for maximal reachability we then proposed a scalable algorithm that facilitates the computation of the viability kernel for high-dimensional LTI systems. This approach also enabled us to formulate a scalable safety-preserving static feedback control strategy. We also provided extensions of this approach to systems with unknown but bounded disturbances or uncertainties.

We demonstrated our techniques on several examples (up to 8D) including a problem of safety in control of anesthesia and flight envelope protection for longitudinal aircraft dynamics. We mention, however, that the second approach is much more scalable: The algorithm can likely be applied to systems with several dozens of state variables.

7.1 Summary of Contributions

We summarize our contributions as follows:

- We proposed two structure decomposition techniques that enable the

computation of the minimal reachable tube and the viability kernel for higher-dimensional LTI systems using Eulerian methods. The decomposition techniques that are available in the literature, while generally applicable to nonlinear systems, assume a certain structure on the system that can be exploited. Our techniques, on the other hand, are designed to *impose* this structure.

- We presented a novel connection between the viability kernel and maximal reachable sets. Owing to this connection efficient and scalable Lagrangian methods can now be used to approximate the viability kernel for high-dimensional systems. Our piecewise ellipsoidal algorithm which was proposed based on this new connection using the ellipsoidal techniques for maximal reachability is capable of computing a guaranteed under-approximation of the viability (discriminating) kernel for high-dimensional (uncertain) LTI systems.

- We showed that the presented connection can also be employed to synthesize safety-preserving control laws. We then proposed a scalable safety-preserving control strategy (again based on the ellipsoidal techniques for maximal reachability and the corresponding optimal control laws) that ensures safety of high-dimensional safety-critical, possibly uncertain LTI systems.

7.2 Future Research Directions

There are several possibilities for future work and further developments.

Structure Decomposition

For both of our decomposition techniques—i.e. Schur-based decomposition (Chapters 3) and Riccati-based decomposition (Chapter 4)—future work includes efforts to reduce potential conservatism in the over-approximation of the minimal reachable tube (or the under-approximation of the viability kernel). One direction is to incorporate the geometric information about the

shape of the target set into the decomposition process so that the projection of the set onto the subspaces of the transformed coordinates does not result in excessive loss of detail. A second direction is an alternative transformation that produces subsystems that may both be manipulated to some degree while still preserving the disjoint property of the input.

Set-Theoretic Methods

While the presented piecewise ellipsoidal algorithm in Chapter 5 has proven to be effective and efficient, it may be subject to excessive conservatism particularly for large time horizons. Quantifying the accuracy of the algorithm is an important future research direction. We are also currently developing alternative approaches that yield a more accurate under-approximation of the viability kernel while still preserving the scalability property. In fact our hope is that the presented connection between the viability kernel and maximal reachable sets encourages the development of scalable and accurate algorithms for the computation of the viability kernel for nonlinear systems. Finally, as our algorithm is already heavily based on intersections, a natural (and straightforward) extension would be to consider hybrid dynamical systems.

Safety-Preserving Control Synthesis

Any Lagrangian technique that can accommodate the synthesis of maximal reachability control laws (and can handle adversarial inputs for the discriminating kernel case) may be used to formulate a safety-preserving controller based on the framework described in Chapter 6. Our proposed control algorithm in that chapter is at early stages of development. Many future research directions are possible:

- We have assumed that safety is only to be preserved over the given finite interval $[0, \tau]$—beyond this point the trajectories may leave the constraint set regardless of what the control input does. We are currently developing a variation of the hybrid controller (6.18) that em-

ploys a pseudo-time variable with varying rates of change and can potentially ensure safety over a horizon larger than τ.

- Suppose that the infinite-horizon discriminating kernel is nonempty, and $Disc_{[0,\tau-t_k]}(\mathcal{K},\mathcal{U},\mathcal{V}) \equiv Disc_{\mathbb{R}^+}(\mathcal{K},\mathcal{U},\mathcal{V})$ for some $k \in \{0,\ldots,|P|\}$. Due to propagation of the approximation error, the set generated by the piecewise ellipsoidal algorithm is not in general guaranteed to converge to the infinite-horizon kernel. However, if the algorithm does converge for at least one terminal direction, then the current formulation of the hybrid controller can be used to synthesize infinite-horizon safety-preserving control laws. We intend to prove this point in the future.

- Another avenue is the extension of the safety-preserving controller to the discrete-time case. While this is certainly possible, the discussions become more involved in the presence of disturbance/uncertainty as the Isaac's condition no longer holds.

- The presented safety-preserving controller is non-restrictive/permissive. That is, the optimal safety control law must be applied only when safety is at stake. Otherwise, any desired/performance-satisfying control law in $\mathcal{U}$ can be chosen. A future research direction would be to determine the "distance" of this controller (in an ordered space of all safety-preserving controllers) to the *least restrictive* controller as defined by [81].

Bibliography

[1] A. Absalom and G. Kenny. Paedfusor pharmacokinetic data set. *British Journal of Anaesthesia*, 95(1):110, 2005.

[2] E. Asarin, O. Bournez, T. Dang, O. Maler, and A. Pnueli. Effective synthesis of switching controllers for linear systems. *Proceedings of the IEEE*, 88(7):1011–1025, 2000.

[3] E. Asarin and T. Dang. Abstraction by projection and application to multi-affine systems. In *Hybrid Systems: Computation and Control, LNCS 2993*, pages 129–132. Springer-Verlag, 2004.

[4] E. Asarin, T. Dang, G. Frehse, A. Girard, C. Le Guernic, and O. Maler. Recent progress in continuous and hybrid reachability analysis. In *Proc. IEEE International Symposium on Computer-Aided Control Systems Design*, Munich, Germany, October 2006.

[5] J.-P. Aubin. *Viability Theory*. Systems and Control: Foundations and Applications. Birkhäuser, Boston, MA, 1991.

[6] J.-P. Aubin. Viability kernels and capture basins of sets under differential inclusions. In *Proc. IEEE Conference on Decision and Control*, pages 4605–4610, Las Vegas, NV, Dec 2002.

[7] J.-P. Aubin, A. M. Bayen, and P. Saint-Pierre. *Viability Theory: New Directions*. Springer Verlag, 2nd edition, 2011.

[8] A. M. Bayen, I. M. Mitchell, M. Oishi, and C. J. Tomlin. Aircraft autolander safety analysis through optimal control-based reach set computation. *Journal of Guidance, Control, and Dynamics*, 30(1):68–77, 2007.

[9] C. Béné, L. Doyen, and D. Gabay. A viability analysis for a bioeconomic model. *Ecological Economics*, 36(3):385–396, 2001.

[10] S. Bibian, C. Ries, M. Huzmeman, and G. A. Dumont. Introduction to automated drug delivery in clinical anesthesia. *European Journal of Control*, 11(6):535–557, Dec 2005.

[11] F. Blanchini. Set invariance in control. *Automatica*, 35(11):1747–1767, 1999.

[12] F. Blanchini and S. Miani. *Set-Theoretic Methods in Control*. Springer, 2008.

[13] F. Borrelli, C. Del Vecchio, and A. Parisio. Robust invariant sets for constrained storage systems. *Automatica*, 45(12):2930–2936, 2009.

[14] G. Bouligand. *Introducion a la geometrie inxnitesimale directe*. Paris, Bourgin: Gauthiers-Villars, 1932.

[15] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[16] A. E. Bryson. *Control of Spacecraft and Aircraft*. Princeton Univ. Press, 1994.

[17] J. Camara, A. Girard, and G. Gossler. Safety controller synthesis for switched systems using multi-scale symbolic models. In *Proc. Conference on Decision and Control and European Control Conference*, pages 520–525, Orlando, FL, 2011.

[18] P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre. Set-valued numerical analysis for optimal control and differential games. In M. Bardi, T. Raghavan, and T. Parthasarathy, editors, *Stochastic and Differential Games: Theory and Numerical Methods*, number 4 in Annals of the International Society of Dynamic Games, pages 177–247, Boston, MA, 1999. Birkhäuser.

[19] K. W. Chang. Singular perturbations of a general boundary value problem. *SIAM Journal on Mathematical Analysis*, 3:520–526, 1972.

[20] A. Chutinan and B. H. Krogh. Computing polyhedral approximations to flow pipes for dynamic systems. In *Proc. IEEE Conference on Decision and Control*, pages 2089–2094, Tampa, FL, December 1998.

[21] P.-A. Coquelin, S. Martin, and R. Munos. A dynamic programming approach to viability problems. In *Proc. IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning (ADPRL 2007)*, pages 178–184, 2007.

[22] T. Dang and O. Maler. Reachability analysis via face lifting. In *Hybrid Systems: Computation and Control, LNCS 1386*, pages 96–109. Springer, 1998.

[23] A. N. Daryin, A. B. Kurzhanski, and I. V. Vostrikov. Reachability approaches and ellipsoidal techniques for closed-loop control of oscillating systems under uncertainty. In *Proc. IEEE Conference on Decision and Control*, pages 6385–6390, San Diego, CA, 2006.

[24] G. Deffuant, L. Chapel, and S. Martin. Approximating viability kernels with support vector machines. *IEEE Transactions on Automatic Control*, 52(5):933–937, 2007.

[25] D. Del Vecchio, M. Malisoff, and R. Verma. A separation principle for a class of hybrid automata on a partial order. In *Proc. American Control Conference*, pages 3638–3643, 2009.

[26] G. Dumont, A. Martinez, and J. Ansermino. Robust control of depth of anesthesia. *International Journal of Adaptive Control and Signal Processing*, 23:435–454, 2009.

[27] L. Evans and P. Souganidis. Differential games and representation formulas for solutions of Hamilton-Jaconbi-Isaacs equations. *Indiana University Mathematics Journal*, 33(5):773–797, 1984.

[28] F. Fadaie and M. E. Broucke. On the least restrictive control for collision avoidance of two unicycles. *International Journal of Robust and Nonlinear Control*, 16(12):553–574, 2006.

[29] A. F. Filippov and F. M. Arscott. *Differential Equations With Discontinuous Righthand Sides*. Mathematics and Its Applications. Kluwer Academic Publishers, 1988.

[30] G. Frehse, C. Le Guernic, A. Donz, S. Cotton, R. Ray, O. Lebeltel, R. Ripado, A. Girard, T. Dang, and O. Maler. SpaceEx: Scalable verification of hybrid systems. In G. Gopalakrishnan and S. Qadeer, editors, *Proc. 23rd International Conference on Computer Aided Verification (CAV)*, pages 1–16. Springer, 2011.

[31] G. Freiling. A survey of nonsymmetric Riccati equations. *Linear Algebra and its Applications*, 351:243–270, 2002.

[32] Z. Gajic and I. Borno. General transformation for block diagonalization of weakly coupled linear systems composed of N-subsystems. *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, 47(6):909–912, 2000.

[33] Y. Gao, J. Lygeros, and M. Quincampoix. The reachability problem for uncertain hybrid systems revisited: a viability theory perspective. In J. Hespanha and A. Tiwari, editors, *Hybrid Systems: Computation and Control, LNCS 3927*, pages 242–256, Berlin Heidelberg, 2006. Springer-Verlag.

[34] T. Gilhuly. *Modeling and control of neuromuscular blockade*. PhD thesis, University of British Columbia, Vancouver, Canada, 2007.

[35] T. Gilhuly, G. Dumont, and B. Macleod. Modelling for computer controlled neuromuscular blockade. In *IEEE Engineering in Medicine and Biology Magazine*, pages 26–29, 2005.

[36] A. Girard. Reachability of uncertain linear systems using zonotopes. In M. Morari, L. Thiele, and F. Rossi, editors, *Hybrid Systems: Computation and Control, LNCS 3414*, pages 291–305. Springer, 2005.

[37] A. Girard. Synthesis using approximately bisimilar abstractions: state-feedback controllers for safety specifications. In *Hybrid Systems: Computation and Control*, Stockholm, Sweden, 2010.

[38] A. Girard. Controller synthesis for safety and reachability via approximate bisimulation. *Automatica*, 48(5):947–953, 2012.

[39] A. Girard and C. Le Guernic. Efficient reachability analysis for linear systems using support functions. In *IFAC World Congress*, Seoul, Korea, July 2008.

[40] A. Girard, C. Le Guernic, and O. Maler. Efficient computation of reachable sets of linear time-invariant systems with inputs. In J. Hespanha and A. Tiwari, editors, *Hybrid Systems: Computation and Control, LNCS 3927*, pages 257–271. Springer-Verlag, 2006.

[41] A. Girard and G. J. Pappas. Approximate bisimulation relations for constrained linear systems. *Automatica*, 43(8):1307–1317, 2007.

[42] A. Girard and G. J. Pappas. Approximation metrics for discrete and continuous systems. *IEEE Transactions on Automatic Control*, 52(5):782–798, 2007.

[43] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, 1996.

[44] M. R. Greenstreet. Verifying safety properties of differential equations. In *Proc. Conference on Computer Aided Verification*, pages 277–287, New Brunswick, NJ, July 1996.

[45] J. Groß. Explicit solutions to the matrix inverse problem $AX = B$. *Linear Algebra and its Applications*, 289:131–134, 1999.

[46] M. R. Hafner, D. Cunningham, L. Caminiti, and D. Del Vecchio. Automated vehicle-to-vehicle collision avoidance at intersections. In *Proc. of World Congress on Intelligent Transport Systems*, Orlando, FL, October 2011.

[47] Z. Han and B. H. Krogh. Reachability analysis of hybrid control systems using reduced-order models. In *Proc. American Control Conference*, pages 1183–1189, Boston, MA, June 2004.

[48] Z. Han and B. H. Krogh. Reachability analysis for affine systems using ϵ-decomposition. In *Proc. IEEE Conference on Decision and Control, and European Control Conference*, pages 6984–6990, Seville, Spain, December 2005.

[49] Z. Han and B. H. Krogh. Reachability analysis of large-scale affine systems using low-dimensional polytopes. In J Hespanha and A. Tiwari, editors, *Hybrid Systems: Computation and Control, LNCS 3927*, pages 287–301, Berlin, Germany, 2006. Springer-Verlag.

[50] Z. Han and B. H. Krogh. Reachability analysis of nonlinear systems using trajectory piecewise linearized models. In *Proc. American Control Conference*, pages 1505–1510, Minneapolis, MN, 2006.

[51] P. A. Ioannou and J. Sun. *Robust Adaptive Control*. Prentice Hall, Englewood Cliffs, NJ, 1996.

[52] C. Ionescu, R. De Keyser, B. Torrico, T. De Smet, M. Struys, and J. Normey-Rico. Robust predictive control strategy applied for propofol dosing using BIS as a controlled variable during anesthesia. *IEEE Transactions on Biomedical Engineering*, 55(9):2161–2170, 2008.

[53] K. J. Johansson, M. Egerstedt, J. Lygeros, and S. Sastry. On the regularization of Zeno hybrid automata. *Systems and Control Letters*, 38:141–150, 1999.

[54] A. Kanade, R. Alur, F. Ivančić, and S. Ramesh. Generating and analyzing symbolic traces of Simulink/Stateflow models. In A. Bouajjani and O. Maler, editors, *Computer Aided Verification, LNCS 5643*, pages 430–445. Springer Berlin Heidelberg, 2009.

[55] S. Kaynama, J. Maidens, M. Oishi, I. M. Mitchell, and G. A. Dumont. Computing the viability kernel using maximal reachable sets. In *Hybrid Systems: Computation and Control*, pages 55–63, Beijing, China, 2012.

[56] S. Kaynama and M. Oishi. A modified Riccati transformation for complexity reduction in reachability analysis of linear time-invariant systems. *IEEE Transactions on Automatic Control*. (accepted; preprint available at www.ece.ubc.ca/∼kaynama).

[57] S. Kaynama and M. Oishi. Schur-based decomposition for reachability analysis of linear time-invariant systems. In *Proc. IEEE Conference on Decision and Control*, pages 69–74, Shanghai, China, December 2009.

[58] S. Kaynama and M. Oishi. Complexity reduction through a Schur-based decomposition for reachability analysis of linear time-invariant systems. *International Journal of Control*, 84(1):165–179, 2011.

[59] S. Kaynama, M. Oishi, I. M. Mitchell, and G. A. Dumont. The continual reachability set and its computation using maximal reachability techniques. In *Proc. IEEE Conference on Decision and Control, and European Control Conference*, pages 6110–6115, Orlando, FL, 2011.

[60] S. Kaynama, M. Oishi, I. M. Mitchell, and G. A. Dumont. Fixed-complexity piecewise ellipsoidal representation of the continual reachability set based on ellipsoidal techniques. In *Proc. American Control Conference*, Montreal, QC, 2012. (to appear).

[61] E. C. Kerrigan. *Robust Constraint Satisfaction: Invariant Sets and Predictive Control*. PhD thesis, University of Cambridge, 2000.

[62] P. V. Kokotović. A Riccati equation for block-diagonalization of ill-conditioned systems. *IEEE Transactions on Automatic Control*, 20(6):812–814, 1975.

[63] P. V. Kokotović, H. K. Khalil, and J. O'Reilly. *Singular Perturbation Methods in Control: Analysis and Design*. SIAM, 1999.

[64] E. K. Kostousova. Control synthesis via parallelotopes: optimization and parallel computations. *Optimization methods and software*, 14(4):267–310, 2001.

[65] B. H. Krogh and O. Stursberg. Efficient representation and computation of reachable sets for hybrid systems. In O. Maler and A. Pnueli, editors, *Hybrid Systems: Computation and Control, LNCS 2623*, pages 482–497, Berlin, Germany, 2003. Springer-Verlag.

[66] W. Kühn. Rigorously computed orbits of dynamical systems without the wrapping effect. *Computing*, 61(1):47–67, 1998.

[67] A. B. Kurzhanski and T.F. Filippova. On the description of the set of viable trajectories of a differential inclusion. *Sov. Math. Doklady*, 34, 1987.

[68] A. B. Kurzhanski, I. M. Mitchell, and P. Varaiya. Optimization techniques for state-constrained control and obstacle problems. *Journal of Optimization Theory and Applications*, 128(3):499–521, 2006.

[69] A. B. Kurzhanski and I. Vályi. *Ellipsoidal Calculus for Estimation and Control*. Birkhäuser, Boston, MA, 1996.

[70] A. B. Kurzhanski and P. Varaiya. Ellipsoidal techniques for reachability analysis. In N. Lynch and B. Krogh, editors, *Hybrid Systems: Computation and Control, LNCS 1790*, pages 202–214, Berlin Heidelberg, 2000. Springer-Verlag.

[71] A. B. Kurzhanski and P. Varaiya. Ellipsoidal techniques for reachability analysis: internal approximation. *Systems & Control Letters*, 41:201–211, 2000.

[72] A. B. Kurzhanski and P. Varaiya. On reachability under uncertainty. *SIAM Journal on Control and Optimization*, 41(1):181–216, 2002.

[73] A. A. Kurzhanskiy and P. Varaiya. Ellipsoidal Toolbox (ET). In *Proc. IEEE Conference on Decision and Control*, pages 1498–1503, San Diego, CA, December 2006.

[74] A. A. Kurzhanskiy and P. Varaiya. Ellipsoidal techniques for reachability analysis of discrete-time linear systems. *IEEE Transactions on Automatic Control*, 52(1):26–38, 2007.

[75] A. A. Kurzhanskiy and P. Varaiya. Computation of reach sets for dynamical systems. In *Chapter for Control Handbook*. 2 edition, 2008.

[76] M. Kvasnica, P. Grieder, M. Baotić, and M. Morari. Multi-Parametric Toolbox (MPT). In R. Alur and G. J. Pappas, editors, *Hybrid Systems: Computation and Control, LNCS 2993*, pages 448–462, Berlin, Germany, 2004. Springer.

[77] C. Le Guernic. *Reachability analysis of hybrid systems with linear continuous dynamics*. PhD thesis, Université Grenoble 1 – Joseph Fourier, 2009.

[78] C. Le Guernic and A. Girard. Reachability analysis of linear systems using support functions. *Nonlinear Analysis: Hybrid Systems*, 4(2):250–262, 2010.

[79] J. Lygeros. On reachability and minimum cost optimal control. *Automatica*, 40(6):917–927, June 2004.

[80] J. Lygeros, D. N. Godbole, and S. Sastry. Verified hybrid controllers for automated vehicles. *IEEE Transactions on Automatic Control*, 43(4):522–539, Apr 1998.

[81] J. Lygeros, C. J. Tomlin, and S. Sastry. Controllers for reachability specifications for hybrid systems. *Automatica*, 35:349–370, 1999.

[82] M. S. Mahmoud and M. G. Singh. *Large Scale Systems Modelling*. Pergamon Press, 1981.

[83] K. Margellos and J. Lygeros. Air traffic management with target windows: An approach using reachability. In *Proc. IEEE Conference on Decision and Control*, pages 145–150, Shanghai, China, Dec 2009.

[84] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36:789–814, 2000.

[85] T. Mendonca, J. Lemos, H. Magalhaes, P. Rocha, and S. Esteves. Drug delivery for neuromuscular blockade with supervised multimodel adaptive control. *IEEE Transactions on Control Systems Technology*, 17(6):1237–1244, November 2009.

[86] I. M. Mitchell. Comparing forward and backward reachability as tools for safety analysis. In A. Bemporad, A. Bicchi, and G. Buttazzo,

editors, *Hybrid Systems: Computation and Control, LNCS 4416*, pages 428–443, Berlin Heidelberg, 2007. Springer-Verlag.

[87] I. M. Mitchell. A toolbox of level set methods. Technical report, UBC Department of Computer Science, TR-2007-11, June 2007.

[88] I. M. Mitchell. Scalable calculation of reach sets and tubes for nonlinear systems with terminal integrators: a mixed implicit explicit formulation. In *Proc. Hybrid Systems: Computation and Control*, pages 103–112, Chicago, IL, 2011. ACM.

[89] I. M. Mitchell, A. M. Bayen, and C. J. Tomlin. A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on Automatic Control*, 50(7):947–957, July 2005.

[90] I. M. Mitchell and J. A. Templeton. A toolbox of Hamilton-Jacobi solvers for analysis of nondeterministic continuous and hybrid systems. In M. Morari and L. Thiele, editors, *Hybrid Systems: Computation and Control, LNCS 3414*, pages 480–494, Berlin, Germany, 2005. Springer-Verlag.

[91] I. M. Mitchell and C. J. Tomlin. Overapproximating reachable sets by Hamilton-Jacobi projections. *Journal of Scientific Computing*, 19(1–3):323–346, 2003.

[92] A. S. Morse. Supervisory control of families of linear set-point controllers—part 1: exact matching. *IEEE Transactions on Automatic Control*, 41:1413–1431, 1996.

[93] T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall. The double description method. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, volume II, pages 51–73, 1953.

[94] G. Nenninger, G. Frehse, and V. Krebs. Reachability analysis and control of a special class of hybrid systems. In *Modelling, Analysis, and Design of Hybrid Systems, LNCIS 279*, pages 173–192. Springer Berlin Heidelberg, 2002.

[95] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, NY, 1999.

[96] M. Oishi, C. J. Tomlin, V. Gopal, and D. N. Godbole. Addressing multiobjective control: Safety and performance through constrained optimization. In *Hybrid Systems: Computation and Control, LNCS 2034*, pages 459–472. Springer-Verlag, 2001.

[97] P. Oliveira, J. P. Hespanha, J. M. Lemos, and T. Mendonça. Supervised multi-model adaptive control of neuromuscular blockade with off-set compensation. In *Proc. European Control Conference*, 2009.

[98] D. Panagou, K. Margellos, S. Summers, J. Lygeros, and K. J. Kyriakopoulos. A viability approach for the stabilization of an underactuated underwater vehicle in the presence of current disturbances. In *Proc. IEEE Conference on Decision and Control*, pages 8612–8617, December 2009.

[99] S. Prajna and A. Jadbabaie. Safety verification of hybrid systems using barrier certificates. In R. Alur and G. Pappas, editors, *Hybrid Systems: Computation and Control*, volume LNCS 2993, pages 477–492, 2004.

[100] S. V. Raković. Set theoretic methods in model predictive control. *Nonlinear Model Predictive Control*, pages 41–54, 2009.

[101] S. V. Raković and D. Q. Mayne. Set robust control invariance for linear discrete time systems. pages 975–980, 2005.

[102] A. Rantzer and S. Prajna. On analysis and synthesis of safe control laws. In *Proc. the Allerton Conference on Communication, Control, and Computing*, pages 1–9, 2004.

[103] C. Radhakrishna Rao and Sujit Kumar Mitra. Generalized inverse of a matrix and its applications. In *Proc. sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages 601–620, 1972.

[104] M. G. Safonov and R. Y. Chiang. A Schur method for balanced-truncation model reduction. *IEEE Transactions on Automatic Control*, 34(7), 1989.

[105] M. G. Safonov, A. J. Laub, and G. Hartmann. Feedback properties of multivariable systems: The role and use of return difference matrix. *IEEE Transactions on Automatic Control*, 26(1):47–65, 1981.

[106] P. Saint-Pierre. Approximation of the viability kernel. *Applied Mathematics and Optimization*, 29(2):187–209, Mar 1994.

[107] K.-H. Shim and M. E. Sawan. Singularly perturbed unified time systems with low sensitivity to model reduction using delta operators. *International Journal of Systems Science*, 37(4):243–251, 2006.

[108] Norihiko Shishido and Claire J. Tomlin. Ellipsoidal approximations of reachable sets for linear games. In *Proc. IEEE Conference on Decision and Control*, pages 999–1004, Sydney, Australia, December 2000.

[109] O. Simanski, A. Schubert, R. Kaehler, M. Janda, J. Bajorat, R. Hofmockel, and B. Lampe. Automatic drug delivery in anesthesia: From the beginning until now. In *Proc. Mediterranean Conf. Contr. Automation*, Athens, Greece, 2007.

[110] J. M. Siret, G. Michailesco, and P. Bertrand. Optimal approximation of high-order systems subject to polynomial inputs. *International Journal of Control*, 26(6):963–971, 1977.

[111] S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control; Analysis and Design*. John Wiley & Sons, West Sussex, UK, 2007.

[112] D. R. Smith. Decoupling and order reduction via the Riccati transformation. *SIAM Review*, 29(1):91–113, 1987.

[113] D. M. Stipanović, I. Hwang, and C. J. Tomlin. Computation of an overapproximation of the backward reachable set using subsystem level set functions. In *Proc. IEE European Control Conference*, Cambridge, UK, September 2003.

[114] G. Strang. *Linear Algebra and Its Applications*. Brooks Cole, 1988.

[115] S. Syafiie, J. Niño, C. Ionescu, and R. De Keyser. NMPC for propofol drug dosing during anesthesia induction. In *Nonlinear Model Predictive Control*, volume 384, pages 501–509. Springer Berlin Heidelberg, 2009.

[116] C. J. Tomlin, J. Lygeros, and S. Sastry. A game theoretic approach to controller design for hybrid systems. *Proceedings of the IEEE*, 88(7):949–970, 2000.

[117] C. J. Tomlin, I. M. Mitchell, A. M. Bayen, and M. Oishi. Computational techniques for the verification and control of hybrid systems. *Proceedings of the IEEE*, 91(7):986–1001, 2003.

[118] P. Varaiya. Reach set computation using optimal control. In *Proc. KIT Workshop on Verification of Hybrid Systems*. Verimag, Grenoble, 1998.

[119] R. Vasudevan, V. Shia, Y. Gao, R. Cervera-Navarro, R. Bajcsy, and F. Borrelli. Safe semi-autonomous control with enhanced driver modeling. In *Proc. American Control Conference*, Montreal, QC, 2012. (to apprear).

[120] R. Vidal, S. Schaffert, J. Lygeros, and S. Sastry. Controlled invariance of discrete time systems. In B. Krogh and N. Lynch, editors, *Hybrid Systems and Control, LNCS 1790*, pages 437–451, Berlin Heidelberg, 2000. Springer-Verlag.

[121] H. Yazarel and G. J. Pappas. Geometric programming relaxations for linear system reachability. In *Proc. American Control Conference*, pages 553–559, Boston, MA, June 2004.

[122] C. C. Zervos and G. A. Dumont. Multivariable self-tuning control based on Laguerre series representation. *Adaptive Control Strategies for Industrial Use*, 137:44–57, 1989.

[123] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, Englewood Cliffs, NJ, 1996.

Appendix A

Supplementary Materials for Schur-Based Decomposition

A.1 On Assumption 3.1 (and 4.2)

Proposition A.1. *With $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{m\times p}$, the equation $XA = B$ has a solution w.r.t. $X \in \mathbb{R}^{m\times n}$ if and only if $\mathscr{C}(B^T) \subseteq \mathscr{C}(A^T)$.*

Proof. By taking the transpose of both sides we have $A^T X^T = B^T$. Let $Y := X^T$. The equation $A^T Y = B^T$ is known to have at least one solution w.r.t. Y if and only if the column space of B^T is in the image (or range) of the linear operator A^T, i.e. $\mathscr{C}(B^T) \subseteq \mathscr{C}(A^T)$. □

To check if the condition holds, we can check if

$$\operatorname{span}\{\vec{b}_1, \ldots, \vec{b}_m\} \subseteq \operatorname{span}\{\vec{a}_1, \ldots, \vec{a}_n\}, \tag{A.1}$$

where $\vec{b}_i$ and $\vec{a}_i$ denote the ith row of the matrices B and A, respectively. Or equivalently, we can check if the following rank condition holds:

$$\operatorname{rank}\left(\left[A^T | B^T\right]\right) = \operatorname{rank}\left(A^T\right). \tag{A.2}$$

A.2 Proof of Proposition 3.4

To prove Proposition 3.4 let us first state a simple lemma.

Lemma A.1. *Consider the backward reachable tube $Reach^{\flat}_{[0,t]}(\mathcal{K},\mathcal{U})$ of (2.12) over the interval $[0,t]$, $t \in [0,\tau]$, for a fixed $\tau \in \mathbb{R}^+$. Denote by $Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K})$*

the backward reachable tube of its corresponding autonomous system

$$\dot{x} = Ax. \tag{A.3}$$

The following inclusions hold:

$$\mathcal{K} \subseteq \mathit{Reach}^{\flat}_{[0,t]}(\mathcal{K},\mathcal{U}) \subseteq \mathit{Reach}^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K}) \quad \forall t \in [0,\tau]. \tag{A.4}$$

Proof. Assume, without loss of generality, that $\mathcal{U}$ is a compact hyper-rectangular subset of $\mathbb{R}^p$ such that $\mathcal{U} = \prod_{i=1}^{p} \mathcal{U}_i$, $u_i \in \mathcal{U}_i = [\underline{\mathcal{U}}_i, \overline{\mathcal{U}}_i]$, $0 \in \mathcal{U}_i$. Notice that the trajectories of the autonomous system (A.3) approach those of the controlled system (2.12) as

$$\xi := \sup\{\|u\| \colon u \in \mathcal{U}\} \tag{A.5}$$

tends to zero. As such, we draw on the level set formulation of the backward reachable tube of system (2.12) and treat (A.3) as a particular form of (2.12) in which the control input u is diminished.

It is well-known [89] that if $\mathcal{K}$ is represented as the zero sub-level set of some bounded and Lipschitz continuous implicit surface function $g \colon \mathbb{R}^n \to \mathbb{R}$, i.e. $\mathcal{K} = \{x \mid g(x) \leq 0\}$, then the backward reachable tube $\mathit{Reach}^{\flat}_{[0,t]}(\mathcal{K},\mathcal{U})$ can be obtained as the zero sub-level set of the viscosity solution $\phi \colon \mathbb{R}^n \times [0,\tau] \to \mathbb{R}$ of the modified terminal value HJB PDE

$$\nabla_t \phi(x,t) = -\min\left\{0, H\left(x, \nabla_x \phi(x,t)\right)\right\}, \quad \phi(x,\tau) = g(x) \tag{A.6}$$

$$H(x,\ell) = \sup_{u \in \mathcal{U}} \langle \ell, Ax + Bu \rangle \tag{A.7}$$

with the Hamiltonian $H(\cdot,\cdot)$ and the costate vector ℓ. Here, $\langle \cdot,\cdot \rangle$ denotes the inner product. Thus, $\mathit{Reach}^{\flat}_{[0,t]}(\mathcal{K},\mathcal{U}) = \{x \mid \phi(x,\tau - t) \leq 0\}$. The optimal Hamiltonian, in this case, can be determined analytically as

$$H^*(x,\ell) = \ell^T A x + \ell^T B u^*, \quad u^* = [u_1^* \cdots u_p^*]^T \tag{A.8}$$

with

$$u_i^* = \begin{cases} \underline{\mathcal{U}}_i & \text{if} \quad \ell^T b_i < 0; \\ [\underline{\mathcal{U}}_i, \overline{\mathcal{U}}_i] & \text{if} \quad \ell^T b_i = 0; \\ \overline{\mathcal{U}}_i & \text{if} \quad \ell^T b_i > 0 \end{cases} \quad , \quad i = 1, \ldots, p \tag{A.9}$$

where b_i is the i-th column vector of matrix B. Notice that the second term on the right hand side of (A.8) is always non-negative, i.e. $\ell^T B u^* \geq 0$. Therefore we have

$$\nabla_t \phi(x,t) = \begin{cases} 0 & \text{if} \quad \ell^T A x \geq -\ell^T B u^*; \\ |\ell^T A x| - \ell^T B u^* & \text{otherwise.} \end{cases} \tag{A.10}$$

When $\xi = 0$, the controlled system (2.12) is equivalent to the autonomous system (A.3) and the Hamiltonian (A.7) becomes $H(x,\ell) = H^*(x,\ell) = \ell^T A x$. Consequently, (A.10) reduces to

$$\nabla_t \phi(x,t)\Big|_{\xi=0} =: \nabla_t \phi^{\mathfrak{A}}(x,t) = \begin{cases} 0 & \text{if} \quad \ell^T A x \geq 0; \\ |\ell^T A x| & \text{otherwise} \end{cases} \tag{A.11}$$

where $\phi^{\mathfrak{A}}(\cdot,\cdot)$ is to denote the implicit surface function whose zero sub-level set determines the backward reachable tube $Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K})$ of (A.3). That is, $Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K}) = \{x \mid \phi^{\mathfrak{A}}(x, \tau - t) \leq 0\}$.

Comparing (A.10) and (A.11) one can observe that not only the interval over which $\nabla_t \phi$ (the rate of surface change in time) is zero is shortened (i.e. $\ell^T A x \geq 0$ as opposed to $\ell^T A x \geq -\ell^T B u^*$), but also its maximum (positive) value is increased (i.e. $|\ell^T A x|$ as opposed to $|\ell^T A x| - \ell^T B u^*$). Therefore, for all $(x,t) \in \mathbb{R}^n \times [0,\tau]$ we have

$$\nabla_t \phi^{\mathfrak{A}}(x, \tau - t) \geq \nabla_t \phi(x, \tau - t) \tag{A.12}$$

$$\Longrightarrow \quad \phi^{\mathfrak{A}}(x, \tau - t) \leq \phi(x, \tau - t) \leq \phi(x, \tau) \leq 0 \tag{A.13}$$

$$\Longleftrightarrow \quad Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K}) \supseteq Reach^{\flat}_{[0,t]}(\mathcal{K}, \mathcal{U}) \supseteq \mathcal{K}. \tag{A.14}$$

□

Notice that this result agrees with the intuitive interpretation that larger control authority (i.e. $\xi \neq 0$) implies a smaller *unsafe* minimal reachable tube. We are now ready to prove Proposition 3.4.

Proof of Proposition 3.4. Using Lemma A.1 we have

$$Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K}) \supseteq Reach^{\flat}_{[0,t]}(\mathcal{K},\mathcal{U}) \quad \forall t \tag{A.15}$$

where $Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K})$ denotes the backward reachable tube of the autonomous system (A.3). Therefore, to prove $Reach^{\flat}_{[0,t]}(\mathcal{K},\mathcal{U}) = \mathcal{K}$, $\forall t \in [0,\tau]$, it is sufficient to show that $Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K}) = \mathcal{K}$, $\forall t \in [0,\tau]$.

Let $S\Lambda S^{-1}$ be the eigen-decomposition of A. Conditions (ii) and (iii) imply $\Lambda = \lambda I_n$, $\lambda \geq 0$. Rewriting the Hamiltonian of the HJB PDE (A.6) for the autonomous system (A.3) and using condition (i) we have

$$\begin{aligned} H(x, \nabla_x \phi^{\mathfrak{A}}(x,t)) &= \langle \nabla_x \phi^{\mathfrak{A}}(x,t), Ax \rangle & \text{(A.16)} \\ &= \langle \nabla_x \phi^{\mathfrak{A}}(x,t), S\Lambda S^{-1} x \rangle & \text{(A.17)} \\ &= \lambda \langle \nabla_x \phi^{\mathfrak{A}}(x,t), x \rangle \geq 0 \qquad \forall (x,t) \in \mathbb{R}^n \times [0,\tau]. & \text{(A.18)} \end{aligned}$$

The non-negativity of the Hamiltonian is due to the fact that $\mathcal{K}$ is convex and $\mathbf{0} \in \mathcal{K}$. Thus, the costate vector $\nabla_x \phi^{\mathfrak{A}}(x,t)$ at every point on the boundary constitutes an acute (hyper-) angle with respect to the trajectory x initiating from that point in forward time. This is schematically illustrated for a trivial planar system in Figure A.1. As a result, for all $(x,t) \in \mathbb{R}^n \times [0,\tau]$ we have

$$\begin{aligned} H(x, \nabla_x \phi^{\mathfrak{A}}(x,t)) \geq 0 &\iff \nabla_t \phi^{\mathfrak{A}}(x,t) = 0 & \text{(A.19)} \\ &\iff Reach^{\flat\mathfrak{A}}_{[0,t]}(\mathcal{K}) = \mathcal{K}. & \text{(A.20)} \end{aligned}$$

This concludes the proof. □

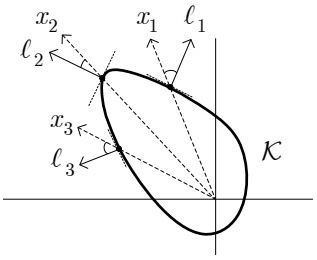


Figure A.1: Three sample costate vectors (ℓ_i) and trajectories (x_i) initiating from the boundary of an arbitrarily-shaped convex target set $\mathcal{K}$ in the phase-plane of a simple planar system in forward time. Notice the non-negativity of $\langle \ell_i, x_i \rangle$ as shown in forward time. In backward time the trajectories are reversed and the eigenvalues are negated, hence the Hamiltonian is still non-negative.

A.3 Decomposed System Matrices for Examples 3.6.2 and 3.6.3

A.3.1 4D Aircraft Dynamics (Example 3.6.2)

The decomposed system matrices are:

$$A_d = \left[\begin{array}{cc:cc} 0.1527 & -0.1511 & 0.0312 & 0 \\ 0.1853 & -0.1536 & -0.0247 & 0.0065 \\ \hdashline 0 & 0 & -0.3169 & -7.5973 \\ 0 & 0 & 0.1028 & -0.4331 \end{array}\right], \quad B_d = \left[\begin{array}{c} 0 \\ 0 \\ \hdashline -0.1785 \\ 1.1598 \end{array}\right].$$

A.3.2 8D Distillation Column (Example 3.6.3)

The system matrices after the first decomposition are:

$$A_d = \left[\begin{array}{cccc:cccc} -0.6460 & -2.7152 & 0.9186 & -1.0340 & -1.5499 & 0.0128 & 0.3436 & 0.0404 \\ 3.0705 & -0.6461 & -0.8642 & 0.6817 & 1.5878 & -0.0215 & -0.0567 & -0.0336 \\ 0 & 0 & -0.8627 & 1.6880 & 2.0531 & 0 & 0.0135 & 0 \\ 0 & 0 & -0.6716 & -0.8627 & -0.5888 & 0 & -0.2143 & 1.2635 \\ \hdashline 0 & 0 & 0 & 0 & -0.7357 & -0.2275 & -0.0082 & -0.0021 \\ 0 & 0 & 0 & 0 & 0 & -0.2259 & 0.0021 & -0.0457 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.0052 & 0.0024 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.0755 \end{array}\right]$$

$$B_d = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \hdashline 1.2886 & -0.0504 \\ 0.3132 & -0.2249 \\ 0.7117 & -0.6994 \\ 0.0599 & -0.3014 \end{array}\right].$$

The unidirectional coupling term in A_d is treated as a disturbance to the upper subsystem. The second level decomposition applied to each 4D subsystem results in

$$A_{d_1} = \left[\begin{array}{cc:cc} -1.3594 & -1.2819 & 0 & 0 \\ 1.0769 & -0.3660 & 0 & 0 \\ \hdashline 0 & 0 & -0.9978 & -2.8164 \\ 0 & 0 & 3.0041 & -0.2943 \end{array}\right], \quad B_{d_1} = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ \hdashline 0 & 0 \\ 0 & 0 \end{array}\right],$$

$$A_{d_2} = \left[\begin{array}{cc:cc} -0.0052 & 0.0026 & 0.0215 & 0.5192 \\ 0 & -0.0755 & -0.1334 & 0.1262 \\ \hdashline 0 & 0 & -0.7479 & -0.1877 \\ 0 & 0 & 0.0339 & -0.2137 \end{array}\right], \; B_{d_2} = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ \hdashline 1.3130 & -0.0574 \\ 0.2260 & -0.2934 \end{array}\right].$$

Appendix B

Supplementary Materials for Riccati-Based Decomposition

B.1 Proofs of Propositions 4.2 and 4.3

Proof of Proposition 4.2. From the matrix inversion lemma, $(Y+UCV)^{-1} = Y^{-1} - Y^{-1}U(C^{-1}+VY^{-1}U)^{-1}VY^{-1}$, with $Y = -(\delta+1)I$, $U = B_1$, $C = I$, and $V = B_1^\dagger$ we have

$$\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1} = -\frac{1}{\delta+1}\Big(I + \frac{1}{\delta}B_1B_1^\dagger\Big). \tag{B.1}$$

Using this, (4.14), (4.21), (4.23), (4.26), multiplicative and triangular inequalities, and the fact that $\|B_1B_1^\dagger\| \geq 1$ we obtain

$$\begin{aligned}
\|\delta\mathscr{F}(Z(\delta))\| &\leq |\delta|\big(\alpha(\|Z_0\| + \|D\|)^2 + \beta(\|Z_0\| + \|D\|)\big) \\
&\leq |\delta|\bigg(\alpha\bigg(\|Z_0\| + \frac{2\|A_0\|\|Z_0\|}{\|A_0\| + \alpha\|Z_0\|}\bigg)^2 \\
&\qquad + \beta\bigg(\|Z_0\| + \frac{2\|A_0\|\|Z_0\|}{\|A_0\| + \alpha\|Z_0\|}\bigg)\bigg) \\
&\leq |\delta|\big(9\alpha\|Z_0\|^2 + 3\beta\|Z_0\|\big) \\
&\leq |\delta|\Big(9\alpha\gamma^2\big\|\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1}\big\|^2 \\
&\qquad + 3\beta\gamma\big\|\big(B_1B_1^\dagger - (\delta+1)I\big)^{-1}\big\|\Big) \\
&\leq |\delta|\bigg(9\alpha\gamma^2\Big|\frac{1}{\delta+1}\Big|^2\Big(1 + \Big|\frac{1}{\delta}\Big|\Big)^2\|B_1B_1^\dagger\|^2 \\
&\qquad + 3\beta\gamma\Big|\frac{1}{\delta+1}\Big|\Big(1 + \Big|\frac{1}{\delta}\Big|\Big)\|B_1B_1^\dagger\|\bigg)
\end{aligned}$$

$$= \frac{1}{|\delta|}\left(\frac{|\delta|+1}{|\delta+1|}\right)^2 a + \left(\frac{|\delta|+1}{|\delta+1|}\right) b \qquad \forall \delta \in \mathbb{R}\backslash\{-1,0\}.$$

□

Proof of Proposition 4.3. Notice from (4.25) and (4.28) that for large values of δ, Z can be closely approximated by its initial value Z_0. Using (B.1),

$$\lim_{\delta\to\pm\infty} \|\delta\mathscr{F}(Z(\delta))\| = \lim_{\delta\to\pm\infty} \Big\| \frac{\delta}{(\delta+1)^2} Q_1 (I + \frac{1}{\delta} B_1 B_1^\dagger) P_1 Q_1 \times (I + \frac{1}{\delta} B_1 B_1^\dagger) + \frac{\delta}{\delta+1} P_2 Q_1 (I + \frac{1}{\delta} B_1 B_1^\dagger) \Big\| \quad \text{(B.2)}$$

$$= \|0 + P_2 Q_1\| = \|\Gamma\| \quad \text{(B.3)}$$

with $Q_1 := \big(B_2 B_1^\dagger A_{12} - A_{22}\big)^{-1}\Gamma$, $P_1 := (A_{12} - B_1 B_1^\dagger A_{12})$, and $P_2 := (B_2 B_1^\dagger A_{12} - A_{22})$. □

B.2 Formulating an Upper-Bound on the Condition Number of the Modified Riccati Transformation

Lemma B.1. *The condition number $\kappa(T)$ of the Riccati transformation matrix $T = T_1 T_2$ is bounded by*

$$\kappa(T) \le \max\big\{1+\mu,\, 1+\nu(1+\mu)\big\} \cdot \max\big\{1+\nu,\, 1+\mu(1+\nu)\big\} \quad \text{(B.4)}$$

with constants

$$\mu := \frac{2\|N_0\|\|M_0\|}{\|N_0\| + \|\delta\mathscr{F}(Z)\|\|M_0\|} + \|M_0\|,$$

$$\nu := \|B_2 B_1^\dagger\| + (1 + \|B_1 B_1^\dagger\|) \left[\frac{2\|A_0\|\|Z_0\|}{\|A_0\| + \|B_1 B_1^\dagger A_{12} - A_{12}\|\|Z_0\|} + \|Z_0\| \right].$$

Proof. Let $A = [A_{ij}]$ be any partitioned matrix. Then $\|A\| \le \big\|\big[\|A_{ij}\|\big]\big\|$.

Since $T = \left[\begin{smallmatrix} I & M \\ -L & I-LM \end{smallmatrix}\right]$ and $T^{-1} = \left[\begin{smallmatrix} I-ML & -M \\ L & I \end{smallmatrix}\right]$, we have

$$\kappa(T) = \|T\|\|T^{-1}\| \leq \left\| \begin{bmatrix} 1 & \|M\| \\ \|L\| & 1+\|L\|\|M\| \end{bmatrix} \right\| \left\| \begin{bmatrix} 1+\|M\|\|L\| & \|M\| \\ \|L\| & 1 \end{bmatrix} \right\|. \tag{B.5}$$

From (4.10) we find $\|L\| \leq \|B_2 B_1^\dagger\| + \|Z\|\big(1 + \|B_1 B_1^\dagger\|\big)$. But from (4.23) and (4.26) we know that $\|Z\|$ is bounded above by $\frac{2\|A_0\|\|Z_0\|}{\|A_0\|+\|B_1 B_1^\dagger A_{12}-A_{12}\|\|Z_0\|} + \|Z_0\|$, and similarly from (4.33) and (4.36), $\|M\|$ is bounded above by $\frac{2\|N_0\|\|M_0\|}{\|N_0\|+\|\delta\mathscr{F}(Z)\|\|M_0\|} + \|M_0\|$. Therefore, $\|L\| \leq \nu$ and $\|M\| \leq \mu$ and consequently, $\kappa(T) \leq \max\{1+\mu,\, 1+\nu(1+\mu)\} \cdot \max\{1+\nu,\, 1+\mu(1+\nu)\}$. $\square$

B.3 Decomposed System Matrices for Example 4.5.3 and Section 4.5.5

B.3.1 Arbitrary 6D System (Example 4.5.3)

The decomposed system matrices are:

$$A'' = \left[\begin{array}{ccc:ccc} 3.3126 & 0.7676 & 2.4511 & 0 & 0 & 0 \\ 0.6223 & -1.5072 & -2.3105 & 0 & 0 & 0 \\ -0.0989 & -0.3285 & 0.6852 & 0 & 0 & 0 \\ \hdashline 0.0944 & 0.0042 & -0.1299 & 0.1880 & 0.0445 & 0.3528 \\ -0.0540 & -0.4170 & -0.0461 & 0.2802 & 0.0888 & 0.1593 \\ -0.1474 & 0.1949 & 0.2443 & 0.0848 & 0.0888 & 0.1197 \end{array}\right],$$

$$B'' = \left[\begin{array}{cc} 0.1469 & 0.2657 \\ -0.7988 & 2.4582 \\ -2.3854 & -0.3955 \\ \hdashline 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{array}\right].$$

B.3.2 Maximal Reachability Example (Section 4.5.5)

The decomposed system matrices are:

$$A'' = \left[\begin{array}{cc:cc} 0.5912 & -0.2477 & 0 & 0 \\ -0.4583 & -0.2017 & 0 & 0 \\ \hdashline 0.0197 & 0.1351 & -0.1905 & -0.1878 \\ -0.0825 & -0.1479 & 0.0012 & -0.0633 \end{array}\right], \quad B'' = \left[\begin{array}{c} 0.6846 \\ 2.4813 \\ \hdashline 0 \\ 0 \end{array}\right].$$

Appendix C

Other Backward Reachability Constructs[1]

Some additional backward constructs formed under the constrained dynamical system (2.1), their connections to one another and to the constructs used within this thesis (see Chapter 2) are presented here, aiming to help the reader attain a more complete picture.

C.1 Definitions and Connections

Definition C.1 (Minimal Reachable Set)**.** *The minimal reachable set at time t is the set of initial states such that, for every input $u(\cdot)$, the trajectories emanating from those states reach $\mathcal{K}$ exactly at time t:*

$$Reach_t^{\flat}(\mathcal{K},\mathcal{U}) := \big\{x_0 \in \mathcal{X} \mid \forall u(\cdot) \in \mathscr{U}_{[0,t]},\, x_{x_0}^u(t) \in \mathcal{K}\big\}. \tag{C.1}$$

Definition C.2 (Invariance Kernel)**.** *The (finite-horizon) invariance kernel of $\mathcal{K}$ is the set of initial states in $\mathcal{K}$ such that the trajectories emanating from those states remain within $\mathcal{K}$ for all time $t \in [0,\tau]$ for all input $u(\cdot)$:*

$$Inv_{[0,\tau]}(\mathcal{K},\mathcal{U}) := \big\{x_0 \in \mathcal{X} \mid \forall u(\cdot) \in \mathscr{U}_{[0,\tau]},\, \forall t \in [0,\tau],\, x_{x_0}^u(t) \in \mathcal{K}\big\}. \tag{C.2}$$

Definition C.3 (Continual Reachable Set [59, 60])**.** *The continual reachable set defined over the time horizon $[0,\tau]$ is the set of initial states in $\mathcal{K}$ for which, for any given time $t \in [0,\tau]$, there exists a $u(\cdot)$ such that the*

[1] A part of this chapter is based on [59, 60]. The main results of these papers, however, have not been presented in this thesis.

trajectories emanating from those states reach $\mathcal{K}$ at t:

$$Reach^{\gamma}_{[0,\tau]}(\mathcal{K},\mathcal{U}) := \big\{x_0 \in \mathcal{X} \mid \forall t \in [0,\tau],\, \exists u(\cdot) \in \mathscr{U}_{[0,t]},\, x^u_{x_0}(t) \in \mathcal{K}\big\}. \tag{C.3}$$

The following inclusions are complementary to $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq \mathcal{K} \subseteq Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Reach^{\sharp}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ described in Proposition 2.1.

Proposition C.1.

$$Inv_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Reach^{\gamma}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq \mathcal{K}. \tag{C.4}$$

Proof. That $Inv_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$ is well-known [5]. To show $Viab_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq Reach^{\gamma}_{[0,\tau]}(\mathcal{K},\mathcal{U})$, take $x_0 \in Viab_{[0,\tau]}(\mathcal{K},\mathcal{U})$. Therefore, $\exists u(\cdot) \in \mathscr{U}_{[0,\tau]}\ \forall t \in [0,\tau]\ \ x^u_{x_0}(t) \in \mathcal{K} \Longrightarrow \forall t \in [0,\tau]\ \exists u(\cdot) \in \mathscr{U}_{[0,t]}\ \ x^u_{x_0}(t) \in \mathcal{K} \Longleftrightarrow x_0 \in Reach^{\gamma}_{[0,\tau]}(\mathcal{K},\mathcal{U})$. To show $Reach^{\gamma}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \subseteq \mathcal{K}$, take $x_0 \in Reach^{\gamma}_{[0,\tau]}(\mathcal{K},\mathcal{U})$ and let $\tau = 0$. x_0 must also belong to $\mathcal{K}$. □

The maximal reachable tube and the invariance kernel are duals of one another as mentioned in Section 1.2:

Proposition C.2 ([18], [79])**.**

$$Reach^{\sharp}_{[0,\tau]}(\mathcal{K}^c,\mathcal{U}) = (Inv_{[0,\tau]}(\mathcal{K},\mathcal{U}))^c. \tag{C.5}$$

Unlike the maximal reachable tube and sets, the minimal reachable tube cannot be constructed from the union of the minimal reachable sets as mentioned in Section 2.2:

Proposition C.3 ([86])**.**

$$Reach^{\flat}_{[0,\tau]}(\mathcal{K},\mathcal{U}) \supseteq \bigcup_{t\in[0,\tau]} Reach^{\flat}_{t}(\mathcal{K},\mathcal{U}). \tag{C.6}$$

Among Lagrangian methods, the technique in [72] has been extended to handle universally quantified inputs. Therefore, it is also capable of computing the minimal reachable sets. As a by-product of this feature, the same technique can also be used to *directly* compute the invariance kernel.

Proposition C.4 ([59, 79])**.**

$$Inv_{[0,\tau]}(\mathcal{K},\mathcal{U}) = \bigcap_{t\in[0,\tau]} Reach_t^{\flat}(\mathcal{K},\mathcal{U}). \tag{C.7}$$

Proof. $x_0 \in \bigcap_{t\in[0,\tau]} Reach_t^{\flat}(\mathcal{K},\mathcal{U}) \iff \forall t \in [0,\tau]\ \forall u(\cdot) \in \mathscr{U}_{[0,\tau]}\ \ x_{x_0}^u(t) \in \mathcal{K} \iff \forall u(\cdot) \in \mathscr{U}_{[0,\tau]}\ \forall t \in [0,\tau]\ \ x_{x_0}^u(t) \in \mathcal{K} \iff x_0 \in Inv_{[0,\tau]}(\mathcal{K},\mathcal{U})$. This can also be verified from (C.5) and (2.9) and the simple fact that $Reach_t^{\flat}(\mathcal{K},\mathcal{U}) = (Reach_t^{\sharp}(\mathcal{K}^c,\mathcal{U}))^c$. □

Finally, the continual reachable set can be expressed in terms of the maximal reachable sets as:

Proposition C.5 ([59])**.**

$$Reach_{[0,\tau]}^{\gamma}(\mathcal{K},\mathcal{U}) = \bigcap_{t\in[0,\tau]} Reach_t^{\sharp}(\mathcal{K},\mathcal{U}). \tag{C.8}$$

Proof. $x_0 \in Reach_{[0,\tau]}^{\gamma}(\mathcal{K},\mathcal{U}) \iff \forall t \in [0,\tau]\ \exists u(\cdot) \in \mathscr{U}_{[0,t]}\ \ x_{x_0}^u(t) \in \mathcal{K} \iff \forall t \in [0,\tau]\ \ x_0 \in Reach_t^{\sharp}(\mathcal{K},\mathcal{U}) \iff x_0 \in \bigcap_{t\in[0,\tau]} Reach_t^{\sharp}(\mathcal{K},\mathcal{U})$. □