Safety-Preserving Control of Systems With Multiplicative Model Uncertainty

With Application to Closed-Loop Anesthesia

by

Mahdi Yousefi

M.A.Sc., The University of British Columbia, 2014
B.Sc., Amirkabir University of Technology (Tehran Polytechnic), 2012

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Electrical and Computer Engineering)

The University of British Columbia
(Vancouver)

December 2018

© Mahdi Yousefi, 2018
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Safety-preserving control of systems with multiplicative model uncertainty with application to closed-loop anesthesia

submitted by Mahdi Yousefi in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering

Examinig Committee:

Guy Dumont, Electrical and Computer Engineering
Supervisor

Ian Mitchell, Computer Science
Supervisory Committee Member

Mark Ansermino, Anaesthesiology, Therapeutic and Pharmacology
Supervisory Committee Member

Bhushan Gopaluni, Chemical and Biological Engineering
University Examiner

Ryozo Nagamune, Mechanical Engineering
University Examiner
Abstract

To convince the public to employ new technologies in their daily life, the reliability and safety of such technologies must be demonstrated. Verification of safety is more crucial in applications which may directly jeopardize human safety. A self-driving car is an example of such a technology in which the safety of passengers and pedestrians depends on the car’s actions made based on its perception of the environment. A closed-loop drug delivery system is another example of a safety-critical technology for which there are risks of drug over/under-dosing and adverse side effects. Uncertainty is the key challenge in safety verification of these technologies. In the self-driving car example, the uncertainty is in the car’s knowledge of the environment. In the context of closed-loop drug delivery systems, uncertainty is in the patient response to drugs due to inter-patient variability (model uncertainty). Unmeasurability of variables indicating the system’s safety (e.g., drug concentrations in the plasma) is another source of problems in safety verification of technologies such as closed-loop anesthesia.

Motivated by closed-loop anesthesia, this thesis aims to develop a mathematical framework for formal safety verification and safety-preserving control of safety-critical systems with model uncertainty. This extends formal methods and existing safety-preserving control techniques to systems with a certain class of multiplicative model uncertainty. Moreover, this work proposes one of the first safety-preserving control schemes for output-feedback control systems in which safety-critical variables are not measurable. We also extend this technique to uncertain output-feedback systems. In this work, we employ the developed techniques to design and formalize a safety system for closed-loop anesthesia. Finally, we propose a novel approach to reduce conservatism of the formalized safety system by
reducing model uncertainty using model falsification. The results discussed in this thesis may facilitate the process of obtaining regulatory approvals for closed-loop anesthesia, which helps the emergence of this technology (and other closed-loop drug delivery systems) in clinical environments.
Lay Summary

During surgery, anesthesiologists are tasked with sufficiently and yet safely sedating a patient. The best combination and dose of drugs for a given patient and procedure is selected and administered initially and then adapted manually during the procedure by an anesthesiologist. Due to the complexity and pressure of the clinical environment, and the large degree of inter-patient variability, the anesthesiologists performance is variable and suboptimal. Patient safety, postoperative outcomes, and recovery times can be negatively affected as a result. Just as autopilots have greatly improved the safety of aviation, an automated control system could improve the safety and limit the variability of anesthesia drug delivery.

Automated Anesthesia drug delivery systems have been implemented and used as investigational devices to anesthetize patients. To establish this technology as a clinical device, regulatory authorities must be convinced of its safety. In this context, safety is guaranteed once one shows that the drug delivery system does not overdose or underdose patients and maintains anesthetics adverse side effects (e.g. hypotension, respiratory depression) within a safe range. This thesis provides mathematical tools that can be employed to verify the safety of anesthesia (and other) drug delivery systems and to design safety systems that guarantee the safety despite variability in patient responses to anesthetics. The proposed safety analysis framework may facilitate the process of obtaining regulatory approval for automated drug delivery systems and the emergence of such technologies in clinical environments.
Preface

Motivated by closed-loop anesthesia, the current thesis proposes a mathematical framework that can be used for formal safety verification of control systems with model uncertainty and unmeasurable states. This work also formalizes safety systems for closed-loop anesthesia. These safety systems may facilitate the process of obtaining regulatory approvals for closed-loop anesthesia to be used as a clinical device, and thereby cause the more rapid emergence of this technology in clinical environments. All of the work presented in this thesis was conducted in BC Children’s Hospital Research Institute. Parts of the work presented in this thesis has been published in or submitted to several international conferences and scientific journals.

Results from Chapter 2 were published in:


I was the lead investigator responsible for all theoretical contributions, technical analysis and manuscripts composition. Dr. Klaske van Heusden was involved in the early stages of concept formulation and contributed to manuscripts edits. Dr. Guy Dumont, Dr. Mark Ansermino and Dr. Ian Mitchell were the supervisory
authors on the papers and were involved in concept formulation and manuscripts compositions.

Results from Chapter 3 were published in:


I was the lead investigator responsible for all theoretical contributions, technical analysis and manuscripts composition. Dr. Klaske van Heusden was involved in the early stages of concept formulation and contributed to manuscripts edits. Dr. Guy Dumont and Dr. Ian Mitchell were the supervisory authors on the papers and were involved in concept formulation and manuscripts compositions.

Results from Chapter 4 were published in:


I was the lead investigator responsible for all technical and clinical analysis and manuscripts composition. Dr. Klaske van Heusden was involved in the early stages of concept formulation and contributed to manuscripts edits. Nicholas West was involved in clinical analysis and contributed to manuscript edits. Dr. Guy Dumont, Dr. Mark Ansermino and Dr. Ian Mitchell were the supervisory authors on the papers and were involved in concept formulation and manuscripts compositions.

Results from Chapter 5 were published in:


I was the lead investigator responsible for all theoretical contributions, technical and clinical analysis and manuscripts composition. Dr. Klaske van Heusden was involved in the early stages of concept formulation and contributed to manuscripts edits. Dr. Guy Dumont, Dr. Mark Ansermino and Dr. Ian Mitchell were the supervisory authors on the papers and were involved in concept formulation and manuscripts compositions.

Some of the results discussed in Chapter 5 are based on the patient models identified in:


Dr. Klaske van Heusden was the lead author responsible for all theoretical contributions, technical and clinical analysis and manuscripts composition. My contribution to this work was limited. I was involved in the early stages of concept formulation and contributed to manuscripts edits. Dr. Guy Dumont, Dr. Mark Ansermino were the supervisory authors on the papers and were involved in concept formulation and manuscripts compositions.
# Table of Contents

Abstract .................................................................................................................. iii  
Lay Summary ........................................................................................................... v  
Preface .................................................................................................................... vi  
Table of Contents .................................................................................................... ix  
List of Tables .......................................................................................................... xii  
List of Figures ......................................................................................................... xiii  
Acknowledgements ................................................................................................ xiv  
1 Introduction ........................................................................................................... 1  
   1.1 Motivation ...................................................................................................... 1  
   1.2 Formal Verification and Safety-Preserving Control ..................................... 2  
   1.3 Objectives and Contributions ...................................................................... 4  
   1.4 Organization of The Thesis ........................................................................ 7  
2 Model-Invariant Safety-Preserving Control ......................................................... 8  
   2.1 Preliminaries ............................................................................................... 9  
      2.1.1 Notations ................................................................................................. 9  
      2.1.2 Safety-preserving control .................................................................... 10  
      2.1.3 Control synthesis .................................................................................. 11  
      2.1.4 Limitations .......................................................................................... 13  

ix
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Model-Invariant Safety Preserving Control</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Problem formulation</td>
<td>13</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Control synthesis</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Model-Invariant Viability Kernel Under-Approximation</td>
<td>19</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Viability kernel approximation</td>
<td>19</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Model-invariant viability kernel under-approximation</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>Conclusion</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>Output-Feedback Safety-Preserving Control</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>Background</td>
<td>31</td>
</tr>
<tr>
<td>3.2</td>
<td>Output-Feedback Safety-Preserving Control</td>
<td>33</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Notations and definitions</td>
<td>34</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Main result</td>
<td>35</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Discriminating kernel approximation</td>
<td>37</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Numerical comparison</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>Extension to Uncertain Systems</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusion</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>Case Study I - Formalized Safety Systems for Closed-loop</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>Background</td>
<td>48</td>
</tr>
<tr>
<td>4.2</td>
<td>Modeling Propofol Effects</td>
<td>49</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Propofol effect on depth of hypnosis</td>
<td>50</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Propofol effect on blood pressure</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>Limiting Plasma/Effect-Site Concentrations</td>
<td>51</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Simulation results</td>
<td>53</td>
</tr>
<tr>
<td>4.4</td>
<td>Extending The Safety System</td>
<td>56</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Simulation results</td>
<td>59</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusion</td>
<td>68</td>
</tr>
<tr>
<td>5</td>
<td>Case Study II - Model-Invariant Blood Pressure Safety</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>System for Closed-Loop Anesthesia</td>
<td></td>
</tr>
<tr>
<td>5.0.1</td>
<td>Assumptions</td>
<td>71</td>
</tr>
<tr>
<td>5.1</td>
<td>Blood-Pressure Model Falsification in Propofol Anesthesia</td>
<td>72</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Blood-pressure model falsification</td>
<td>73</td>
</tr>
</tbody>
</table>
5.2 Falsified Robust Safety-Preserving Control of Propofol Anesthesia 75
  5.2.1 Simulation results 76
  5.3 Conclusion 83

6 Conclusion and Future Work 86
  6.1 Conclusions 86
  6.2 Future Work 87

Bibliography 89
List of Tables

Table 4.1  Closed-loop anesthesia induction time in the presence of safety systems  .................................................. 64
List of Figures

Figure 2.1  Computation time comparison: model-invariant viability kernel vs. its under-approximation ........................ 27
Figure 2.2  Model-invariant viability kernel vs. its under-approximation . 28
Figure 2.3  Evolution of the model-invariant viability kernel ........ 29

Figure 3.1  Conservatism comparison: output-feedback safety-preserving control vs. feedback invariant ...................... 41
Figure 3.2  Trajectory comparison: output-feedback safety-preserving control vs. feedback invariant .......................... 42
Figure 3.3  Control input: output-feedback safety-preserving control . . 42

Figure 4.1  Block diagram of closed-loop anesthesia with formalized safety system .................................................. 52
Figure 4.2  PKPD model viability kernel .......................... 54
Figure 4.3  Closed-loop anesthesia safety system: Scenario 1 ........ 55
Figure 4.4  Closed-loop anesthesia safety system: Scenario 2 ........ 56
Figure 4.5  Closed-loop anesthesia safety system: Scenario 3 ........ 57
Figure 4.6  Closed-loop anesthesia: clinical example ................. 58
Figure 4.7  Block diagram of closed-loop anesthesia with extended formalized safety system ................................. 61
Figure 4.8  Closed-loop anesthesia simulation results: unconstrained response .......................................................... 62
Figure 4.9  Closed-loop anesthesia simulation results: with formalized safety system ................................. 63
Figure 4.10  Closed-loop anesthesia simulation results: with extended for-
malized safety system ............................................. 65
Figure 4.11  Disturbance profile used in prolonged stimulation scenario . . . 66
Figure 4.12  Closed-loop anesthesia extended safety system: prolonged stim-
ulation scenario ......................................................... 67

Figure 5.1  Blood pressure model falsification: clinical example .............. 74
Figure 5.2  Model-invariant viability kernel: BP PKPD model set .......... 77
Figure 5.3  Block diagram of closed-loop anesthesia with individualized
BP safety system ......................................................... 77
Figure 5.4  Effect of different BP constraints on DoH .............................. 78
Figure 5.5  Closed-loop anesthesia simulation results: with individualized
BP safety system ......................................................... 79
Figure 5.6  Block diagram of closed-loop anesthesia with model-invariant
BP safety system ......................................................... 81
Figure 5.7  Closed-loop anesthesia simulation results: with model-invariant
BP safety system ......................................................... 82
Figure 5.8  Block diagram of closed-loop anesthesia with falsified BP model-
variant safety system .................................................. 83
Figure 5.9  Closed-loop anesthesia simulation results: with falsified model-
variant BP safety system .......................................... 84
Figure 5.10  Conservatism comparison: model-invariant vs. falsified model-
invariant ................................................................. 85
Acknowledgements

I would like to express my deepest appreciation to Prof. Guy Dumont for his incredible supervision and mentorship over the past six years, from the day I started my Master’s to the last day of my PhD program. He has trained me to be a leading researcher yet kind, responsible and helpful to my fellow researchers. I would like to show my sincere gratitude to Prof. Mark Ansermino for his invaluable mentorship. His feedback has always been motivating for me to be a better researcher, speaker and writer. My appreciation extends to Prof. Ian Mitchell for his encouraging ideas and insightful feedback. I have had a wonderful collaboration with Dr. Klaske van Heusden. I will always appreciate the time and efforts she has dedicated to my research throughout the process. I couldn’t finish the program in four years if it was not her professional and critical feedback and comments.

Last but not least, I thank my parents for their never-ending love, support and constant encouragement over the years. I would like to dedicate this thesis to my dearest friend and partner, Nazanin. She has been with me in happiness and sadness since I was an undergraduate student and helping me to grow and become who I am now. Thank you, Nazanin!
To Nazanin.
Chapter 1

Introduction

1.1 Motivation

Although closed-loop control and automation have had massive impacts on almost all industries, from pulp and paper industry to aerospace and automotive industries, control theory has yet to emerge in clinical environments. In current clinical practice, most tasks concerning health care are conducted manually by clinicians. Nevertheless, control theory is capable of revolutionizing health care by bringing automation to clinical environments in the form of medical devices.

The benefits of feedback control over manual control are known to engineers and there is no need for more emphasis. Over the past few years, engineers and clinicians have put considerable efforts in bringing the benefits of feedback control into medical devices to improve patient care in hospitals [9–12]. Led by researchers at the University of Padova and the University of Virginia, the artificial pancreas project is a prime example of using feedback control in medical devices. This technology which is designed for patients diagnosed with type-1 diabetes, conducts automated administration of insulin based on feedback of blood glucose [13, 14]. Closed-loop anesthesia is another example of using control theory to improve patient care in clinical practice. Closed-loop anesthesia systems perform automated administration of anesthetic and analgesic drugs based on feedback from a measure of clinical effects [15–18].

Mainly due to safety concerns, these devices have not been available to clini-
cians. To obtain regulatory approval and to gain clinicians’ trust to use these technologies in clinical practice their safety must be guaranteed and demonstrated. To verify the safety of medical devices and to evaluate the performance of safety systems for medical technologies, realistic clinical scenarios in normal and faulty conditions are commonly employed in simulation [19–21]. While useful, this process is time consuming and costly. Moreover, this verification approach is based on a finite number of scenarios. While a system may meet the safety requirements in the simulated scenarios, other medical scenarios may lead to unsafe situations. Therefore, verification based on simulation scenarios does not provide formal proof of safety. Motivated by closed-loop anesthesia, this work aims to provide a basic theoretical framework for verifying and guaranteeing safety of medical devices without the need for simulation scenarios. Nevertheless, the theoretical contributions of this work are generic and not limited to medical control systems.

1.2 Formal Verification and Safety-Preserving Control

Unlike Verification using simulation scenarios, formal model verification techniques are capable of verifying safe operation for all possible states and inputs and providing a formal proof of safety [22, 23]. These techniques study the evolution of the states of a dynamical system under given conditions (e.g. constraints, perturbations, uncertainty), and address questions concerning the system’s behaviour, including reachability and safety. For a safety-critical control system, safety is verified and guaranteed if we can show that the states of the system never violate a set of safety constraints. There are two common approaches to study safety in safety-critical control systems using formal methods: 1) to investigate the existence of a control input which preserves safety, 2) to investigate if a given feedback controller maintains safety.

Safety-preserving control techniques (approach 1) formulate a control policy which keeps the system states within the safe region. These techniques rely on approximation of the viability kernel [23], which is a set of initial states for which there exists a control action that keeps trajectories of a system starting from those states within the safe region. The viability kernel has traditionally been approximated using Eulerian methods [24] and level set approaches [25, 26]. These tech-
Techniques require gridding the state space, causing the computational complexity to increase as the state dimension increases \cite{27}. In contrast, Lagrangian techniques have been successfully tested for viability kernel approximation of high-dimension systems \cite{28-30}.

A controller is called safety-preserving if it generates a viable trajectory. If the viability kernel is empty, no controller can provide safety. However, if the set is not empty, one needs to synthesize a controller to preserve safety. Lygeros et al. in a series of papers \cite{26, 31, 32} as well as Mitchell et al. in \cite{25} used optimal control formulations based on Hamilton-Jacobi equations to synthesize a controller which satisfies safety specifications. Girard \cite{33} used approximate bisimulation to design a safety-preserving controller. He showed that a controller which preserves safety of an approximately bisimilar abstraction\textsuperscript{1} of a system also maintains safety of the original system. Kaynama et al. in \cite{34} combined a safety-preserving control law with an arbitrary controller (performance controller). This hybrid scheme is capable of satisfying performance criteria while preserving the system safety.

To address safety using approach 2, one needs to approximate the feedback invariant for a given controller. The feedback invariant is a set of initial conditions for which the controller maintains trajectories starting from those states within the safe region. Raković et al. \cite{35-37} used this approach to address safety in model predictive control. Artstein et al. \cite{38} employed invariant sets to verify safety in feedback systems with a given controller. Lesser and Abate \cite{39} employed the feedback invariant to verify safety of an output-feedback safety system.

The techniques mentioned can be interpreted stochastically as well \cite{40-44}. In this framework, a control action is called safety-preserving if it maximizes the probability that the corresponding trajectory starting from the safe region remains within the set. This approach, which is mostly used for formal verification of hybrid stochastic systems \cite{45}, provides a probabilistic proof of safety. To provide a probabilistic proof of safety, a stochastic process model is required which is not always available. For instance, models reported in the literature for automated drug delivery systems are mostly deterministic \cite{46-48}.

\textsuperscript{1}An approximately bisimilar abstraction of a system is a simple model (bisimilar abstraction) of the system, used for control synthesis purposes.
1.3 Objectives and Contributions

As mentioned previously, this work is motivated by closed-loop anesthesia. Despite a number of publications on formal verification of control systems, these methods cannot be directly employed to verify safety of closed-loop anesthesia and design formalized safety systems for this technology. The safety verification and safety-preserving control techniques mentioned previously, have been developed based on implicit assumptions that 1) a system’s model is known and there is no parametric model uncertainty, 2) system’s states are fully measurable. The existing techniques cannot cope with model uncertainty, but solutions have been proposed for systems with additive state uncertainty (e.g. [34, 49]). Although model uncertainty can be represented in the form of additive state uncertainty, applying the existing methods to this formulation may result in infeasible or very conservative solutions. Moreover, discussion of safety when the states cannot be measured, is limited to a small number of publications, e.g. [39, 50] (the limitations will be discussed in the next chapters).

In the closed-loop anesthesia problem, neither of the above-mentioned assumptions holds. The existing models which describe the effects of anesthetics on patients (e.g. [46–48]) are uncertain due to inter-patient variability. These models are population-based and represent an average response of patients to anesthetics. Identifying an individualized model for each patient prior to the closed-loop administration of anesthetics is not practical, either. In addition, the states of these models which indicate patient safety (e.g. plasma/effect-site concentrations), are not measurable. These challenges prevent us from using existing safety verification techniques in the case of closed-loop anesthesia.

The main objective of this work is to develop a mathematical framework for formal verification of closed-loop anesthesia and other safety-critical applications which suffer from model uncertainty. In this work, we propose a model-invariant safety-preserving control technique which extends the existing safety-preserving control techniques to cases with multiplicative model uncertainty. Given a multi-model description of model uncertainty, we synthesize a safety-preserving controller which maintains the states of all members of the model set within a set of viability constraints. We call such controllers model-invariant safety-preserving.
The second contribution of this work is to introduce a safety-preserving control scheme for systems in which the safety-critical states are not measurable. Called output-feedback safety-preserving, this control strategy guarantees to maintain the estimated states of an output-feedback control system within a constraint set which is eroded by the evolution of the estimation error dynamics. We show that this maintains the system’s states within the actual safe set. We also extend the proposed output-feedback safety-preserving control technique to cases with model uncertainty.

The last major contribution of this work is to employ the developed mathematical frameworks to formalize safety systems for closed-loop anesthesia. First, we formalize the safety system proposed by van Heusden et al. [19]. This safety system specifies constraints on predicted propofol concentrations in the plasma and effect-site based on the therapeutic window of propofol. This safety system minimizes the risk of drug over/under-dosing. We also formalize the extension of the mentioned safety system which includes constraints on blood pressure [51]. The mentioned safety systems limit the patient states which are predicted using population-based models in an open-loop fashion. Once these safety systems are formalized, the predicted states are guaranteed to be within the safety constraints during operation; however, there is no guarantee that the actual states of patients remain safe due to inter-patient variability. In the last step, we propose and formalize a model-invariant safety system for closed-loop anesthesia which guarantees safety despite model uncertainty. Due to the fact that this safety system considers inter-patient variability, it introduces conservatism. We employ model falsification to reduce model uncertainty. We show that this reduces conservatism of the model-invariant safety system.

To summarize, the main contributions of this work are:

- Extending existing safety-preserving control techniques and formal methods to cases with multiplicative model uncertainty.
- Extending existing safety-preserving control techniques and formal methods to uncertain output-feedback control systems.

\(^2\)Propofol is an anesthetic agent commonly used in intravenous anesthesia.
• Formalizing existing safety systems for closed-loop anesthesia.

• Formalizing a model-invariant safety system for closed-loop anesthesia with reduced conservatism using model falsification.

The following papers have been written out of the contents of this thesis:

• Refereed journal papers:


• Refereed conference papers:


1.4 Organization of The Thesis

This thesis is outlined as follows. Chapter 2 provides a summary of the existing safety-preserving control techniques and discusses their limitations. In this chapter, we introduce our model-invariant safety-preserving control technique which extends the existing methods to cases with model-uncertainty. In Chapter 3, we propose the output-feedback safety-preserving control technique which provides a guarantee of safety for output-feedback systems by maintaining the estimated states within a contracted safe region. Moreover, we discuss the output-feedback safety-preserving control technique in the presence of model uncertainty. Chapter 4 formalizes the existing safety systems in the literature for closed-loop anesthesia. In Chapter 5, we propose and formalize a model-invariant safety system for closed-loop anesthesia based on the results of Chapter 2. Moreover, this chapter proposes a novel approach to reduce conservatism of the model-invariant safety system by reducing model uncertainty using model falsification. Finally, Chapter 6 concludes the thesis.
Chapter 2

Model-Invariant
Safety-Preserving Control

Employing safety-preserving control techniques to guarantee safety of safety critical control systems has mainly been discussed for systems with known dynamics, and discussion of safety for uncertain systems is limited to systems with additive state uncertainty. Although parametric model uncertainty can be represented in the form of additive state uncertainty, a feasible solution may not exist or it might be very conservative depending on the size of uncertainty. This chapter introduces model-invariant safety-preserving control that can be used to guarantee safety of systems with a certain class of multiplicative uncertainty. The model-invariant safety-preserving control technique, on the other hand, provides safe control for systems for which other safety-preserving methods cannot provide feasible solutions.

The main contributions in this chapter are as follows:

- Introducing the model-invariant safety-preserving control scheme: this technique extends the existing safety-preserving control schemes to cases with model uncertainty.

- Proposing a computationally-efficient method to under-approximate the model-invariant viability kernel: the model-invariant safety-preserving control technique relies on the model-invariant viability kernel whose computational
complexity is high. We propose a recursive formula to under-approximate the model-invariant viability kernel with a significantly reduced computational burden.

- Proposing an optimization problem to minimize conservatism of the model-invariant viability kernel under-approximation.

## 2.1 Preliminaries

### 2.1.1 Notations

The Minkowski sum of any two non-empty convex sets \( \mathcal{P} \subset \mathbb{R}^n \) and \( \mathcal{Q} \subset \mathbb{R}^n \) is

\[
\mathcal{P} \oplus \mathcal{Q} := \{ p + q \mid p \in \mathcal{P}, q \in \mathcal{Q} \}.
\]

Their Minkowski difference (the erosion of \( \mathcal{P} \) by \( \mathcal{Q} \)) is

\[
\mathcal{P} \ominus \mathcal{Q} := \{ p \in \mathcal{P} \mid \forall q \in \mathcal{Q}, p + q \in \mathcal{P} \}.
\]

For vector \( x \in \mathbb{R}^n \), \( \| \cdot \|_2 : \mathbb{R}^n \to \mathbb{R} \) is the vector 2-norm of \( x \). For matrix \( A \in \mathbb{R}^{n \times n} \), \( \| \cdot \|_2 : \mathbb{R}^{n \times n} \to \mathbb{R} \) is the induced 2-norm, which is equivalent to the maximum singular value of \( A \) (\( \| A \|_2 = \bar{\sigma}(A) \)). The set

\[
\mathcal{B}(\kappa) := \{ x \in \mathbb{R}^n \mid \| x \|_2 \leq \kappa \}
\]

denotes the closed 2-norm ball of radius \( \kappa > 0 \).

**Lemma 2.1.1.** For any two non-empty convex sets \( \mathcal{P} \subset \mathbb{R}^n \) and \( \mathcal{Q} \subset \mathbb{R}^n \), we have

\[
(\mathcal{P} \ominus \mathcal{Q}) \oplus \mathcal{Q} \subseteq \mathcal{P},
\]

\[
\mathcal{P} \ominus (\mathcal{P} \oplus \mathcal{Q}) \subseteq \mathcal{P} \cap \mathcal{Q}.
\]

**Lemma 2.1.2.** For any set defined as \( \mathcal{I} = \mathcal{B}(\kappa) \subset \mathbb{R}^n \), we have

\[
\bar{\sigma}(H) \mathcal{I} \subseteq H \mathcal{I} \subseteq \bar{\sigma}(H) \mathcal{I}, \quad H \in \mathbb{R}^{n \times n}.
\]
In above, $\sigma(H)$ denotes the minimum singular value of matrix $H$.

2.1.2 Safety-preserving control

Consider the following linear time-invariant dynamics:

$$X : \dot{x}(t) = Ax(t) + Bu(t).$$

(2.7)

In the above equation, $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ are respectively the state and input vectors. The matrices $A$ and $B$ are in $\mathbb{R}^{n \times n}$ and $\mathbb{R}^{n \times m}$. The time $t$ ranges throughout a continuous time domain $T = [0, \tau]$ with a finite horizon ($0 < \tau < \infty$). The input is constrained by a non-empty, compact and convex subset of $\mathbb{R}^m$ ($u(t) \in \mathcal{U} \subset \mathbb{R}^m$). Viability theory is concerned with ensuring that the system’s states remain within a set of viability constraints ($x(t) \in \mathcal{X} \subset \mathbb{R}^n$) over the course of operation. The viability constraint set $\mathcal{X}$ is a convex compact set. Accordingly, the safety is guaranteed if there exists $u(t) \in \mathcal{U}$ such that the states of $X$ remain within $\mathcal{X}$ for all $t \in T$. Viability theory provides a guarantee of safety through the use of the viability kernel [23].

**Definition 1** (Viability kernel). The finite-horizon viability kernel of the viability constraint set $\mathcal{X}$ for the system $X$ is a subset of $\mathcal{X}$ starting from which there exists a constrained input ($u(t) \in \mathcal{U}$) that maintains the states of $X$ inside $\mathcal{X}$ over a finite time interval ($\forall t \in T$):

$$\text{Viab}_T(\mathcal{X}, \mathcal{U}, X) = \{x_0 \in \mathcal{X} | x(0) = x_0, \exists u(\cdot) : T \to \mathcal{U} \text{ s.t. } \forall t \in T, x(t) \in \mathcal{X}\}.$$  

(2.8)

Accordingly, we can define $\text{Viab}_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)$ as a set of states at time $t$ starting from which there exists safety-preserving control over the period $[t, \tau]$.

**Definition 2** (Safety-preserving control). An admissible input signal $u(\cdot) : T \to \mathcal{U}$ is called safety-preserving if it maintains the states of $X$ within $\mathcal{X}$ for all $t$ in $T$.

For a given system with a given constraint set, if the viability kernel is empty, no control action exists to preserve safety. Thereby, safety cannot be guaranteed. On the other hand, if the set is non-empty, one can find a safety-preserving control action that maintains the safety of the system.
2.1.3 Control synthesis

Safety-preserving controllers are capable of maintaining trajectories of systems within the safe region of the state space. The control synthesis problem, finding a safety-preserving controller, has received significant attention from researchers. A classical approach to synthesize a safety-preserving controller is to use the information of the shape of the computed viability kernel. Accordingly, a safety-preserving control policy is formulated based on the proximal normal at every point of the viability kernel boundaries [53, 54]. Kurzhanski et al. [55] approached the safety-preserving control synthesis problem through set-valued functions as well as dynamic programming techniques. Lygeros et al. [31, 32] used optimal control formulations to synthesize a safety-preserving controller. Mitchell et al. [56] employed the Hamilton-Jacobi partial differential equation to synthesize a controller for sampled data systems. The viability kernel can also be used as a terminal constraint of the receding horizon optimization problem in an MPC framework [36, 57–59]. Due to the fact that the feasibility of the optimization is guaranteed with a non-empty viability kernel, the controller can be considered safety-preserving.

Here, we employ the hybrid scheme proposed by Kaynama et al. [34] that meets closed-loop performance criteria while preserving safety. Kaynama et al. [34] combined a closed-loop controller (to meet desired closed-loop criteria) with a safety-preserving controller (to guarantee safety) as follows:

\[
  u(t) = \begin{cases} 
  u_{pr}(t), & x(t) \in \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X); \\
  u_{sp}(t), & x(t) \notin \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X). 
\end{cases}
\] (2.9)

In the above equation, \( \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X) \) refers to the interior of \( \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X) \), \( u_{pr}(\cdot) \) is an input calculated by a closed-loop controller, \( u_{sp}(\cdot) \) is an input provided by a safety-preserving controller. In this work, we utilize the safety-preserving control policy formulated by Kurzhanski and Valyi [55]. According to [55], assuming the viability kernel and the input constraint set are compact, convex and
The h-continuous following control policy is safety-preserving:

\[
    u_{sp}(t) = \arg\min_{u(t)} \{< l^0(x(t), Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)), Bu(t) > \mid u(t) \in \mathcal{U} \}, \tag{2.10}
\]

where

\[
    l^0(x(t), Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)) = \arg\max_{\|l\| \leq 1} \{< l, x(t) > - \rho(l, Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)) \}. \tag{2.11}
\]

In the above equations, \(< \cdot, \cdot > \) denotes the inner product of two vectors. In (2.10), \(B\) is the gain matrix of \(X\). In (2.11), \(\rho(l, Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X))\) is the support function of \(Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)\) in direction \(l\):

\[
    \rho(l, Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)) = \max\{< l, z > \mid z \in Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X) \}. \tag{2.12}
\]

This control policy assures that the states of \(X\) remain within \(\mathcal{X}\). To prevent chattering in the hybrid control input (2.9), Kaynama et al. [34] suggested to use a convex combination of \(u_{pr}(t)\) and \(u_{sp}(t)\):

\[
    u(t) = (1 - \beta_\alpha)u_{pr}(t) + \beta_\alpha u_{sp}(t), \tag{2.13}
\]

where \(\beta_\alpha\) is calculated as

\[
    \beta_\alpha = \begin{cases} 
        0 & , \alpha < d(x(t)); \\
        \frac{\alpha - d(x(t))}{\alpha} & , 0 < d(x(t)) \leq \alpha; \\
        1 & , d(x(t)) \leq 0.
    \end{cases} \tag{2.14}
\]

In the above equation, \(\alpha \in [0, 1]\) is a design variable and \(d(x(t))\) is the minimum distance between \(x(t)\) (the states of \(X\)) and boundaries of \(Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)\). The distance is negative if \(x(t) \notin Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)\). Accordingly, as the states approach the boundaries of \(Viab_{[t, \tau]}(\mathcal{X}, \mathcal{U}, X)\), the safety-preserving controller contributes more and pushes the states back inside the safe region.

The viability kernel can be calculated off-line. The safety-preserving control

\[\text{1The definition and characteristics of h-continuous functions are given in [60].}\]
policy can be implemented in real-time, using the pre-calculated viability kernel, the input matrix B and feedback from the full system states $x(t)$.

2.1.4 Limitations

The majority of safety-preserving control techniques provide a guarantee of safety based on the following implicit assumptions:

- the state-space model is known and there is no model uncertainty,
- the states are fully measurable,
- there is no measurement noise.

Safety-preserving control is a model-based method which requires an accurate model of a system. Safety-preserving control of linear systems with uncertain models and formal verification of such systems have mostly been discussed in the context of additive state uncertainty of the following form (see [34, 40, 42, 61]):

$$X : \dot{x}(t) = Ax(t) + Bu(t) + d(t),$$

(2.15)

where $d(t) \in \mathcal{D}$ is the state uncertainty and $\mathcal{D} \subset \mathbb{R}^n$ is a convex and compact set. Although uncertainty in the matrices $A$ and $B$ can be written in the form of additive uncertainty, it may result in an empty viability kernel or very conservative solutions depending on the size of uncertainty. The above-mentioned assumptions limit applications of safety-preserving control as these techniques cannot be employed to address safety of safety-critical systems with parametric model uncertainty and output-feedback control systems.

2.2 Model-Invariant Safety Preserving Control

2.2.1 Problem formulation

Consider an uncertain dynamics of the following form:

$$X_u : \dot{x}(t) = Ax(t) + \alpha Bu(t).$$

(2.16)
Here, \( \tilde{B} \in \mathbb{R}^{n \times m} \) denotes the nominal gain matrix. \( A \in \mathcal{A} \) and \( \alpha \in \mathcal{G} \) are uncertain parameters. \( \mathcal{A} \subset \mathbb{R}^{n \times n} \) and \( \mathcal{G} \subset \mathbb{R}^{1 \times 2} \) are compact sets and we assume the maximum singular value of all members of \( \mathcal{A} \) is bounded. Consider \( \mathcal{M} \) as the set of all instances of \( X_u \):

\[
\mathcal{M} = \{ X | X : \dot{x}(t) = Ax(t) + \alpha \tilde{B}u(t), A \in \mathcal{A}, \alpha \in \mathcal{G} \}. \tag{2.17}
\]

This notation refers to model sets with a finite or infinite number of members, including \( \mathcal{A} \) and/or \( \mathcal{G} \) which are continuous. Model-invariant safety-preserving control provides a control action that keeps the states of any member of \( \mathcal{M} \) within the safe set \( \mathcal{K} \).

**Definition 3** (Model-invariant safety-preserving control). A control action \( u(\cdot) : \mathbb{T} \rightarrow \mathcal{U} \) is model-invariant safety-preserving over \( \mathbb{T} \) if it maintains the states of all members of \( \mathcal{M} \) within the constraint set \( \mathcal{K} \) for all time in \( \mathbb{T} \).

The corresponding model-invariant viability kernel is defined as follows:

**Definition 4** (Model-invariant viability kernel). The finite-horizon model-invariant viability kernel of \( \mathcal{K} \) for the model set \( \mathcal{M} \) is a subset of \( \mathcal{K} \) which includes initial conditions starting from which there exists a constrained control input that maintains the states of all members of \( \mathcal{M} \) inside \( \mathcal{K} \) for all time in \( \mathbb{T} \):

\[
\text{Viab}_{\mathbb{T}}(\mathcal{K}, \mathcal{U}, \mathcal{M}) = \{ x_0 \in \mathcal{K} | x(0) = x_0, \exists u(\cdot) : \mathbb{T} \rightarrow \mathcal{U} \text{ s.t. } \forall X \in \mathcal{M} \text{ & } \forall t \in \mathbb{T}, x(t) \in \mathcal{K} \}. \tag{2.18}
\]

Accordingly, \( \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) \) is defined as a set of states at time \( \tau \) starting from which there exists model-invariant safety-preserving control for the model set \( \mathcal{M} \) over \([t, \tau] \).

\( ^2\mathbb{R}^{\geq 1} \) is a set of real numbers which are greater than 1.
2.2.2 Control synthesis

Let $\mathcal{I}_{[t, \tau]}$ denote the intersection of the viability kernels of all individual set members:

$$\mathcal{I}_{[t, \tau]} = \bigcap_{X \in \mathcal{M}} \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{X}_j).$$  \hspace{1cm} (2.19)

The following theorem formulates a model-invariant viability kernel policy for the model set $\mathcal{M}$:

**Theorem 2.2.1.** Consider the model set $\mathcal{M}$ defined in (2.17), safe region $\mathcal{K}$, and the input constraints $\mathcal{U}$. Assume the intersection of the viability kernels of its individual members $\mathcal{I}_{[t, \tau]}$ is not empty and that $x(t) \in \mathcal{I}_{[t, \tau]}$. Define the following control policy

$$u(t) = \arg \min_{u(t) \in \mathcal{U}} \left\{ l^0 \left( x(t), \mathcal{I}_{[t, \tau]} \right), Bu(t) > \right\},$$  \hspace{1cm} (2.20)

where $l^0 \left( x(t), \mathcal{I}_{[t, \tau]} \right)$ is defined as (2.11). Then the control policy $u(t)$ is model-invariant safety-preserving.

**Proof.** For $X_j \in \mathcal{M}$,

$$X_j: \ \dot{x}(t) = A_j x(t) + \alpha_j Bu(t), \ A_j \in \mathcal{A}, \ \alpha_j \in \mathcal{G},$$  \hspace{1cm} (2.21)

let define

$$V(t) = \text{Dist}^2 \left( x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) \right),$$  \hspace{1cm} (2.22)

where $\text{Dist} \left( x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) \right)$ is the Hausdorff distance measuring the dis-
tance of $x(t)$ from $\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)$ \cite{55}:

$$
\text{Dist} \left( x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) \right) = \\
\min \{ \| x(t) - v \|_2 | v \in \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) \} = \\
\max \{ < l, x(t) > - \rho (l | \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) | \| l \|_2 \leq 1 \} = \\
< I^0 (x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)), x(t) > - \\
\rho (I^0 (x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)) | \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)). 
$$

\begin{align*}
\text{(2.23)}
\end{align*}

Kurzhanski et al. \cite{55} showed that assuming $\frac{d}{dt} V(t)$ exists, starting from any point in $\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)$ the following control policy is safety preserving over $[t, \tau]$ for $X_j$:

$$
u(t) = \arg \min_{u} \left\{ \frac{d}{dt} V(t) \mid u \in \mathcal{U} \right\}. 
$$

\begin{align*}
\text{(2.24)}
\end{align*}

If there is no model uncertainty, this control policy is simplified to (2.10). Due to the fact that $\text{Dist} \left( x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) \right) \geq 0$, (2.24) can be expressed as

$$
u(t) = \arg \min_{u(t) \in \mathcal{U}} \left\{ \frac{d}{dt} \text{Dist} \left( x(t), \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j) \right) \right\}. 
$$

\begin{align*}
\text{(2.25)}
\end{align*}

Since $\mathcal{S}_{[t, \tau]} \subseteq \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)$, we can use $\mathcal{S}_{[t, \tau]}$ in (2.25) as an under approximation of $\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)$ and rewrite the safety-preserving control policy (2.25) as:

$$
u(t) = \arg \min_{u \in \mathcal{U}} \left\{ \frac{d}{dt} \text{Dist} \left( x(t), \mathcal{S}_{[t, \tau]} \right) \right\}. 
$$

\begin{align*}
\text{(2.26)}
\end{align*}

Accordingly, starting from any point in $\mathcal{S}_{[t, \tau]}$, (2.26) is safety-preserving over $[t, \tau]$ for $X_j$. The derivative of the Hausdorff distance can be simplified as

$$
\frac{d}{dt} \text{Dist} \left( x(t), \mathcal{S}_{[t, \tau]} \right) = < I^0 (x(t), \mathcal{S}_{[t, \tau]}), \dot{x}(t) > \\
- \frac{\partial}{\partial t} \rho (I^0 (x(t), \mathcal{S}_{[t, \tau]} ) | \mathcal{S}_{[t, \tau]}). 
$$

\begin{align*}
\text{(2.27)}
\end{align*}

According to \cite{55} and due to the fact $\mathcal{S}_{[t, \tau]} \subseteq \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j)$, the partial
The derivative of the support function can be written as
\[
\frac{\partial}{\partial t} \rho(l^0(x(t), J_{t, \pi})) = \langle l^0(x(t), J_{t, \pi}), Ax(t) \rangle + \rho(l^0(x(t), J_{t, \pi}) | \alpha_j \bar{B} \mathcal{U}).
\] (2.28)

Substituting (2.28) in (2.27) yields
\[
\frac{d}{dt} \text{Dist} \left( x(t), J_{t, \pi} \right) = \\
\langle l^0(x(t), J_{t, \pi}), Ax(t) + \alpha_j \bar{B}u(t) \rangle - \\
\langle l^0(x(t), J_{t, \pi}), Ax(t) - \rho(l^0(x(t), J_{t, \pi}) | \alpha_j \bar{B} \mathcal{U}) \rangle = \\
\langle l^0(x(t), J_{t, \pi}), \alpha_j \bar{B}u(t) \rangle - \rho(l^0(x(t), J_{t, \pi}) | \alpha_j \bar{B} \mathcal{U}).
\] (2.29)

Consequently, we can express (2.26) as
\[
u(t) = \arg \min_{\nu(t) \in \mathcal{U}} \{ \langle l^0(x(t), J_{t, \pi}), \alpha_j \bar{B}u(t) \rangle \}.\] (2.30)

Due to the fact that \( \alpha_j > 0 \) and (2.30) is convex, the minimizer of (2.30) is independent of \( \alpha_j \). Thus, we can simplify (2.30) to the following optimization problem:
\[
u(t) = \arg \min_{u(t) \in \mathcal{U}} \{ \langle l^0(x(t), J_{t, \pi}), \bar{B}u(t) \rangle \}.\] (2.31)

Since \( J_{t, \pi} \subseteq \text{Viab}_{t, \pi}(\mathcal{X}, \mathcal{U}, X_j) \) for all \( X_j \in \mathcal{M} \), (2.28) holds for all members of \( \mathcal{M} \) and (2.26) is simplified to (2.31) for all \( X_j \in \mathcal{M} \). Thus, (2.31) maintains the states of all members of \( \mathcal{M} \) over \([0, \tau]\) within \( \mathcal{K} \). Consequently, according to Definition 3, (2.31) is model-invariant safety preserving.

It follows by definition that the intersection is a subset of the model-invariant viability kernel:
\[
J_{t, \pi} \subseteq \text{Viab}_{t, \pi}(\mathcal{X}, \mathcal{U}, \mathcal{M}).
\] (2.32)

Note that \( \bar{B} \) in (2.17) is known, \( J_{t, \pi} \) can be calculated offline, the model-invariant safety-preserving control policy can be implemented in real-time using feedback.
from the measured states $x(t)$, and the same model-invariant safety preserving controller can be used for any system $X_j \in \mathcal{M}$.

**Corollary 2.2.2.** The model-invariant viability kernel for the model set $\mathcal{M}$ defined in (2.17) is equivalent to the intersection of the viability kernels of all members of the model set:

$$\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) = \mathcal{I}_{[t, \tau]}.$$  \hspace{1cm} (2.33)

**Proof.** According to (2.32), to show (2.33) holds, it is sufficient to show that

$$\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) \subseteq \mathcal{I}_{[t, \tau]}.$$  \hspace{1cm} (2.34)

Suppose (2.34) does not hold:

$$\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) \nsubseteq \mathcal{I}_{[t, \tau]}.$$  \hspace{1cm} (2.35)

Accordingly,

$$\exists x' \in \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) \text{ s.t. } x' \notin \mathcal{I}_{[t, \tau]}.$$  \hspace{1cm} (2.36)

Due to the fact that $\mathcal{I}_{[t, \tau]} = \bigcap_{X_i \in \mathcal{M}} \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_i)$, (2.36) implies that

$$\exists X_j \in \mathcal{M} \text{ s.t. } x' \notin \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, X_j).$$  \hspace{1cm} (2.37)

Therefore, starting from $x'$ there is no safety-preserving control action to preserve safety of $X_j$ over $[t, \tau]$. According to the definition of the model-invariant viability kernel (Definition 4), $x'$ cannot be a member of $\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M})$. This means

$$\forall x' \in \text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) \text{ s.t. } x' \notin \mathcal{I}_{[t, \tau]}.$$  \hspace{1cm} (2.38)

Equation (2.38) contradicts the assumption that (2.34) does not hold. Therefore, (2.34) holds by contradiction. Comparing (2.32) and (2.34) yields

$$\text{Viab}_{[t, \tau]}(\mathcal{K}, \mathcal{U}, \mathcal{M}) = \mathcal{I}_{[t, \tau]}.$$  \hspace{1cm} (2.39)
2.3 Model-Invariant Viability Kernel Under-Approximation

We showed that for the model set defined in (2.17), the intersection of the viability kernels of the model set members corresponds to the model-invariant viability kernel of the uncertain system. However, as the number of set members increases, approximating the viability kernels of all members of the model set becomes computationally expensive. Moreover, computing the model-invariant viability kernel is infeasible in a case of infinite or continuous model sets. In this section, we propose an efficient approach to under-approximate the model-invariant viability kernel assuming the constraints are given in the form of 2-norm balls.

2.3.1 Viability kernel approximation

This section briefly describes a recursive formulation used in Lagrangian algorithms to compute the viability kernel. In the next section, we will employ this formulation to under-approximate the model-invariant viability kernel. Lagrangian algorithms rely on geometric representations of the constraint sets such as polytopes [29], ellipsoids [28, 62], zonotopes [63]. Since these methods do not require state-space gridding, as required by other methods such as Eulerian methods [25], Lagrangian methods are computationally feasible to analyze systems with high dimensions. Maidens et al. in [27] summarized some of the Lagrangian algorithms for computing the viability kernel and discussed their differences. Viability kernel calculation for both continuous and discrete time systems requires time discretization. In this section, we only focus on the viability kernel calculation of discrete time systems; however, the results can be extended to continuous time systems with minor modifications [27].

In Lagrangian methods, the viability kernel is approximated using recursive calculations of backward reachable sets.

Definition 5 (Backward reachable set). The backward reachable set of the system \( X \) from a set \( \mathcal{H} \subset \mathbb{R}^n \) over the time domain \( \mathbb{T} \) with finite horizon \( \tau \) is a set of initial
conditions starting from which there exists a constrained input $u(t) \in \mathcal{U}$ such that the states of $X$ reach $\mathcal{H}$ at time $t = \tau$:

$$
\text{Reach}_\tau(\mathcal{H}, \mathcal{U}, X) = \{ x_0 \in \mathbb{R}^n | x(0) = x_0, \exists u(\cdot) : \mathbb{T} \to \mathcal{U} \text{ s.t. } x(\tau) \in \mathcal{H} \}.
$$

(2.40)

Consider the following discrete-time system,

$$
X : \quad x(t+1) = Ax(t) + Bu(t),
$$

(2.41)

where the time $t$ ranges a discrete time domain $\mathbb{T} = [0, \tau] \cap \mathbb{Z}^+$. For the sake of simplicity, we assume all time intervals given in the rest of this section are discrete-time intervals. Consider

$$
\mathcal{K}_i := \text{Viab}_{[i-1,i]}(\mathcal{H}, \mathcal{U}, X), \quad i \in [0, \tau].
$$

(2.42)

Maidens et al. [27] reformulated Saint-Pierre’s recursive formula [24] and approximated the viability kernel in terms of the backward reachable set over one discrete time step.

**Theorem 2.3.1.** The sequence of finite horizon viability kernels $\mathcal{K}_i$ can be computed recursively in terms of reachable sets as

$$
\begin{cases}
\mathcal{K}_0 = \mathcal{H}, \\
\mathcal{K}_{i+1} = \mathcal{K}_0 \cap \text{Reach}_1(\mathcal{K}_i, \mathcal{U}, X).
\end{cases}
$$

(2.43)

**Proof.** See [27].

The reachable set over a single time step can be calculated as follows [27]:

$$
\text{Reach}_1(\mathcal{H}, \mathcal{U}, X) = A^{-1}(\mathcal{H} \oplus B^- \mathcal{U}),
$$

(2.44)

where $B^- = -B$. In the above equation $A^{-1}(\cdot)$ is the preimage of a set under the map $A : \mathbb{R}^n \to \mathbb{R}^n$. In the rest of this paper, we will assume that the matrix $A$ is nonsingular; therefore, the preimage of $A$ can be calculated by applying the matrix inverse $A^{-1}$ to the set $\mathcal{H} \oplus B^- \mathcal{U}$. 


2.3.2 Model-invariant viability kernel under-approximation

Here, we find a subset of the model-invariant viability kernel without the need for calculating the viability kernel of individual set members. To do so, we will calculate a common subset of the viability kernels of the entire model set. Then we show that this set is a subset of the model-invariant viability kernel.

Consider the following model set which represents model-uncertainty in a discrete time system:

\[ M = \{ X | X : x(t + 1) = Ax(t) + \alpha \tilde{B}u(t), A \in \mathcal{A}, \alpha \in \mathcal{G} \}. \]  

(2.45)

Similar to (2.17), \( \mathcal{A} \subset \mathbb{R}^{n \times n} \) and \( \mathcal{G} \subset \mathbb{R}_{\geq 1} \) are compact sets and we assume the maximum singular values of all members of \( \mathcal{A} \) are bounded. Here, we assume that the constraints limit the 2-norm of \( x(t) \) and \( u(t) \):

\[ K = B(l), \quad l \in \mathbb{R}, \]  

(2.46)

\[ U = B(g), \quad g \in \mathbb{R}. \]  

(2.47)

According to (2.44), we can calculate the reachable set over a single time step for any member of the model set \( M \) as follows:

\[ \text{Reach}_1(\mathcal{K}, \mathcal{U}, X) = A^{-1}(\mathcal{K} \oplus \alpha \tilde{B}^{-1} \mathcal{U}), \quad X \in M, \]  

(2.48)

where \( \tilde{B}^{-} = -\tilde{B} \). We aim to find a subset of (2.48) which is invariant to the members of \( M \). Consider \( B^\dagger(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) as an operator which under-approximates a given set by a 2-norm ball:

\[ B^\dagger(\mathcal{K}) \subseteq \mathcal{K}, \quad \mathcal{K} \subset \mathbb{R}^n. \]  

(2.49)
For any member of $\mathcal{M}$, we can under-approximate (2.48) as follows:

$$\bar{B}^- \subseteq \alpha \bar{B}^- \cup \mathcal{H}, \quad \forall \alpha \in \mathcal{A}, \quad (\alpha \geq 1)$$

$$\Rightarrow \mathcal{H} \oplus \bar{B}^- \subseteq \mathcal{H} \oplus \alpha \bar{B}^- \cup \mathcal{H}, \quad \forall \alpha \in \mathcal{A},$$

$$\Rightarrow \mathcal{B}^{-1} (\mathcal{H} \oplus \bar{B}^- \cup \mathcal{H}) \subseteq \mathcal{H} \oplus \alpha \bar{B}^- \cup \mathcal{H}, \quad \forall \alpha \in \mathcal{A},$$

$$\Rightarrow (\bar{\sigma}(A^{-1}))^{-1} \mathcal{B}^{-1} (\mathcal{H} \oplus \bar{B}^- \cup \mathcal{H}) \subseteq A^{-1} (\mathcal{H} \oplus \alpha \bar{B}^- \cup \mathcal{H}), \quad \forall \alpha \in \mathcal{A}, A \in \mathcal{A}, \quad (\text{lemma 2.1.2})$$

$$\Rightarrow (\bar{\sigma}(A))^{-1} \mathcal{B}^{-1} (\mathcal{H} \oplus \bar{B}^- \cup \mathcal{H}) \subseteq A^{-1} (\mathcal{H} \oplus \alpha \bar{B}^- \cup \mathcal{H}), \quad \forall \alpha \in \mathcal{A}, A \in \mathcal{A}. \quad (2.50)$$

Consider $\bar{\sigma}(\mathcal{A}) := \max \{ \bar{\sigma}(A) \mid A \in \mathcal{A} \}$. Accordingly,

$$(\bar{\sigma}(\mathcal{A}))^{-1} \mathcal{B}^{-1} (\mathcal{H} \oplus \bar{B}^- \cup \mathcal{H}) \subseteq A^{-1} (\mathcal{H} \oplus \alpha \bar{B}^- \cup \mathcal{H}), \quad \forall \alpha \in \mathcal{A}, \forall A \in \mathcal{A}. \quad (2.51)$$

Define

$$\mathcal{R}_1 := (\bar{\sigma}(\mathcal{A}))^{-1} \mathcal{B}^{-1} (\mathcal{H} \oplus \bar{B}^- \cup \mathcal{H}). \quad (2.52)$$

According to (2.51), for the model set defined in (2.17) with the constraints defined in (2.46) and (2.47) we have:

$$\mathcal{R}_1 \subseteq \text{Reach}_1 (\mathcal{H}, \mathcal{U}, X), \quad \forall X \in \mathcal{M}. \quad (2.53)$$

**Theorem 2.3.2.** For the model set $\mathcal{M}$ defined in (2.17) and the constraints defined in (2.46) and (2.47), the sequence of the viable sets $\mathcal{H}_i^\downarrow$ computed using the formula below is a subset of the model-invariant viability kernel of the model set $\mathcal{M}$:

$$\left\{ \begin{array}{l}
\mathcal{H}_0^\downarrow = \mathcal{H}, \\
\mathcal{R}_i = (\bar{\sigma}(\mathcal{A}))^{-1} \mathcal{B}^{-1} (\mathcal{H}_i^\downarrow \oplus \bar{B}^- \cup \mathcal{H}), \\
\mathcal{H}_{i+1}^\downarrow = \mathcal{H}_0^\downarrow \cap \mathcal{R}_i.
\end{array} \right. \quad (2.54)$$
Proof. Using (2.53) and the fact that $\mathcal{K}_{0} = \mathcal{K}_0 = \mathcal{K}$,

$$
\mathcal{K}_1 \subseteq \text{Reach}_1(\mathcal{K}_0, \mathcal{U}, X), \quad \forall X \in \mathcal{M},
$$

$$
\Rightarrow \mathcal{K}_0^\perp \cap \mathcal{K}_1 \subseteq \mathcal{K}_0 \cap \text{Reach}_1(\mathcal{K}_0, \mathcal{U}, X), \quad \forall X \in \mathcal{M},
$$

$$
\Rightarrow \mathcal{K}_1^\perp \subseteq \mathcal{K}_1, \quad \forall X \in \mathcal{M}. \quad (2.55)
$$

By induction,

$$
\mathcal{K}_t^\perp \subseteq \mathcal{K}_t, \quad \forall X \in \mathcal{M}, \quad (2.56)
$$

According to Theorem 2.3.1, we can write

$$
\mathcal{K}_t^\perp \subseteq \text{Viab}_{[0,t]}(\mathcal{K}, \mathcal{U}, X), \quad \forall X \in \mathcal{M}. \quad (2.57)
$$

According to (2.33), the above equation follows:

$$
\mathcal{K}_t^\perp \subseteq \text{Viab}_{[0,t]}(\mathcal{K}, \mathcal{U}, \mathcal{M}). \quad (2.58)
$$

The recursive formula (2.54) can be used to under-approximate the model-invariant viability kernel of any set as long as $\bar{s}(\mathcal{A})$ is known.

Reducing conservatism

The level of conservatism that the proposed approach introduces is directly related to how much $(\bar{s}(\mathcal{A}))^{-1}$ shrinks $\mathcal{B}^\perp(\mathcal{K} \oplus \mathcal{B}^{-} \mathcal{U})$ in (2.51). This can yield a very conservative, or even an empty solution, depending on the size of the maximum singular value and the number of steps in (2.54). Furthermore, the under-approximation of $\mathcal{K} \oplus \mathcal{B}^{-} \mathcal{U}$ by a 2-norm ball is an additional source of conservatism. However, the assumptions that $\mathcal{K}$ and $\mathcal{U}$ are 2-norm balls decreases conservatism caused by $\mathcal{B}^\perp(\mathcal{K} \oplus \mathcal{B}^{-} \mathcal{U})$.

To minimize conservatism due to the size of $\bar{s}(\mathcal{A})$, we decompose members of $\mathcal{A}$ into a fixed (nominal) part $\bar{A} \in \mathbb{R}^{n \times n}$ and a variable part $\hat{A} \in \mathbb{R}^{n \times n}$ with the
variable part to be minimized:

\[ A = \tilde{A}(I + \tilde{A}), \quad A \in \mathcal{A}. \]  

(2.59)

In the above equation, \( I \in \mathbb{R}^{n \times n} \) denotes the identity matrix. Accordingly, we can define \( \mathcal{A} \) as:

\[ \mathcal{A} = \{ \tilde{A} \in \mathbb{R}^{n \times n} \mid I + \tilde{A} = \tilde{A}^{-1}A, \tilde{A} \in \mathbb{R}^{n \times n}, A \in \mathcal{A} \}. \]  

(2.60)

Here, we assume \( \tilde{A} \) and \( I + \tilde{A} \) are non-singular. According to this definition, if there is no uncertainty, \( \mathcal{A} = \emptyset \) and \( A = \tilde{A} \) for all members of \( \mathcal{A} \). In other words, the set \( \mathcal{A} \) reduces to \( \mathcal{A} = \{ \tilde{A} \} \). Considering \( \bar{\sigma}(I + \tilde{A}) := \max\{\bar{\sigma}(I + \tilde{A}) \mid \tilde{A} \in \mathcal{A} \} \), we can modify (2.50) as follows:

\[ \mathcal{K} \oplus \tilde{B}^{-1} \mathcal{U} \subseteq \mathcal{K} \oplus \alpha \tilde{B}^{-1} \mathcal{U}, \quad \forall \alpha \in \mathcal{G}, \]

\[ \Rightarrow \tilde{A}^{-1}(\mathcal{K} \oplus \tilde{B}^{-1} \mathcal{U}) \subseteq \tilde{A}^{-1}(\mathcal{K} \oplus \alpha \tilde{B}^{-1} \mathcal{U}), \quad \forall \alpha \in \mathcal{G}, \]

\[ \Rightarrow \mathcal{K}_G(\tilde{A}^{-1}(\mathcal{K} \oplus \tilde{B}^{-1} \mathcal{U})) \subseteq \tilde{A}^{-1}(\mathcal{K} \oplus \alpha \tilde{B}^{-1} \mathcal{U}), \quad \forall \alpha \in \mathcal{G}, \]

\[ \Rightarrow \bar{\sigma}(I + \tilde{A})^{-1}\mathcal{K}_G(\tilde{A}^{-1}(\mathcal{K} \oplus \tilde{B}^{-1} \mathcal{U})) \subseteq (I + \tilde{A})^{-1}\tilde{A}^{-1}(\mathcal{K} \oplus \alpha \tilde{B}^{-1} \mathcal{U}), \quad \forall \alpha \in \mathcal{G}, \tilde{A} \in \mathcal{A}, \]

(2.61)

Subsequently, we can redefine \( \mathcal{R}_1 \) as

\[ \mathcal{R}_1 := (\bar{\sigma}(I + \mathcal{A}))^{-1}\mathcal{K}_G(\tilde{A}^{-1}(\mathcal{K} \oplus \tilde{B}^{-1} \mathcal{U})), \]  

(2.62)

and \( \mathcal{R}_i \) in (2.54) as:

\[ \mathcal{R}_i := (\bar{\sigma}(I + \mathcal{A}))^{-1}\mathcal{K}_G(\tilde{A}_i^{-1}(\mathcal{K}_i \oplus \tilde{B}^{-1} \mathcal{U})), \]  

(2.63)
In this new formulation, conservatism of the model-invariant viability kernel under-approximation depends on conservatism of $A^1(\mathcal{X} \oplus B^-)$ as well as the size of $\bar{\sigma}(I + \omega)$. According to Lemma 2.1.2, the under-approximation of $A^{-1}(\mathcal{X} \oplus B^-)$ by a 2-norm ball depends on directionality and the size of the minimum singular value of $A^{-1}$. Furthermore, $(\bar{\sigma}(I + \omega))^{-1}$ may shrink $A_{i}$ in each step of (2.54). To reduce conservatism of the model-invariant viability kernel under-approximation, we employ the following optimization problem to choose decomposition of $A$ in cases where the model set is finite:

$$
\begin{align*}
\min_{A^{-1}, A_i} & \quad \beta \gamma - (1 - \beta) \lambda \\
\text{subject to} & \quad \bar{\sigma}(A_i) \leq \gamma, \\
& \quad \bar{\sigma}(A^{-1}) \geq \lambda, \\
& \quad I + A_i = A^{-1} A_i, \\
& \quad i = 1, \ldots, p.
\end{align*}
$$

(2.64)

In (2.64), $p$ refers to the number of models in $\mathcal{M}$ and $\beta \in [0, 1]$ is a tuning variable. The above optimization problem decreases $\bar{\sigma}(I + \omega)$ by minimizing $\bar{\sigma}(A_i)$ (increasing $(\bar{\sigma}(I + \omega))^{-1}$) and maximizes the minimum singular value of $A^{-1}$ (increases the condition number of $A^{-1}$). Maximizing the minimum singular value is a non-convex problem [64], thus the above-mentioned optimization problem is non-convex. To remedy this, we employ a convex-concave relaxation of this problem as discussed in [64]. Solving the relaxed version of (2.64) does not provide a globally optimal solution. However, we will show that decomposing members of $A$ by solving the convex-concave relaxation of (2.64) significantly decreases conservatism of the model-invariant viability kernel under-approximation. Note that (2.64) can be used to reduce conservatism of the under-approximation only in cases where the model-set is finite.

**Computational example**

Here, we apply the proposed technique to under-approximate the model-invariant viability kernel of a system with multiplicative model-uncertainty. This example is based on the example described in [59]. Consider the following uncertain state-
space model:

\[
\dot{x}(t + 1) = (A^0 + \Delta)x(t) + \alpha B^0 u(t),
\]

(2.65)

where,

\[
A^0 = \begin{bmatrix} 0.261 & -1.098 \\ 0.891 & 0.419 \end{bmatrix}, \quad B^0 = \begin{bmatrix} 0.319 \\ -1.308 \end{bmatrix},
\]

and

\[
\Delta \in Co\{ \begin{bmatrix} 0 & 0.125 \\ -0.125 & 0 \end{bmatrix}, \begin{bmatrix} -0.025 & 0 \\ 0.125 & -0.025 \end{bmatrix}, \begin{bmatrix} -0.025 & -0.125 \\ 0 & 0.025 \end{bmatrix} \}, \quad \alpha \in [1, 1.5].
\]

In the above model, \( Co\{ \cdot \} \) denotes the convex hull of the set. Consider the following constraints on the states and input of (2.65):

\[
\|x(t)\|_2 \leq 41.7, \quad (2.66)
\]

\[
\|u(t)\|_2 \leq 12.5. \quad (2.67)
\]

We uniformly sample from the continuous uncertainty to derive a multi-model description of uncertainty in the form of (2.17). Moreover, we employ the convex-concave relaxation of (2.64) to identify the nominal and uncertainty matrices in the forms of (2.59) and (2.60). We choose \( \beta = 0.69 \) as it results in the least conservative decomposition.

This example is compatible with the description of model uncertainty specified in (2.16), so the proposed approach can be applied. For the viability kernel calculation, we conduct operations on sets using the Ellipsoidal Toolbox (ET) version 1.1.3 [65] on a 3.1 GHz Intel Core i7 with 16 GB RAM running MacOS 10.10.5 and Matlab R2015b.

For the model set to comprehensively describe the uncertain model, a large number of samples from the uncertain model is required. The viability kernel of all members of the model set is required to calculate the model-invariant viability
kernel. However, as the number of models increases, the computational complexity increases. Fig. 2.1 shows a significant increase in time required to calculate the model-invariant viability kernel for (2.65) as the number of model set members increases. In this example, the finite horizon is assumed to be $\tau = 100$. Fig. 2.1 also shows the computation time to under-approximate the model-invariant viability kernel. This includes the time to solve the optimization problem (2.64) in addition to the time to calculate the recursive formula (2.54). Fig. 2.2 compares the model-invariant viability kernel and its under-approximation for the case when $p = 7$. It can be seen that the proposed approach introduced conservatism; however, the time to under-approximate the model-invariant viability kernel is significantly lower. Fig. 2.2 also compares the case where (2.64) is used to reduce conservatism with the case where (2.64) is not used. It can be seen that the optimization using (2.64) significantly reduces conservatism of the under-approximation.

The other source of conservatism in (2.54) is the number of recursions. Each iteration of (2.54) may shrink the under-approximation. Fig. 2.3 illustrates the evolution of the model-invariant viability kernel under-approximation over time.

**Figure 2.1:** Computation time of the model-invariant viability kernel calculation (orange bar) vs. computation time of the model-invariant viability kernel under-approximation using the proposed method (blue bar).
Figure 2.2: Model-invariant viability kernel vs. its under-approximation ($p = 7$). Red region: $\mathcal{K}$; dashed blue region: $\text{Viab}_{[0,100] \cap \mathbb{Z}}^{+}(\mathcal{K}, \mathcal{U}, \mathcal{M})$; dashed-dotted back region: the under-approximation of $\text{Viab}_{[0,100] \cap \mathbb{Z}}^{+}(\mathcal{K}, \mathcal{U}, \mathcal{M})$ using the proposed approach; dotted green region: the under-approximation of $\text{Viab}_{[0,100] \cap \mathbb{Z}}^{+}(\mathcal{K}, \mathcal{U}, \mathcal{M})$ with reduced conservatism using (2.64).

Due to the fact that viability kernel calculation is a backward time analysis, the evolution starts at $t = 100$ and finishes at $t = 0$ in Fig. 2.3. Accordingly, the under-approximation converges fairly fast and conservatism due to the number of iterations is not significant. Despite the convergence of the under-approximation in this example, convergence of the recursive formula (2.54) is problem dependent and cannot be generalized.

2.4 Conclusion

This chapter introduced the model-invariant safety-preserving control technique. This technique extends safety-preserving control methods to cases with multiplicative model uncertainty. Given a multi-model description of model uncertainty,
the model-invariant safety-preserving control method guarantees the existence of a control action which maintains safety of all members of the model set. The proposed control technique relies on the model-invariant viability kernel. We showed that for the model set $\mathcal{M}$ defined in the form of (2.17), the intersection of the viability kernels of all members of the model set is indeed the model-invariant viability kernel. However, as the size of the model set increases, the computational complexity of the model-invariant viability kernel calculation increases. To remedy this, we proposed a method to under-approximate the model-invariant viability kernel without the need for computing the viability kernel of each member of the model set. Moreover, we proposed an optimization problem to minimize the conservatism of the under-approximation in cases where the model set is finite. The model-invariant safety-preserving control technique extends the applicability of safety-preserving control to systems with multiplicative model-uncertainty.

Figure 2.3: Evolution of the model-invariant viability kernel under-approximation (blue region). Dashed red line: upper/lower limits on the states.
Chapter 3

Output-Feedback

Safety-Preserving Control

In most safety-preserving control and safety verification techniques, there is an implicit assumption that the system states are fully measurable. Only a few papers in the literature discuss safety verification of output-feedback control systems. For instance, Lesser and Abate in [39] proposed a general framework to approximate the feedback invariant for output-feedback control systems (see also [50] and [66]). For a given observer, they suggest to calculate an upper bound on the estimation error and reduce the size of the safe region accordingly. They show that for a given controller which only sees an estimate of the states, if we choose initial conditions of the output-feedback system from the feedback invariant calculated based on the contracted safe region, the controller keeps the states of the system within the actual safe region. There are two major concerns regarding this method. First, in output-feedback systems we do not have access to true initial conditions. However, if we quantify the difference between initial conditions of the system and initial conditions of the observer, we can address this concern by further reducing the size of the safe region. Second, the feedback invariant is controller-specific, which means it can only be used for the controller used to approximate the feedback invariant. It does not provide us with any extra information about other feedback control laws that may result in a bigger set of viable trajectories.

The main contribution in this chapter is to introduce a safety-preserving control
scheme for output-feedback control systems. Given a system with a stable state observer and a bound on the difference between initial conditions of the system and the observer, we quantify a viable tube for the observer based on constraints on the actual states. We prove that a control input which keeps trajectories of the observer (observed trajectories) within the specified viable tube also keeps trajectories of the actual system within the safe region. This approach is not limited to a specific type of controller and any safety-preserving control techniques can be employed to maintain the viability of the estimated states.

The other major contribution is to extend the proposed output-feedback safety-preserving control technique to systems with measurement noise and model uncertainty. Assuming the measurement noise is bounded, the multi-model description of model uncertainty is given, a given observer is stable for all members of the model set, we calculate a viable tube for the observer. Similarly, we show that a control input which maintains the observed trajectories within the calculated viable tube also maintains the states of all members of the model set within the safe region.

3.1 Background

This section summarizes the general framework Lesser and Abate proposed in [39] for safety verification of output-feedback control systems. Consider $U \subset \mathbb{R}^m$ and $K \subset \mathbb{R}^n$ as compact convex sets specifying constraints on inputs and states of a system, respectively. To verify that a closed-loop system is safe using the approach proposed in [39], one needs to show that the closed-loop controller provides a constrained control input which maintains the system’s states inside $K$ for all $t \in \mathbb{T}$, where $\mathbb{T} = [0, \tau]$, and $\tau$ specifies the final time.

Consider the following state-space equation:

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

(3.1)

with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$. In (3.1), $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ denote vectors of the states, inputs and outputs, respectively. The system $X$ is
assumed to be observable. Consider $X_g$ as a closed-loop system which is formed by combining the system $X$ with the closed-loop control policy $g(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ($u(t) = g(x(t))$).

**Definition 6** (Feedback invariant). The finite-horizon feedback invariant set for the closed-loop system $X_g$ is the set of all initial states starting from which the feedback controller ($u(t) = g(x(t))$) maintains the states of $X_g$ within the safe region for all time in $\mathbb{T}$ without violating the input constraint $\mathcal{U}$:

$$\mathcal{F}_T(\mathcal{X}, \mathcal{U}, X_g) = \{ x_0 \in \mathcal{X} \mid x(t) = x_0, u(t) = g(x(t)) \in \mathcal{U}, \forall t \in \mathbb{T}, x(t) \in \mathcal{X} \}. \quad (3.2)$$

Using the feedback invariant set to guarantee safety requires full knowledge of the states. For state-space model (3.1) in which only $y(t)$ is measurable, an observer can be implemented to estimate the states. Then, the question is if the controller which only sees the estimated states ($u(t) = g(\hat{x}(t))$), still satisfies the feedback invariance property. Relying on fast convergence of high-gain observers, Lesser and Abate in [39] employed a high-gain observer of the following form to estimate the states of (3.1):

$$O : \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + H(\varepsilon)(y(t) - C\hat{x}(t)). \quad (3.3)$$

The observer gain $H(\varepsilon)$ is characterized by a positive parameter $\varepsilon$:

$$H(\varepsilon) = \left[ \frac{\alpha_1}{\varepsilon} \quad \frac{\alpha_2}{\varepsilon^2} \quad \cdots \quad \frac{\alpha_n}{\varepsilon^n} \right]^T. \quad (3.4)$$

The $\alpha_i$s are selected such that the following polynomial is Hurwitz:

$$s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_{n-1} s + \alpha_n. \quad (3.5)$$

The selection of $\alpha_i$s guarantees that the error dynamics defined over the error signal,

$$e(t) = x(t) - \hat{x}(t), \quad (3.6)$$

32
is stable. By choosing $\varepsilon$ arbitrarily small, the error signal converges to zero arbitrarily fast [67]. Lesser and Abate quantified an upper bound on the estimation error ($e(t)$) as a function of $\varepsilon$:

$$
\|e(t)\|_2 = \|x(t) - \hat{x}(t)\|_2 \leq \delta(\varepsilon).
$$

The following procedure is proposed in [39] to calculate the feedback invariant set for the output-feedback control system $X_g$ in which the controller sees the estimate of the states:

1. Compute $\delta(\varepsilon)$ for the high-gain observer,

2. Assuming the states of $X_g$ are fully measurable, compute $\mathcal{F}_T(\mathcal{H}, \mathcal{U}, X_g)$ using formal methods with the state constraint set $\mathcal{K} = \mathcal{K} \ominus \mathcal{B}(\delta(\varepsilon))$.

Lesser and Abate [39] proved that closed-loop trajectories of $X_g$ starting from $x(0) \in \mathcal{F}_T(\mathcal{H}, \mathcal{U}, X_g)$ and under the control policy $u(t) = g(\hat{x}(t))$, stay within the safe set $\mathcal{K}$. One of the drawbacks of this approach is that $x(0)$ is unknown and we cannot show that $x(0) \in \mathcal{F}_T(\mathcal{H}, \mathcal{U}, X_g)$. However, if the uncertainty on $x(0)$ can be specified as $\|x(0) - \hat{x}(0)\|_2 \leq \lambda$, by choosing $\hat{x}(0)$ from

$$
\mathcal{F} = \mathcal{F}_T(\mathcal{H}, \mathcal{U}, X_g) \ominus \mathcal{B}(\lambda),
$$

it can be shown that $x(0) \in \mathcal{F}_T(\mathcal{H}, \mathcal{U}, X_g)$.

### 3.2 Output-Feedback Safety-Preserving Control

One of the shortcomings of the approach described in the previous section is that the feedback invariant set is limited to a specific controller and cannot be used in general. Moreover, there might be a bigger set of initial conditions which result in viable trajectories with a different choice of controller. This larger set may result in less conservative solutions. In this section, we propose a more general approach to preserve safety of control systems in which the states are not fully measurable.
3.2.1 Notations and definitions

Consider the safe tube $\mathcal{K}_T$ which characterizes constraints on trajectories of $X$:

$$\mathcal{K}_T = \{ x(\cdot) \mid \forall t \in T, x(t) \in \mathcal{K}_t \}. \quad (3.9)$$

$\mathcal{K}_t \subset \mathbb{R}^n$ specifies constraints on the states of $X$ at time $t$. Since the state constraints do not change over time, $\mathcal{K}_t = \mathcal{K}$ in (3.9).

**Definition 7** (Output-feedback safety-preserving control). *Output-feedback safety-preserving control guarantees that there is a constrained input that keeps trajectories of $X$ within $\mathcal{K}_T$ despite the fact that the states of $X$ are not fully measurable and only their estimates are available.*

We use a state observer of the following form to estimate the states of $X$:

$$O : \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(y(t) - C\hat{x}(t))$$

$$= (A - LC)\hat{x}(t) + Bu(t) + LCx(t). \quad (3.10)$$

In the above equation $L \in \mathbb{R}^{n \times p}$ must be designed such that $O$ is stable. Since $H(\varepsilon)$ stabilizes $O$, we can also use $H(\varepsilon)$ defined in (3.4) instead of $L$ in (3.10). Consider the estimation error defined in (3.6) as the difference between the states of $X$ and their estimates. We can formulate the error dynamics illustrating the evolution of the estimation error, as follows:

$$E : \dot{e}(t) = (A - LC)e(t). \quad (3.11)$$

Consider $\varepsilon_0 \subset \mathbb{R}^n$ as a convex compact set of all possible initial errors $e(0)$ and $\varepsilon_t \subset \mathbb{R}^n$ as a set of all reachable errors at time $t$ calculated based on the evolution of the error dynamics starting from $\varepsilon_0$. Now, we can define the error tube which is the set of all error trajectories ($e(\cdot)$) starting from $\varepsilon_0$:

$$\mathcal{E}^E = \{ e(\cdot) \mid \forall t \in T, e(t) \in \varepsilon_t \}, \quad (3.12)$$

Due to the stability of the error dynamics, $\varepsilon_t$ becomes smaller as time goes forward, and if $\tau$ is sufficiently large, $\varepsilon_t$ converges to zero. According to the observer
dynamics defined in (3.10), the states of $X$ have a direct influence on the states of the observer. However, since we can specify the set of all estimation errors, the observer dynamics can be reformulated independently of the states of $X$:

$$O: \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + LCe(t), \ e(t) \in \mathcal{E}. \quad (3.13)$$

By looking at (3.1) and (3.13), we can see that not only $X$ and $O$ share the same input, but also have the same dynamics. The only difference is that the states of $O$ are perturbed by the external variable $e(t) \in \mathcal{E}$. 

**Definition 8** (Discriminating kernel). *The finite-horizon discriminating kernel of $\mathcal{K}_T$ for $O$ defined in (3.13) is a subset of $\mathcal{K}_0$ ($\mathcal{K}$ at $t = 0$) which specifies all initial conditions for which there exists an admissible input $u(\cdot) \in \mathcal{U}_T$ that keeps trajectories of $O$ starting from those initial states within $\mathcal{K}_T$ for all $e(\cdot) \in \mathcal{E}^T$:*

$$\text{Disc}(\mathcal{K}_T, \mathcal{U}_T, \mathcal{E}^T, O) = \{ \hat{x}_0 \in \mathcal{K}_0 | \dot{\hat{x}}(0) = \hat{x}_0, \ \exists u(\cdot) \in \mathcal{U}_T \text{ s.t. } \forall e(\cdot) \in \mathcal{E}^T, \ \dot{\hat{x}}(\cdot) \in \mathcal{K}_T \}. \quad (3.14)$$

In (3.19), $\mathcal{U}_T = \{ u(\cdot) | \forall t \in T, u(t) \in \mathcal{U} \}$ is a set of all admissible inputs over $T$. Without loss of generality, we can reformulate Definition 7 as follows:

**Definition 9** (Output-feedback safety-preserving control). *Output-feedback safety-preserving control guarantees that there is a constrained input which keeps the state of $O$ within a given set of constraints (a set which we can under-approximate), and that input will also keep the states of $X$ within $\mathcal{K}$.*

### 3.2.2 Main result

The next proposition formulates output-feedback safety-preserving control assuming $\text{Disc}(\mathcal{K}_T, \mathcal{U}_T, \mathcal{E}^T, O)$ is non-empty, where

$$\hat{\mathcal{K}}_T = \mathcal{K}_T \cup \mathcal{E}^T. \quad (3.15)$$

**Proposition 3.2.1.** *For the state-space model $X$ defined in (3.1), the state observer $O$ formulated in (3.13), the error tube $\mathcal{E}^T$ and the safe tube $\hat{\mathcal{K}}_T$ defined in (3.15), assuming $\text{Disc}(\hat{\mathcal{K}}_T, \mathcal{U}_T, \mathcal{E}^T, O)$ is not empty, a safety-preserving control input that
keeps trajectories of $O$ (estimated states) within $\hat{\mathcal{K}}_T$, also maintains trajectories of $X$ within $K_T$.

**Proof.** Since $\text{Disc}(\hat{\mathcal{K}}_T, U_T, \mathcal{E}^E, O)$ is not empty, we can define the following viable tube as a set of all viable trajectories of the observer inside $\hat{\mathcal{K}}_T$:

$$\mathcal{T}_O^T = \{\hat{x}(\cdot) | \forall t \in T, \hat{x}(t) \in \text{Disc}(\hat{\mathcal{K}}_{[t,T]}, U_{[t,T]}, \mathcal{E}^E_{[t,T]}, O)\},$$

(3.16)

According to Definition 8, $\text{Disc}(\hat{\mathcal{K}}_{[t,T]}, U_{[t,T]}, \mathcal{E}^E_{[t,T]}, O)$ is a set of states at time $t$ starting from which there exists $u(\cdot) \in U_{[t,T]}$ that keeps trajectories of $O$ within $\hat{\mathcal{K}}_{[t,T]}$. Due to the fact that the defined trajectories start from $\text{Disc}(\hat{\mathcal{K}}_T, U_T, \mathcal{E}^E, O)$, $\mathcal{T}_O^T$ is non-empty, and there exists $u(\cdot) \in U_T$ to keep trajectories of $O$ viable. Moreover, since $\mathcal{T}_O^T$ includes trajectories of $O$ inside $\hat{\mathcal{K}}_T$, we have $\mathcal{T}_O^T \subseteq \hat{\mathcal{K}}_T$. As we discussed previously, the set $\mathcal{T}_E^E$ defined in (3.36) is a set of all trajectories of the error between the actual states and the estimated states starting from $E_0$. Moreover, based on the definition of the estimation error (3.6), we can calculate the states of $X$ as $x(t) = \hat{x}(t) + e(t)$. Thus, using the same input which generates $\mathcal{T}_O^T$, we can characterize all possible trajectories of $X$ as follows:

$$\mathcal{T}_X^T = \mathcal{T}_O^T \oplus \mathcal{T}_E^E.$$

(3.17)

Since $\mathcal{T}_O^T \subseteq \hat{\mathcal{K}}_T$ and subsequently $\mathcal{T}_O^T \subseteq K_T \oplus \mathcal{E}^E$, we have

$$\mathcal{T}_X^T = \mathcal{T}_O^T \oplus \mathcal{T}_E^E \subseteq (K_T \oplus \mathcal{E}^E) \oplus \mathcal{E}^E,$$

(3.18)

and according to Lemma 2.1.1, $\mathcal{T}_X^T \subseteq \mathcal{K}_T$. 

According to Proposition 1, in an output-feedback control system with an observer, a safety-preserving control input that keeps the states of the observer within the contracted version of the safe region, also maintains the states of the actual system within the safe region. In this proposition, we assume that $\text{Disc}(\hat{\mathcal{K}}_T, U_T, \mathcal{E}^E, O)$ is not empty. However, if the set is empty, the proposed technique cannot find a controller to preserve safety for a given output-feedback system with a given set of constraints. Once $\text{Disc}(\hat{\mathcal{K}}_T, U_T, \mathcal{E}^E, O)$ is calculated, one can employ a safety-preserving control policy (such as (2.10)) to preserve the safety of the estimated states and consequently, the safety of the actual states. Note that in the proposed
output-feedback safety-preserving control scheme, \( \text{Disc}(\mathcal{K}_T, \mathcal{U}_T, \mathcal{F}_T, O) \) must be used instead of \( \text{Viab}_T(\mathcal{H}, \mathcal{U}, X) \) to formulate safety-preserving controllers. Kaynama et al. in [34] summarized how to employ the discriminating kernel in safety-preserving control formulations.

The main difference between the proposed approach and the one described in the previous section is that the proposed approach is not controller-specific. Once \( \text{Disc}(\mathcal{K}_T, \mathcal{U}_T, \mathcal{F}_T, O) \) is approximated, we can formulate a variety of safety-preserving controllers to maintain trajectories of the observer within \( \mathcal{K}_T \) (see [1, 31, 34]). Moreover, to approximate the feedback invariant set for output-feedback systems, the safe region needs to be eroded by an upper bound on the estimation error. However, the proposed approach suggests to calculate the evolution of the error dynamics and erode the safe region by a set of all possible errors. Accordingly, due to the stability of the error dynamics, the error set gets smaller as time goes forward which results in less erosion of the safe region and consequently, less conservative results.

### 3.2.3 Discriminating kernel approximation

Similar to the viability kernel calculation discussed in the previous section, Kaynama et al. [34] showed that the discriminating kernel can be under-approximated by recursive approximation of the maximal reachable sets. Due to the fact that \( \hat{\mathcal{K}}_t \) changes over time, we cannot employ this method to approximate \( \text{Disc}(\hat{\mathcal{K}}_T, \mathcal{U}_T, \mathcal{F}_T, O) \). Here, we extend the algorithm described in [34] to the case where the safe region (3.15) is time-varying. To do so, we need to extend the definition of the backward reachable set (Definition 5) to systems in the form of (3.13).

**Definition 10** (Backward reachable set). The backward reachable set of the system \( O \) defined in (3.13) from set \( \mathcal{H} \subset \mathbb{R}^n \) over the time domain \( \mathbb{T} \) with the finite horizon \( \tau \) is a set of initial conditions starting from which there exists a constrained input \( (u(\cdot) \in \mathcal{U}_T) \) that the states of \( O \) reach \( \mathcal{H} \) at time \( t = \tau \):

\[
\text{Reach}_\tau(\mathcal{H}, \mathcal{U}_T, \mathcal{F}_T, O) = \{ \hat{x}_0 \in \mathbb{R}^n | \hat{x}(0) = \hat{x}_0, \exists u(\cdot) \in \mathcal{U}_T \text{ s.t. } \forall e(\cdot) \in \mathcal{F}_T, \hat{x}(\tau) \in \mathcal{H} \}.
\] (3.19)
To show that (3.22) holds, we need to show for all $t_0, t_1, \ldots, t_l \in \mathbb{T}$, $t_0 < t_1 < \cdots < t_l$ and $t_0 = \min_{t \in \mathbb{T}} t = 0$ and $t_l = \max_{t \in \mathbb{T}} t = \tau$. Moreover, we define $|\mathbb{P}_T|$ as $|\mathbb{P}_T| := \max\{t_{k+1} - t_k \mid k = 0, 1, \ldots, l - 1, \ t_k \in \mathbb{P}_T\}$. Assume $O$ is bounded by $M$ on $\hat{\mathcal{K}}_T$, $\mathcal{U}_T$ and $\mathcal{R}^E_T$, which means $\forall t \in \mathbb{T}$, $\forall \hat{x}(t) \in \hat{\mathcal{K}}_T$, $\forall u(t) \in \mathcal{U}_T$, and $\forall e(t) \in \mathcal{E}_T$, $\|\dot{x}(t)\|_2 \leq M$. We define an under-approximation of $\hat{\mathcal{K}}_T$ as:

$$\hat{\mathcal{K}}_{T}^\perp = \{\hat{x}(\cdot) \mid \forall t \in \mathbb{T}, \ x(t) \in \hat{\mathcal{K}}_{T}^\perp\} \tag{3.20}$$

where $\hat{\mathcal{K}}_{T}^\perp = (\hat{\mathcal{K}}_T \ominus \mathcal{D}) \ominus \mathcal{D}(M|\mathbb{P}_T|)$, and $\mathcal{D} = \arg \max \{\text{vol}(\mathcal{Z}) \mid \mathcal{Z} = \hat{\mathcal{K}}_{T_{l-1}} \ominus \hat{\mathcal{K}}_{T_1}, \ t_l \in \mathbb{P}_T, \ t \in [t_{l-1}, t_l]\}$. In the above equation, $\text{vol}(\mathcal{Z})$ denotes the volume of the set $\mathcal{Z}$. $\mathcal{D}$ shows the maximum contraction of $\hat{\mathcal{K}}_{T}$ in the interval $[t_{l-1}, t_l]$. In the next proposition, we show that if we keep the states of $O$ for all $t_l \in \mathbb{P}_T$ within $\hat{\mathcal{K}}_{T_l}^\perp$, they will stay within $\hat{\mathcal{K}}_{T_l}^\perp$ for all $t \in \mathbb{T}$.

Consider the following $l$-step recursion:

$$\hat{\mathcal{R}}_l = \hat{\mathcal{K}}_{T_l}^\perp, \tag{3.21}$$

$$\hat{\mathcal{R}}_{l-1} = \hat{\mathcal{R}}_l \cap \text{Reach}_{t_l-t_{l-1}} \left( \hat{\mathcal{R}}_{T_l}, \mathcal{U}_{[t_{l-1}, t_l]}, \mathcal{R}_T^{E}, O \right),$$

for $i = l, l - 1, \ldots, 1$.

**Proposition 3.2.2.** For the state-space equation defined in (3.13) which is bounded by $M$ on $\hat{\mathcal{K}}_T$, $\mathcal{U}_T$ and $\mathcal{R}_T^{E}$, the final set $\hat{\mathcal{R}}_0$ calculated by the recursive formula (3.21) satisfies:

$$\hat{\mathcal{R}}_0 \subseteq \text{Disc}(\hat{\mathcal{K}}_T, \mathcal{U}_T, \mathcal{R}_T^{E}, O). \tag{3.22}$$

**Proof:** Calculating $\hat{\mathcal{R}}_0$ using the recursive relation (3.21) indicates that starting from any points in $\hat{\mathcal{R}}_{l-1}$ there exists $u_t(\cdot) \in \mathcal{U}_{[t_{l-1}, t_l]}$ such that the states of $O$ can reach $\hat{\mathcal{R}}_{T_l}$ at $t = t_l - t_{l-1}$. By taking the concatenation of the inputs $u_t(\cdot)$ for all $t_l \in \mathbb{P}_T$ [27], we can define an input $u_T(\cdot) \in \mathcal{U}_T$ which, starting from any points in $\hat{\mathcal{R}}_0$, the states of $O$ all $t_l \in \mathbb{P}_T$ stay within $\hat{\mathcal{R}}_{l}$:

$$x(t_l) \in \hat{\mathcal{R}}_{T_l} \subseteq \hat{\mathcal{K}}_{T_l}^\perp \subseteq \hat{\mathcal{K}}_{T_l}. \tag{3.23}$$

To show that (3.22) holds, we need to show for all $t \in \mathbb{T}$, $u_T(\cdot)$ maintains $\dot{x}(t) \in \hat{\mathcal{K}}_T$. 

38
Since any $t \in \mathbb{T}$ lies in some interval $(t \in [t_i, t_{i+1}])$ and due to the fact that $O$ is bounded by $M$, we have:

$$
\|\hat{x}(t) - \hat{x}(t_i)\|_2 \leq \| \int_{t_i}^{t} \dot{x} \, dt \|_2 \leq M(t - t_i) \\
< M(t_{i+1} - t_i) \leq M|\mathbb{T}|.
$$

(3.24)

According to (3.24), $v = \dot{x}(t) - \dot{x}(t_i) \in \mathcal{B}(M|\mathbb{T}|)$. Since $\dot{x}(t_i) \in \mathcal{F}^{\perp}$ and according to the definition of $v$ as well as Lemma 1, we have:

$$
\dot{x}(t) \in \mathcal{F}^{\perp} \oplus \mathcal{B}(M|\mathbb{T}|) \\
\subseteq \left( (\mathcal{F} \oplus \mathcal{D} \oplus \mathcal{B}(M|\mathbb{T}|)) \oplus \mathcal{B}(M|\mathbb{T}|) \right) \\
\subseteq \mathcal{F} \oplus \mathcal{D} \subseteq \mathcal{F} \oplus (\mathcal{F} \oplus \mathcal{F}) \subseteq \mathcal{F}.
$$

(3.25)

This result implies that starting from any point in $\mathcal{R}_0$ there is an input to keep the states of $O$ inside $\mathcal{F}$ at time $t$. Thus, (3.22) holds.

\[\square\]

\subsection*{3.2.4 Numerical comparison}

In this section, we apply the proposed output-feedback safety-preserving control to the example Lesser and Abate discussed in [39]. We regenerate the results presented in [39] and compare them with the results of the technique proposed in this work. We show that there exists a safety-preserving control input for the closed-loop system defined below that generates a viable trajectory starting from some initial conditions from outside of $\mathcal{F}$ (defined in (3.8)) but inside $\text{Disc}(\mathcal{F}, \mathcal{U}, \mathcal{X}^E, O)$.

Consider the dynamics of the double integrator:

$$
\begin{bmatrix}
\dot{x}_1(t) \\
\dot{x}_2(t)
\end{bmatrix} =
\begin{bmatrix}
0 & 1 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1(t) \\
x_2(t)
\end{bmatrix} +
\begin{bmatrix}
0 \\
1
\end{bmatrix} u(t),
$$

$$
y =
\begin{bmatrix}
1 & 0
\end{bmatrix}
\begin{bmatrix}
x_1(t) \\
x_2(t)
\end{bmatrix}.
$$

(3.26)
A controller of the following form is used in [39] to stabilize (3.26):

$$u(t) = \begin{bmatrix} -\beta & -\beta \end{bmatrix} x(t). \quad (3.27)$$

Since $x_2(t)$ cannot be measured, the following high-gain observer is designed in [39] to estimate the states of (3.26):

$$\dot{\hat{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \hat{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} \frac{\alpha_1}{\varepsilon^2} \\ \frac{\alpha_2}{\varepsilon^2} \end{bmatrix} (x_1(t) - \hat{x}_1(t)). \quad (3.28)$$

Accordingly, (3.27) needs to be reformulated as

$$u(t) = \begin{bmatrix} -\beta & -\beta \end{bmatrix} \hat{x}(t). \quad (3.29)$$

In this example, the states are required to remain within the safe region $\mathcal{X} = \{ x(t) | |x_1(t)| \leq 4, |x_2(t)| \leq 3 \}$ over $\mathcal{T} = \{ t | t \in [0, 10] \}$. We assume the uncertainty on $x(0)$ to be $\| x(0) - \hat{x}(0) \|_2 = 0.5$. The absolute value of the input is restricted to be less than 1, $\mathcal{U} = \{ u(t) | |u(t)| \leq 1 \}$. The parameters of the observer are set to $\alpha_1 = \alpha_2 = 4$ and $\varepsilon = 0.01$. For this observer the upper bound on the 2-norm of the estimation error is $\delta(\varepsilon = 0.01) = 0.1768$ [39]. Lesser and Abate in [39] showed that for the system and controller defined in (3.26) and (3.27), setting $\beta = 0.2$ results in the largest feedback invariant set.

To approximate the discriminating kernel for the observer defined in (3.28), we calculate the evolution of the error dynamics defined below with the initial error set $\mathcal{E}_0 = \{ e(0) \in \mathbb{R}^2 | \| e(0) \|_2 \leq 0.5 \}$:

$$\dot{e}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} e(t) - \begin{bmatrix} \alpha_1/\varepsilon \\ \alpha_2/\varepsilon^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} e(t). \quad (3.30)$$

In this work, we employ the ellipsoidal techniques [55] implemented in [65] to represent sets and conduct operations on them. Moreover, we utilize the level set toolbox developed by Mitchell [68] for reachability analyses. Following the steps we described in sections [3.1] and [3.2.3] we calculate the feedback invariant ($\mathcal{F}$) for the system described in (3.26), (3.27) and (3.28), as well as the discriminating
Figure 3.1: Conservatism comparison; Solid-line red ellipse: Ellipsoidal under-approximation of safe region $\mathcal{K}$, dashed-line black ellipse: $\text{Disc}(\mathcal{K}^T, \mathcal{U}_T, \mathcal{T}_E^T, O)$, dotted-line blue ellipse: $\hat{\mathcal{F}}$.

Kernel for the observer ($\text{Disc}(\mathcal{K}^T, \mathcal{U}_T, \mathcal{T}_E^T, O)$). The results are illustrated in Fig. (3.1). Comparing $\hat{\mathcal{F}}$ with $\text{Disc}(\mathcal{K}^T, \mathcal{U}_T, \mathcal{T}_E^T, O)$ depicts that we can even have a bigger viable set if we use a different controller rather than the one defined in (3.27). In other words, Fig. 3.1 shows that there is a bigger set of initial conditions that can result in viable trajectories.

In the next step, we employ the safety-preserving controller described in Section 2.1.3 to maintain the states of the observer defined in (3.28) within the viable tube $\mathcal{K}^T$. We employ the controller defined in (3.27) with $\beta = 0.2$ as the performance controller in the proposed output-feedback safety-preserving control scheme. Fig. 3.2 shows an observed trajectory of (3.26) under safety-preserving control starting from $\hat{x}(0) = (1.5, 2)$ which is inside $\text{Disc}(\mathcal{K}^T, \mathcal{U}_T, \mathcal{T}_E^T, O)$ (dotted green line). Fig. 3.2 also depicts the actual trajectory of (3.26) generated with the same input sequence but a different initial condition ($x(0) = (2, 2)$) due to the uncertainty (solid black line). Due to the fast convergence of the high-gain observer, the estimated states rapidly converge to the actual states. In Fig. 3.2, we can see that the safety-preserving control input (solid blue line in Fig. 3.3) that keeps the states of the observer viable preserves safety for (3.26). Fig. 3.2 illustrates a
Figure 3.2: Closed-loop trajectories. Red ellipse: the safe region \( \mathcal{H} \); solid black line: the trajectory of \( X \) controlled by output-feedback safety-preserving control; dotted green line: the observed trajectory of \( X \); dashed blue line: the trajectory of \( X \) controlled by the state-feedback controller (3.27).

Figure 3.3: Control input. Solid blue line: a control input sequence generated by the output-feedback safety-preserving control; dashed black line: a control input sequence generated by the feedback controller (3.27); dotted red lines: upper and lower bounds.

trajectory of (3.26) under the feedback control law (3.29) with the same choice of \( \beta = 0.2 \) (dashed blue line). As we expected, since (1.5, 2) is not inside \( \mathcal{F} \), the controller cannot keep the trajectory inside the safe region.
3.3 Extension to Uncertain Systems

Consider the following uncertain dynamics perturbed by the measurement noise:

\[
    \dot{x}(t) = Ax(t) + Bu(t), \quad A \in \mathcal{A}, \quad B \in \mathcal{B},
\]

\[
y(t) = Cx(t) + n(t). \tag{3.31}
\]

\(C \in \mathbb{R}^{p \times n}\) is assumed to be known. \(\mathcal{A}\) and \(\mathcal{B}\) are compact sets and we assume the maximum singular values of all members of \(\mathcal{A}\) are bounded. This definition includes a wider class of uncertain systems compared to (2.16). \(n(t) \in \mathcal{N}\) is the measurement noise where \(\mathcal{N} \subset \mathbb{R}^{p}\) is a compact convex set. Consider the following nominal model to be used to formulate an observer for (3.31):

\[
    \bar{X} : \quad \dot{x}(t) = \bar{A}x(t) + \bar{B}u(t), \tag{3.32}
\]

where \(\bar{A}\) and \(\bar{B}\) are known members of \(\mathcal{A}\) and \(\mathcal{B}\), respectively. Suppose the following state-observer is robust (stable for all members of the model set):

\[
    O : \quad \dot{\hat{x}}(t) = (\bar{A} - LC)\hat{x}(t) + \bar{B}u(t) + LCx(t) + Ln(t). \tag{3.33}
\]

The corresponding error dynamics can be formulated as \((e(t) = x(t) - \hat{x}(t))\):

\[
    E : \quad \dot{e}(t) = (A - LC)e(t) + (A - \bar{A})\hat{x}(t) + (B - \bar{B})u(t) - Ln(t), \quad A \in \mathcal{A}, \quad B \in \mathcal{B}. \tag{3.34}
\]

Due to the uncertainty in the system, (3.34) is also uncertain. The robustness of the observer implies that (3.34) is stable for all \(A \in \mathcal{A}\). Consider \(\mathcal{M}_E\) as a set of all instances of the error dynamics:

\[
    \mathcal{M}_E = \{ E | E : \quad \dot{e}(t) = (A - LC)e(t) + (A - \bar{A})\hat{x}(t) + (B - \bar{B})u(t) - Ln(t), \quad A \in \mathcal{A}, \quad B \in \mathcal{B} \}. \tag{3.35}
\]

For \(E \in \mathcal{M}_E\) and given \(u(\cdot) \in \mathcal{U}_T\), \(n(t) \in \mathcal{N}\), and assuming \(\hat{x}(\cdot)\) stays within a compact convex viable tube \(\mathcal{V}_T\), consider \(\mathcal{E}^E_T \subset \mathbb{R}^{n}\) as a set of all reachable errors.
at time $t$ starting from $E_0 \subset \mathbb{R}^n$. The corresponding error tube is defined as:

$$
\mathcal{T}_E^t = \{ e(\cdot) | E \in \mathcal{M}_E, E_0^E = E_0, \forall t \in T, e(t) \in E_t^E \}.
$$

(3.36)

Due to the stability of the error dynamics, the error tube converges to a compact set for sufficiently large $\tau$. Let define $E_{t}^{\mathcal{M}_E}$ as

$$
E_{t}^{\mathcal{M}_E} = \bigcup_{E \in \mathcal{M}_E} E_t^E.
$$

(3.37)

$E_{t}^{\mathcal{M}_E}$ is a set of all reachable errors at time $t$ despite the measurement noise and uncertainty. Now, we can reformulate the observer defined in (3.33) as follows:

$$
O: \dot{\hat{x}}(t) = \bar{A}\hat{x}(t) + \bar{B}u(t) + L\bar{e}(t), \; \bar{e}(t) \in \bar{E}_{t}^{\mathcal{M}_E},
$$

(3.38)

where

$$
\bar{e}(t) = Ce(t) + n(t),
$$

(3.39)

and

$$
\bar{E}_{t}^{\mathcal{M}_E} = CE_{t}^{\mathcal{M}_E} \oplus \mathcal{N}.
$$

(3.40)

Accordingly, we can define a maximal error tube $\mathcal{F}_{t}^{\mathcal{M}_E}$ as

$$
\mathcal{F}_{t}^{\mathcal{M}_E} = \{ e(\cdot) | \forall t \in T, e(t) \in \bar{E}_{t}^{\mathcal{M}_E} \}.
$$

(3.41)

**Proposition 3.3.1.** For the uncertain system $X_u$ defined in (3.31), the state observer $O$ formulated in (3.38), the viable tube $\mathcal{V}_T$ and the maximal error tube $\mathcal{F}_{t}^{\mathcal{M}_E}$, assuming $\text{Disc}(\mathcal{V}_T, \mathcal{U}_T, \mathcal{F}_{t}^{\mathcal{M}_E}, O)$ is not empty, a safety-preserving control input that keeps trajectories of $O$ (estimated states) within $\mathcal{V}_T$ maintains trajectories of $X_u$ within $\mathcal{V}_T \oplus \mathcal{F}_{t}^{\mathcal{M}_E}$.

One can take the similar steps discussed in the proof of Proposition 3.2.1 to show the above-mentioned proposition holds. Here, the question is how to find the
set \( \mathcal{V}_T \) such that:

\[
\mathcal{V}_T \oplus \mathcal{P}^{me}_T \subseteq \mathcal{K}_T.
\]  

(3.42)

Algorithm 1 illustrates our recursive formulation to calculate \( \mathcal{V}_T \). In this algorithm, \( \mathcal{V}_T \) is calculated as a 2-norm ball; however, it can be extended to other topologies to reduce conservatism. Starting from a subset of \( \mathcal{K}_T \), Algorithm 1 searches for a set which satisfies (3.42) by shrinking or expanding the initial set. The convergence rate and also conservatism of Algorithm 1 depends on \( \gamma \). A small choice of \( \gamma \) results in a less conservative approximation of \( \mathcal{V}_T \) with the cost of more computational time. Furthermore, if the algorithm converges to an empty set, there will no control action to preserve safety of \( X_u \) in the proposed output-feedback framework. Calculating \( \mathcal{V}_T \) using Algorithm 1 satisfies (3.42). Thus, according to Proposition 3.3.1 starting from \( \text{Disc}(\mathcal{Y}_T, \mathcal{U}_T, \mathcal{F}_T^{me}, O) \) where \( \mathcal{F}_T^{me} \) is calculated based on the assumption that \( \hat{x}(t) \in \mathcal{V}_T \), a safety preserving control action which keeps the estimated states \( (\hat{x}(t)) \) within \( \mathcal{V}_T \), also keeps the states of the uncertain system \( X_u \) within \( \mathcal{K}_T \).

3.4 Conclusion

In this chapter, we have introduced a novel safety-preserving control scheme for output-feedback control systems. Given an observer that estimates the states of an output-feedback control system, we showed that a safety-preserving controller which maintains the estimated states within a predefined subset of the safe region, also keeps the states of the actual system within the safe region. The subset of the safe region in which the estimated states are viable is calculated by eroding the safe region by a maximal set of all possible estimation errors. We showed that the proposed output-feedback safety-preserving control scheme provides less conservative solutions compared to the existing techniques in the literature. Furthermore, this technique is independent of the choice of closed-loop controller.

We also extended the proposed framework to uncertain systems with measurement noise. Given a robust observer, we quantified a maximal estimation error set in the presence of model uncertainty and measurement noise. Similarly, we showed that maintaining observed states of an uncertain system within the safe re-
Algorithm 1 Recursive calculation of the observer’s viable tube

1: procedure
2: Initialization:
3: choose $\lambda_0 > 0$ s.t. $\mathcal{B}(\lambda_0) \subseteq \mathcal{K}$.
4: choose $\gamma > 0$.
5: Loop:
6: define $\mathcal{V}^{T\lambda_i}_T = \{x(\cdot) \mid \forall t \in T, x(t) \in \mathcal{B}(\lambda_i)\}$.
7: calculate $\mathcal{E}_{\mathcal{M}_E}^T$ for all $E \in \mathcal{M}_E$.
8: calculate $\mathcal{T}_T^{\mathcal{M}_E}$.
9: if $\mathcal{V}^{T\lambda_i}_T \oplus \mathcal{T}_T^{\mathcal{M}_E} \subseteq \mathcal{K}_T$ then
10: if $\mathcal{V}^{T\lambda_i}_{T-1} \oplus \mathcal{T}_T^{\mathcal{M}_E} \subseteq \mathcal{K}_T$ then
11: return $\mathcal{V}^{T\lambda_i}_{T-1}$.
12: else
13: $\lambda_{i+1} \leftarrow \lambda_i - \gamma$.
14: if $\lambda > 0$ then
15: $i \leftarrow i + 1$.
16: goto Loop.
17: else
18: return $\mathcal{V}^{T\lambda_i}_T = \emptyset$.
19: else
20: $\lambda_{i+1} \leftarrow \lambda_i + \gamma$.
21: $i \leftarrow i + 1$.
22: goto Loop.

gion eroded by the maximal error set results in the safety of the uncertain system.
Chapter 4

Case Study I - Formalized Safety Systems for Closed-loop Anesthesia

Anesthesia control systems perform automated administration of anesthetic drugs. In closed-loop intravenous anesthesia, drug infusion is manipulated based on feedback of measured clinical effects to induce and maintain a certain level of anesthesia [12, 69]. Appropriately designed and implemented control systems in anesthesia are robust to inter-patient variability [12] and provide less variability in desired clinical effects than manual practice of anesthesia performed by anesthesiologists [70]. Moreover, closed-loop anesthesia minimizes the risk of drug under/overdosing. The feasibility of closed-loop anesthesia has been shown with an experimental system in several clinical studies (e.g. [15, 16, 69, 71–73]).

Despite these efforts, this technology has not been adopted into routine clinical environments. To receive regulatory approval to employ anesthesia control systems as a medical device, patient safety must be guaranteed and demonstrated. To evaluate a safety system for a closed-loop drug delivery system, extensive simulation testing under normal conditions and in the presence of faults is commonly carried out [19–21]. This is a time-consuming process. Moreover, only a finite number of scenarios can be considered in such simulations. Consequently, even if the system meets the safety specifications in the simulated scenarios, other clinical
scenarios may lead to unsafe situations. In contrast, safety-preserving control and formal model verification techniques are capable of verifying safe operation for all possible states and inputs [22, 23].

The main contribution in this chapter is to formalize existing safety systems for closed-loop anesthesia using the methods discussed in Section 2.1. The safety systems reported in the literature for closed-loop anesthesia constrain an open-loop prediction of patient’s states [19, 51]. In these safety systems, model uncertainty and inter-patient variability are not considered [19, 51].

4.1 Background

Dumont et al. in [18] designed a PID controller for administration of propofol, an anesthetic drug used to induce and maintain anesthesia, in closed loop. This controller is shown to be robust to large inter-patient variability [18] and its performance was clinically evaluated [74], [73]. Van Heusden et al. [19] proposed a safety system for the mentioned anesthesia control system by specifying safety constraints on predicted propofol concentrations in the plasma and effect-site. The constraints are defined based on the therapeutic window of propofol. The propofol concentrations are predicted using a population-based model. This safety system minimizes the risk of drug under/overdosing. To improve the performance of the PID controller when the safety constraints are active and the states are saturated, van Heusden et al. [19] used anti-windup (AW) strategies. The first step in this section is to formalize the above-mentioned safety system. The formalized safety system guarantees that the predicted plasma and effect-site concentrations remain within the therapeutic window of propofol during closed-loop anesthesia. We evaluate performance of the formalized safety system for closed-loop anesthesia by examining the system in the clinical scenarios proposed by van Heusden et al. [19].

The safety system mentioned above focuses on the depth of hypnosis as a clinical effect. Most anesthetics can have serious side effects, including respiratory and cardiovascular depression. Including safety constraints on additional physiological variables affected by the propofol infusion have been proposed to further improve patient safety [51]. Low blood pressure is common in the period following induc-
tion of anesthesia. High infusion rates during the induction period may decrease the blood pressure \[75\]; however, it is unknown whether low blood pressure can be avoided for all patients while achieving sufficient anesthesia. We extend the safety system discussed in \[19\] by adding constraints on the predicted blood pressure of patients as suggested in \[51\]. This is expected to improve the system’s safety particularly for patients at risk of cardiovascular depression; however, it is unclear whether these bounds will still allow for a clinically acceptable induction time of anesthesia for the majority of patients in a target population.

**Definition 11.** The induction of anesthesia is completed if a measure of depth of hypnosis\(^1\) reaches the index of 60 and stays below 60 for more than 30 seconds.

We formalize the extended safety system and investigate the effect of safety constraints corresponding to a maximal blood pressure decrease of 50% as suggested by Khosravi \[51\]. We employ a set of 44 patient models identified from clinical data \[76\] and simulate their responses during closed-loop anesthesia with the extended safety system in place. We use the simulated responses to illustrate the feasibility of achieving sufficient depth of anesthesia in the presence of the blood pressure constraints.

### 4.2 Modeling Propofol Effects

This section describes the models we use for formal safety verification in closed-loop anesthesia. The effect of propofol anesthesia is generally described by a 3-compartment PK model, and a first-order model followed by a nonlinear function (PD model). The PK model relates propofol infusion rates to the drug concentration in the plasma \((C_p)\). The PD model shows the relation between the plasma concentration and the propofol effect. The PKPD model can be represented as a 4-state state-space equation with a nonlinear function on the output:

\[
\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = C_e(t); \\
E(C_e(t)) = \frac{C_e(t)^\lambda}{EC_{50}^\lambda + C_e(t)^\lambda}. \tag{4.2}
\]

\(^1\)Measures of depth of hypnosis provide an index between 0 and 100. The index is between 90 and 100 for awake patients.
In (4.1), \(x(t) \in \mathbb{R}^4\) is the state vector which includes drug concentrations in the plasma \((x_1(t) = C_p(t))\), fast \((x_2(t))\) and slow \((x_3(t))\) compartments, and effect-site \((x_4(t) = C_e(t))\). \(u(t) \in \mathbb{R}\) is the propofol infusion rate. \(A\) and \(B\) are the state-space matrices of appropriate sizes. In (4.2), \(EC_{50}\) is the concentration resulting in 50% of the overall effect, \(\lambda \in \mathbb{R}\) is the cooperativity coefficient, and \(E(C_e(t)) \in [0, 1]\) is the propofol effect. The propofol effect used as feedback in closed-loop propofol anesthesia is depth-of-hypnosis (DoH). In this section, we are also interested in the propofol effect on blood pressure (BP).

### 4.2.1 Propofol effect on depth of hypnosis

A measure of DoH is commonly used as feedback in closed-loop anesthesia [18, 71]. Different methods of measuring DoH have been discussed in [77, 78]. DoH is generally described by an index between 0 to 100:

\[
\text{DoH} = 100(1 - E(C_e^{(\text{DoH})}(t))).
\] (4.3)

\(C_e^{(\text{DoH})}\) is the DoH effect-site concentration. For a conscious patient, the DoH index is between 90 and 100. In this paper, we employ the population-based PKPD model identified by Schnider et al. [46]. The safety system proposed in [19] limits \(C_p^{(\text{DoH})}\) and \(C_e^{(\text{DoH})}\) (the plasma and effect-site concentrations corresponding to the DoH PKPD model). This minimizes the risk of drug under/overdosing. For the sake of simplicity, in the rest of this section, we use the plasma and effect-site concentrations \((C_p\) and \(C_e)\) to refer to the concentrations corresponding to the Schnider’s et al. PKPD model \((C_p^{(\text{DoH})}\) and \(C_e^{(\text{DoH})}\)).

### 4.2.2 Propofol effect on blood pressure

To the best of our knowledge, the PKPD models identified by Kazama et al. [48] and Jeleazcov et al. [79] are the only published models which describe the effect of propofol infusion rates on blood pressure. Jeleazcov et al. [79] identified a relation between pre-induction blood pressure and demographics in addition to the model parameters. Extrapolation of these relations results in a baseline systolic pressure of 223 mmHg for a 60 y old patient, indicating significant overfitting, and this model cannot be generalized. In this work, we employ the BP PKPD model
identified by Kazama et al. [48]. Kazama et al. [48] specified the propofol effect on blood pressure as % change from the baseline with the maximal drug effect limited to a systolic blood pressure of 80 mmHg. In [48], the percentage of the blood pressure decrease is specified as:

\[ \% \text{ BP decrease} = 100 \frac{E(C_{BP}(t))}{E(C_{BP})}. \]  

(4.4)

In (4.4), \( C_{BP}(t) \) denotes the effect-site concentration corresponding to blood pressure. Kazama et al. [48] used Gept’s et al. PK model [80] to predict \( C_p \) and identified the blood pressure PD model based on clinical data [48]. They identified the PD model for 4 different age groups (20-39 yr, 40-59 yr, 60-69 yr, 70-80 yr).

### 4.3 Limiting Plasma/Effect-Site Concentrations

Here, our goal is to formalize the safety system proposed in [19] which guarantees that the predicted states of a population-based PKPD model remain within safety constraints. In this safety system, safety constraints are defined for the predicted drug effect concentrations associated with DoH [19]. This safety system uses the PKPD models described in section 4.2.1 that are commonly used to guide drug infusion in target controlled drug infusion systems. The following safety constraint is suggested in [19] for closed-loop propofol anesthesia:

\[ 1.5 \text{mg/l} \leq C_e \leq 8 \text{mg/l}. \]  

(4.5)

The propofol infusion rate is limited by:

\[ 0 \leq u(t) \leq 600 \text{ml/h}. \]  

(4.6)

The initial peak in the plasma concentration after the induction of anesthesia can result in drug toxicity (typically hypotension) [81]. Thus, the following constraint is specified on \( C_p \) in [19]:

\[ 0 \leq C_p \leq 10 \text{mg/l}. \]  

(4.7)

The above-mentioned safety constraints do not necessarily represent clinical
Figure 4.1: Block diagram of closed-loop anesthesia with the formalized safety system limiting plasma and effect-site concentrations

hard bounds. Ensuring the drug dosing remains within this region of the therapeutic window reduces the risk of under- or overdosing for the majority of the population. However, the bounds are expected to be reached for outliers, indicating that the response of these patients or their anesthetic requirements are not well represented by the population average. Activation of the safety system indicates this outlier behaviour and allows the anesthesiologist to make a clinical decision, while the closed-loop system will not, autonomously, compromise patient safety.

We employ the safety-preserving control technique discussed in Section 2.1 to design a safety-preserving controller to maintain viability of the state of the PKPD model. We use the PID controller developed by [74] as the performance controller in the hybrid safety-preserving control policy described in Section 2.1.3. We also include back-calculation AW to improve performance of the PID controller [19]. By adding the formalized safety system to the PID controller we guarantee that the closed-loop system keeps the states of the PKPD model within the safety bounds while having all the benefits of the clinically-evaluated PID controller [73], [74]. The main difference between the formalized safety system and other safety systems (e.g. [19]) is that the formal method is capable of verifying safe operation for all possible states and inputs, eliminating the need for extensive simulation testing. Fig. 4.7 shows the block diagram of the anesthesia closed-loop control system which includes the safety-preserving controller (SP control block).
4.3.1 Simulation results

To evaluate performance of the formalized safety system, we simulate the realistic clinical scenarios proposed in [19]. The patient models in these scenarios are selected from a set of 44 patient models identified by Bibian et al. [76] from clinical data. This set includes models for a wide range of patients (i.e. 18 to 60 years old, male and female, ASA\(^{2}\) I-II), providing a sufficient description of inter-patient variability for our analysis. We include the dynamics of the NeuroSENSE monitor (NeuroWave Systems Inc., Cleveland Heights, OH) which provides a measure of DoH [78]. The setpoint for all cases is 50. We formalize the safety system by calculating the viability kernel and formulating the safety-preserving controller for each patient. The PKPD models described in section 4.2 depend on patient demographics and are subsequently patient dependent. Thus, the formalized safety system is patient dependent. We employ the recursive approach described in Section 2.3.1 to approximate the viability kernel. We use convex polytopes to represent sets. To conduct operations on polytopes for viability kernel approximation, we employ the Multi-Parametric Toolbox 3.0 [82]. Since our goal is to preserve safety during maintenance, the PID controller initially starts administering propofol to induce anesthesia with the safety controller deactivated. Once the predicted states of the PKPD model enter the viability kernel, we turn the safety-preserving controller on and let the safety system maintain the states within the safe region.

Scenario 1, linear model

This scenario uses the linear model of patient #5 of the 44-patient model set. The constraint on the effect-site concentration is active in part of this simulation.

Fig. 4.2 illustrates the viability kernel of the PKPD model for Scenario 1 projected onto \(C_p-C_e\) space (green region). The red lines specify the constraints on \(C_e(t)\) and \(C_p(t)\). As shown in Fig. 4.2, the viability kernel covers most of the safe region. Since the relation between the viability kernel and the safe region is similar in the other scenarios, we omit this plot for those scenarios.

Fig. 4.3 shows the simulation results for Scenario 1. When the effect-site concentration approaches the upper bound, classical AW keeps \(C_e(t)\) at the upper bound.

\(^{2}\)American society of anesthesiologists classification index.
bound. The safety-preserving controller brings it back inside the safety bound.

**Scenario 2, stimulation**

In this scenario, closed-loop anesthesia is simulated for the model of patient #35. A disturbance is added 5 minutes after the start of induction of anesthesia which results in reaching the upper bound of the effect-site concentration.

Fig. 4.4 shows the simulation results for *Scenario 2*. In this scenario, the PID controller increases the infusion rate to reject the disturbance; however, the safety-preserving controller and classical AW correct the infusion rate to prevent overdosing.

**Scenario 3, low clearance**

To simulate reduced clearance, non-linear model #15 is multiplied by

\[ G(s) = 1 + \frac{0.8}{(700s + 1)(800s + 1)^2}. \]  

(4.8)
This increases the gain at low frequency. Due to the reduced clearance, the lower bound of $C_e$ is reached. A disturbance is added at $t = 75\text{min}$.

Fig. 4.5 shows the simulation results for Scenario 3. In this scenario, we simulate a patient model with clearance lower than the average population. In this case, the safety-preserving controller as well as classical AW maintain the effect-site concentration at the lower bound for a prolonged time.

In this scenario, because of the lower clearance of the patient model, the performance controller decreases the infusion rate to keep DoH at 50. The reduced infusion rate causes the predicted effect-site concentration to go below the lower bound. Due to the fact that the objective is to maintain the effect-site concentration within the safety bound, the formalized safety system maintains the effect-site concentration at the lower bound.

As illustrated in Fig. 4.3 Fig. 4.4 and Fig. 4.5 the PID controller with AW saturates the effect-site concentration at the lower/upper bound and maintains the
Figure 4.4: Scenario 2. Dotted green line: unconstrained response, dashed blue line: constrained response with classical AW, red solid line: constrained response with the safety-preserving controller.

effect-site concentration within the safety bound. However, this does not necessarily mean that this controller is capable of maintaining viability of the predicted states in all situations. On the other hand, the formalized safety system assures safe operation for all admissible inputs and states. Moreover, the methods used to formalize a safety system are capable of capturing any unsafe situation without the use of simulation scenarios.

4.4 Extending The Safety System

The safety system described previously focuses on the depth of hypnosis as a clinical effect. Anesthetic drugs, aiming to provide adequate depth of hypnosis, have serious potential side effects including cardiopulmonary depression. The administration of propofol depresses blood pressure, and hypotension following induction of anesthesia is common. Compared to the drug effect on DoH, according to the
Kazama model [48] propofol reaches its peak effect on blood pressure at a later time, i.e. the pharmacodynamics are slower. These characteristics can be exploited in a safety system.

The case shown in Figure 4.6 was part of the clinical study described in [83]. Propofol infusion was closed-loop controlled, guided by the $WAV_{CNS}$ measure of depth of hypnosis provided by the NeuroSENSE monitor (NeuroWave Systems Inc., Cleveland Heights, OH). Remifentanil infusion was manually controlled using a target controlled infusion. Figure 4.6 shows closed-loop induction of anesthesia for a 65 year old male, with body mass index 26 and ASA status III. Following induction of anesthesia, the $WAV_{CNS}$ remained slightly above the setpoint for $\approx 5$ minutes. Airway instrumentation during this time caused sustained stimulation, and some response was observed as an increase in blood pressure. After $\approx 10$ minutes, airway instrumentation was completed and followed by a prolonged period of

---

3 American Society of Anesthesiologists (ASA) physical status
Figure 4.6: Example of closed-loop controlled propofol infusion. Top figure: Black, measured WAVCNS. Blue, WAVCNS setpoint. Middle figure: Black, closed-loop controlled propofol infusion. Blue, corresponding predicted effect-site concentration ($C_e$) (Schnider model). Blue dashed, the $C_e$ bounds defined based on the therapeutic window of propofol. Bottom figure: Black, systolic, mean and diastolic blood pressure. The vertical line indicates the time when ephedrine was administered.

surgical preparation without stimulation. Drug dosing remained relatively elevated during airway instrumentation, and while dosing remained within the therapeutic window, the resulting drug accumulation caused hypotension in the period of low stimulation, requiring intervention (ephedrine was administered at $t = 19.25$min). This clinical scenario, with strong stimulation during airway instrumentation, followed by a prolonged period without stimulation, is common as patients are prepared for surgery.

A safety system with an additional constraint on blood pressure may prevent the drug accumulation following induction of anesthesia, and limit the decrease in blood pressure by exploiting the differences in the dynamic responses of blood pressure and depth of hypnosis to propofol infusion. Such a safety system is ex-
expected to be beneficial for patients at risk of cardiovascular depression; however, it may introduce conservatism and limit the system’s performance, particularly by increasing the time for induction of anesthesia. Slow induction of anesthesia (>10min) may lead to patient discomfort. It will also delay the start of surgery which is not acceptable due to logistics. Moreover, slow induction of anesthesia will delay the anesthesiologists ability to control the airway. This may compromise patient safety.

In this part, we extend the safety system formalized in the previous section by adding safety constraints on blood pressure. We formalize the extended safety system to ensure that $C_p$, $C_e$ and blood pressure remain within the safety constraints. We limit the blood pressure decrease to less than 50% as suggested in [51]. For the sake of simplicity in the formal verification analysis, we specify equivalent constraints on $C_e^{(BP)}$. Due to the monotonically increasing characteristic of (4.2), we use the inverse of (4.2) to map the constraints on the blood pressure decrease to constraints on $C_e^{(BP)}$:

$$BP\text{ decrease } \leq 50\% : \begin{cases} 0 \leq C_e^{(BP)} \leq 4.61\text{mg/l, } 20 - 39\text{yr,} \\ 0 \leq C_e^{(BP)} \leq 4.13\text{mg/l, } 40 - 59\text{yr,} \\ 0 \leq C_e^{(BP)} \leq 3.96\text{mg/l, } 60 - 69\text{yr,} \\ 0 \leq C_e^{(BP)} \leq 2.09\text{mg/l, } 70 - 85\text{yr,} \end{cases} \tag{4.9}$$

To calculate the above limits, we use the model explained in section 4.2.2.

### 4.4.1 Simulation results

In this section, we illustrate the effectiveness of the extended safety system in maintaining the predicted states of patients within the safe zone using simulation results. We employ the 44-patient model set. We formalize the extended safety system by calculating the viability kernel and formulating the safety-preserving controller for each patient. We simulate the patients’ responses while the extended safety system is in place. In addition, we simulate the patients’ responses in closed-loop with the formalized safety system discussed in the previous section (no blood pressure constraint). We use the simulations to show the effectiveness of the extended safety system in maintaining $C_e$, $C_p$ and blood pressure within the safety constraints while
providing sufficient anesthesia. Moreover, we employ a clinical scenario in simulation and illustrate how the extended safety system improves patient’s safety in similar situations.

For each patient, we concatenate the DoH and BP PKPD models and formalize the extended safety system using the integrated model:

\[
\begin{bmatrix}
\dot{x}_{\text{DoH}}(t) \\
\dot{x}_{\text{BP}}(t)
\end{bmatrix}
= \begin{bmatrix}
A_{\text{DoH}}(t) & 0_{4 \times 4} \\
0_{4 \times 4} & A_{\text{BP}}(t)
\end{bmatrix}
\begin{bmatrix}
x_{\text{DoH}}(t) \\
x_{\text{BP}}(t)
\end{bmatrix}
+ \begin{bmatrix}
B_{\text{DoH}}(t) \\
B_{\text{BP}}(t)
\end{bmatrix}
u(t). \quad (4.10)
\]

In the above equation, DoH and BP subscripts refer to the parameters of the DoH and BP PKPD models, respectively \(x_{1,\text{DoH}} = C_p \in [0, 10 \text{mg/l}], x_{4,\text{DoH}} = C_e \in [1.5 \text{mg/l}, 8 \text{mg/l}], x_{4,\text{BP}} = C^{(\text{BP})}_e \in [0, 4.61 \text{mg/l}] \) (20-39 yr) \& \(\in [0, 4.13 \text{mg/l}] \) (40-59 yr)). In (4.10), we use two different PK models for the DoH and BP PKPD models as Kazama et al. [48] used Gept’s et al. PK model [80] in their identified PKPD model instead of Schneider’s PK model [47]. To calculate the viability kernel for the concatenated PKPD model, the constraint set must be convex and compact. For the sake of viability kernel approximation, we limit the other states of the PKPD models to be \([0, 10 \text{mg/l}]\).

The PKPD models described in section 4.2 depend on the patient’s demographics (e.g. weight) and are subsequently patient dependent. Thus, similar to the safety system discussed in the previous section, the extended safety system and its formalized version are patient dependent. We use convex polytopes to represent the constraint sets. We employ the Multi-Parametric Toolbox 3.0 [82] for viability kernel approximation. For each patient, we formulate the PKPD model based on patient’s demographics and compute the corresponding viability kernel. The calculated viability kernel in all cases is non-empty. We use the calculated viability kernel in the safety-preserving control formulation discussed in section 2.1.3 to maintain the predicted \(C_e, C_p\) and blood pressure within the safety constraints.

Fig. 4.7 shows the block diagram of the simulated closed-loop anesthesia control system which includes the extended safety system. We employ the PID controller proposed in [74] for closed-loop anesthesia. We also include the back-calculation anti-windup strategy as suggested by van Heusden et al. [19]. We include the dynamics of the NeuroSENSE monitor. The setpoint is fixed to 50.
We use (4.10) to predict the states of the BP and DoH PKPD models of each patient. Since the goal is to maintain safety during the maintenance of anesthesia, the PID controller initially starts administering propofol to induce anesthesia while the safety system is inactive. Once the predicted states enter the viability kernel, we activate the extended safety system.

Fig. 4.8 shows the unconstrained responses of the patients. The patients are separated according to the age limits used by Kazama et al. [48] (Group 1: 20-39 year; Group 2: 40-59 year). All patients corresponding to the model set identified in [76] fit within these two age groups. Patients corresponding to each age group share the same BP PD model [48]. According to Fig. 4.8, the safety constraints are violated for several patients in both age groups. The proposed safety system is expected to maintain the states within the safety constraints while allowing the closed-loop controller to sufficiently anesthetize the patient (DoH index of 50 is achieved). Fig. 4.9 illustrates the case when the formalized safety system discussed in the previous section is added to the unconstrained closed-loop anesthesia. This safety system assures that the predicted $C_e$ and $C_p$ remain within the therapeutic window of propofol while blood pressure is unconstrained. According to 4.9 limiting $C_e$ and $C_p$ to the propofol therapeutic window increases the induction time for some patients in group 2; however, the DoH index of 50 is achieved for all the patients. Fig. 4.10 depicts the simulated responses of the 44 patients during closed-loop anesthesia in the presence of the extended safety system. The extended safety system maintains $C_e$, $C_p$ and blood pressure within the safety constraints while the DoH index of 50 is achieved in all cases. Limiting blood pressure
Figure 4.8: The unconstrained simulated responses of the 44 patients during closed-loop anesthesia (Group 1: 20-39 year; Group 2: 40-59 year).
Figure 4.9: The simulated responses of the 44 patients during closed-loop anesthesia with the formalized safety system formalized in Section 4.3 (blood pressure is unconstrained). The dotted red lines specify the upper and lower bounds (Group 1: 20-39 year; Group 2: 40-59 year).
in addition to $C_e$ and $C_p$ further increases the induction time for older patients. According to Fig. 4.10, slow induction cannot be avoided in all cases when blood pressure is constrained in closed-loop propofol anesthesia. Comparing Fig. 4.8 and Fig. 4.10 shows that limiting the blood pressure increases the induction time for group-2 patients whose blood pressure is more sensitive to propofol. Accordingly, the trade-off for maintaining the predicted states within the safety constraints is the longer induction time for this population. Table 4.1 summarizes the average induction time in cases with and without the safety systems.

Table 4.1: Comparing the induction time of anesthesia during closed-loop anesthesia in cases with and without safety systems based on the simulated responses of the 44-patient model set. $T_{ind}$: induction time. $\bar{T}_{ind}$: average induction time.

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (20-39 year)</th>
<th>Group 2 (40-59 year)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of cases</td>
<td>$T_{ind}$ [min-max]</td>
</tr>
<tr>
<td>Without safety system</td>
<td>27</td>
<td>4.59 min [3.43-6.05]</td>
</tr>
<tr>
<td>With safety system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>formalized in [5]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With extended</td>
<td>27</td>
<td>4.65 min [3.47-6.1]</td>
</tr>
<tr>
<td>safety system</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Slow induction of intravenous propofol anesthesia (>10 min) is associated with patient discomfort. It may also cause delays in the anesthesiologists ability to control the airway which compromises patient safety. Moreover, the recommended range of the DoH index during general anesthesia is 40-60 [84]. For some cases in group 2, Fig. 4.10 shows that the DoH index rises above 60 after the induction is completed, which again is not clinically desirable. In these cases, the blood pressure decrease is still significant after the induction of anesthesia. Thus, the extended safety system limits the infusion rate to avoid low blood pressure, which causes the index to go above 60. In current anesthesia practice, to avoid slow induction and low blood pressure at the same time, treatment for hypotension is commonly provided [85]. Similarly, to have a fast induction of anesthesia while avoiding hypotension during closed-loop anesthesia, clinical intervention is likely required for more sensitive patients. The extra intervention could be carried out by a control system. This control system could provide recommendations to anesthesiologists when the intervention is required. Moreover, the system could intervene
Figure 4.10: The simulated responses of the 44 patients during closed-loop anesthesia with the extended safety system. The dotted red lines specify the upper and lower bounds (Group 1: 20-39 year; Group 2: 40-59 year).
directly by providing vasoactive medication in closed-loop based on feedback from a measure of blood pressure. To achieve this, further clinical and technical studies are required on the feasibility of such a multi-drug/multi-feedback closed-loop anesthesia system.

In the proposed formalized safety systems, we employ an open-loop prediction of the patient states to maintain safety. The models we use to predict the states are population-based models and do not necessarily provide an accurate description of patient responses to propofol. To formalize a safety system which maintains safety of the actual patient states, uncertainty in patient responses to propofol must be taken into account.

Clinical scenario

Here, we discuss a realistic clinical scenario based on the recent clinical study reported in [83]. We employ the proposed scenarios in simulation to show the improvement achieved with the extended safety system in terms of patient safety.

This scenario is based on the case discussed at the beginning of this section. Closed-loop induction of anesthesia is simulated for patient #5 within the 44-patient model set. A disturbance is added 5 minutes after the start of the induction. The disturbance lasts for 5 minutes (Fig. 4.11). This simulates prolonged stimulation during airway instrumentation. The simulation length is 30 min. Fig. 4.12 shows the simulated results with and without the extended safety system. In the case without the extended safety system, the constraints on $C_e$, $C_p$ and blood pressure are all violated. In the case when the safety system is included, all the

Figure 4.11: Disturbance profile used in prolonged stimulation scenario.
Figure 4.12: Prolonged stimulation scenario. Dashed line: the closed-loop response without the extended safety system. Solid line: the closed-loop response with the extended safety system in place. The dotted red lines specify the upper and lower bounds.
states remain within the safety constraints while the DoH index of 50 is achieved.

4.5 Conclusion

In this chapter, we formalized the safety system proposed by [19] for closed-loop anesthesia. The formalized safety system ensures that the predicted plasma and effect-site concentrations remain within the therapeutic window at all times. Furthermore, we combined and formalized the safety system with the blood pressure safety system discussed in [51]. The extended safety system is capable of maintaining blood pressure in addition to the plasma and effect-site concentrations within safety constraints. We formalized the safety systems using formal verification techniques. Using patient models in simulation, we illustrated the effectiveness of the formalized safety systems in maintaining the predicted patient states within the safety limits. Using the simulation results we showed that the DoH index of 50 is achieved in all cases in the presence of constraints on $C_e$, $C_p$ and blood pressure; however, limiting these states increases the induction time for older patients. Moreover, we discussed the effectiveness of the formalized safety systems in preserving safety using a realistic clinical scenario.
Chapter 5

Case Study II - Model-Invariant Blood Pressure Safety System for Closed-Loop Anesthesia with Reduced Conservatism

The safety systems discussed previously guarantee safety with regards to the predicted states of patients. In these safety systems, the patient states are predicted using population-based models in an open-loop fashion. Due to model uncertainty and inter-patient variability, maintaining the safety of the predicted states does not yield the safety of the actual patient states. In this chapter, we propose a formalized safety system for closed-loop anesthesia which guarantees that blood pressure of individual patients remains within a safety bound despite model uncertainty. In Chapter 2 we introduced a model-invariant verification technique that can be used to synthesize a safety-preserving controller for systems with multiplicative uncertainty. Given a multi-model description of model uncertainty, a model-invariant safety-preserving controller satisfies constraints for all members of a model-set. Model-invariant safety-preserving control relies on calculation of the model-invariant viability kernel. We showed that the model-invariant viability kernel of an uncertain system is equivalent to the intersection of the viability
kernels of all members of the corresponding model set represented in the form of (2.17). Conservatism of model-invariant safety preserving control depends on the level of model uncertainty. As the uncertainty increases, the size of the model-invariant viability kernel decreases. Moreover, the techniques used for viability kernel approximation may result in further conservatism [27, 62].

Although safety-preserving control techniques rely on the off-line calculation of the viability kernel and control synthesis, the use of online data has been recommended to improve performance of the formalized safety systems. For instance, Gillula et al. [86] paired machine learning algorithms with formal methods and achieved high tracking performance while safety was guaranteed. In [87], they used the same approach to learn disturbances online to reduce conservatism of formalized safety systems. Haesaert et al. [88] and Abate [89] discussed data-driven verification techniques for systems with partially known dynamics. In this work, we employ online measurements to decrease conservatism of model-invariant safety-preserving control using model falsification [90]. Given a set of models which describe an uncertain system, online data can be used to falsify inappropriate members of the model set. This reduces model uncertainty. Model falsification inherently deals with missing data and limited excitation. If falsification criteria are not met, no model will be falsified and safety will not be jeopardized. In the context of model-invariant safety-preserving control, as members of the model set are falsified, the model-invariant viability kernel can be recalculated online as the intersection of the viability kernels of unfalsified models. This yields less conservative model-invariant safety-preserving control if the viability kernels of the falsified models are more restrictive compared to the viability kernel of the true model.

The main contribution in this chapter is to formalize a novel model-invariant safety system for closed-loop anesthesia with reduced conservatism using model falsification. We use a model set identified in [7], which illustrates the effect of propofol infusion on blood pressure of 10 at-risk patients. We employ the model-invariant verification technique and formalize a blood pressure safety system for this population. The model-invariant formalized safety system ensures that blood pressure of patients remains within a safety bound despite model uncertainty. In the next step, we employ blood pressure measurements to falsify members of the
model set to decrease conservatism. Due to the lack of excitation, blood pressure measurements cannot be used to identify the true model of each patient. However, using clinical data, we show that the blood pressure data can be used to falsify irrelevant models.

As mentioned in the previous chapter, it is unknown whether closed-loop anesthesia can provide sufficient anesthesia while avoiding hypotension for all patients, especially at-risk patients. Van Heusden et al. [8] have recently identified a set of models which relate propofol infusion rates to depth of hypnosis for the same population discussed in [7]. This model set enables us to study the feasibility of sufficient anesthesia during closed-loop anesthesia in the presence of the blood pressure model-invariant safety system. This model set also helps us to better demonstrate conservatism of the formalized model-invariant safety system and the improvement we achieve in terms of performance when falsification is included. This chapter demonstrates a proof-of-concept for combining the model-invariant safety-preserving control technique with model falsification to decrease conservatism. The results discussed in this chapter are limited to the above-mentioned model set.

5.0.1 Assumptions

Here, we aim to provide a safety system for blood pressure and formalize it based on the following assumptions:

A1) The PKPD models of individual patients are unknown.
A2) A population-based PK model ($G_{PK}(s)$) is available.
A3) Individual patient’s PKPD models are described by

$$G_{PKPD}(s) = G_{PK}(s)G_{PD}(s),$$

where $G_{PK}(s)$ is known and the blood pressure PD model is known to be part of a finite set of models:

$$G_{PD}(s) \in \mathcal{M}_{PD} = \{G_i(s) | i = 1, \ldots, p\},$$

(5.1)
A4) A noisy measure of blood pressure is available.
A5) Outliers are removed from the measurements.
A6) The measurement noise is bounded and an upper limit of the bound is known.
A7) The PK states are not measurable.
A8) The PK states can be predicted at all times.
A9) The BP from all PD models in the finite set can be simulated and predicted at all times.

In this chapter, we employ the blood pressure PKPD model set identified in [7] to represent uncertainty in the blood pressure response of patients to propofol, and to formalize a model-invariant safety system. This safety system requires feedback from all patient states. Since the PK states are not measurable, a population-based model is formulated based on the demographics of each patient to predict the plasma concentration and the other states of the PK model [47]. The state of the PD model can be predicted for all models in the uncertainty set. In addition, a noisy measure of blood pressure is available to the safety system to be used to falsify irrelevant members of the PD model set to reduce conservatism of the safety system.

5.1 Blood-Pressure Model Falsification in Propofol Anesthesia

The concept of model falsification was introduced in the context of model validation for robust control [90]. Given an a priori nominal model and uncertainty description, the validation problem was cast as a falsification problem; if measured time-domain data is inconsistent with the nominal model and uncertainty bounds, the model is falsified (or invalidated). This methodology uses the philosophical principle that a scientific theory can never be proven to be true, but false hypotheses can be falsified by observations [91].

The falsification concept has been used in, for example, data-driven control [92, 93], robust adaptive control [94], and multi-model switching control [95]. In biomedical applications, controller falsification has been proposed for control of neuromuscular blockade [96]. If a finite number of models/controllers is consid-
ered, falsification requires verification of consistency with data for each entry. This approach can be computationally intensive if the initial model/controller space is large or gridding is fine. An analytical approach to controller falsification has been proposed [93] to reduce the computational load and to extend the methodology to an infinite set of candidate controllers.

### 5.1.1 Blood-pressure model falsification

In this section, we formulate a falsification policy which will be used to falsify members of the BP model set identified in [7] assuming A1-A9 of section 5.0.1.

Suppose measured blood pressure $BP(t)$ is generated as follows:

$$BP(t) = \mathcal{L}^{-1}\{G_{PK}(s)G_{PD}(s)u(s) + n(s)\},$$  \hspace{1cm} (5.2)

where $u(s)$ is the propofol infusion and $n(s)$ describes the bounded measurement noise. $\mathcal{L}^{-1}\{\cdot\}$ is the inverse Laplace transform. $G_{PK}(s)$ is known, $G_{PD}(s)$ is unknown and satisfies the assumption in equation (5.2). For each model $G_i(s)$ in the uncertainty set, a corresponding blood pressure can be predicted:

$$\hat{BP}_i(t) = \mathcal{L}^{-1}\{G_{PK}(s)G_i(s)u(s)\}.$$  \hspace{1cm} (5.3)

The following policy can then be used for model falsification.

**Falsification policy**: $G_i(s) \in \mathcal{M}_{PD}$ is falsified if

$$|BP(t) - \hat{BP}_i(t)| > \gamma,$$  \hspace{1cm} (5.4)

where $\gamma$ is defined as

$$\gamma = \max\{|n(t)| \mid \forall t\}.$$  \hspace{1cm} (5.5)

The assumptions that the patient’s BP model exists in the model set and $\gamma$ is known, guarantee that the patient’s true BP model is never falsified. For the BP PD model set, $\gamma$ was identified as the maximal error between the measurements and
Figure 5.1: Examples of model falsification using clinical data; blue lines: responses of unfalsified models; dashed red line: responses of falsified models; black line: measured BP.

model prediction for the 10 patients:

\[
\gamma = 17\%.
\] (5.6)

The assumption that the patient’s model is a member of the model set may not be realistic in practice, as no a priori knowledge of the true patient model exists prior to induction of anesthesia. However, to demonstrate the proof-of-concept, this assumption guarantees that not all members of the model set are falsified. Fig. 5.1 shows two examples where clinical data was used to falsify members of the BP model set. According to Fig. 5.1, the first few minutes of measurement data contain sufficient information to falsify irrelevant models.
5.2 Falsified Robust Safety-Preserving Control of Propofol Anesthesia

While \textit{a priori} information is insufficient to establish which individual model $G_i(s)$ is controlled, data collected during operation can be used to reduce model uncertainty and consequently conservatism introduced by model-invariant safety-preserving control.

This scheme takes advantage of the characteristics of falsification, where limited excitation and missing data are dealt with naturally. If the data does not contain sufficient information to determine consistency with the model, it cannot be falsified. While in this situation falsification may not reduce conservatism, it does not affect the performance of model-invariant safety-preserving control. When data is missing due to, for example, sensor faults, there is no information to determine consistency and again, no model can be falsified. These characteristics are particularly important in biomedical applications, where excitation is limited and missing data is not uncommon.

To illustrate the effectiveness of the proposed method, closed-loop anesthesia is simulated using three safety systems:

1. Closed-loop anesthesia with an individualized safety system. We formalize the individualized safety system assuming the patients models are known. This scheme results in minimum conservatism as we assume no model uncertainty. Although this assumption does not comply with the assumption of Section 5.0.1, we employ this safety system as a benchmark to discuss conservatism of the model-invariant safety-preserving control technique.

2. Closed-loop anesthesia with a model-invariant safety system. This safety system which guarantees safety for all patients within the model set, can be implemented in practice, but introduces significant conservatism.

3. Closed-loop anesthesia with a model-invariant safety system using falsification. This safety system includes feed-back from the measured blood pressure to falsify models. This system can be implemented in practice and it will be shown that it significantly reduces conservatism compared to the
model-invariant safety system and nearly matches the individualized safety system.

The recommended range of the DoH index during general anesthesia is 40-60 [84]. The index of 50 is typically targeted in closed-loop anesthesia [78]. However, it is unknown if the index of 50 is achievable for all patients including at-risk patients in the presence of blood pressure constraints. Here, we study the feasibility of sufficient anesthesia (DoH index of 50) in the presence of various blood pressure constraints.

5.2.1 Simulation results

Closed-loop anesthesia with an individualized safety system

Here, we formalize a blood pressure safety system for each patient assuming:

1. The BP (and DoH) PKPD model is known for each patient.
2. All states of the PKPD model are measurable or predicted.
3. There is no measurement noise.

The above assumptions are sufficient to formalize an individualized safety system with minimum conservatism. However, existing methods to approximate the viability kernel can result in a certain level of conservatism. Maiden et al. in [27] compared conservatism of these methods. Here, we employ convex polytopes to represent constraint sets. We use the Multi-Parametric Toolbox 3.0 [82] to conduct operations on polytopes to calculate the viability kernel.

We employ the set of 9 BP PKPD models identified in [7] for which van Heusden et al. [8] identified DoH PKPD models. We use the BP PKPD model of each patient to formalize a blood pressure safety system for that specific patient. We employ the corresponding DoH PKPD models to simulate the response of each patient during closed-loop anesthesia and in the presence of the BP safety system. We employ the safety constraints proposed in [19] which limit the PK states to the therapeutic window of propofol:

\[ C_{p}, x_2, x_3 \in [0, 10\text{mg/l}] \]  

(5.7)
Figure 5.2: Model-invariant viability kernel of the PKPD model set projected onto BP-$C_P$ space. Green region: model-invariant viability kernel; red region: safe region; blue lines polytopes: the viability kernels of individual patients.

Figure 5.3: Block diagram of closed-loop anesthesia with the individualized safety system.

The following constraint on the infusion rate of propofol is also suggested in [19]:

$$u(t) \in [0, 600\text{ml/h}].$$  \hspace{1cm} (5.8)
We discuss three different constraints on BP:

\[ \text{BP drop} \leq 30\%, 40\%, 50\%. \quad (5.9) \]

We employ the PID controller robustly tuned by Dumont et al. [74] to achieve the closed-loop goal. We include the back-calculation anti-windup suggested by van Heusden et al. [19] to improve the performance of the PID controller when the safety constraints are active. For each patient, we formalize the safety system to guarantee that the states of the PKPD model remain within the safety constraints. Fig. 5.2 illustrates the viability kernels of the patient models. Fig. 5.3 illustrates the block diagram of the system implemented in simulation.

Fig. 5.4 shows the box plot of the DoH index of patients with different constraints on BP decrease at \( t = 20\) min following the start of propofol infusion. Induction of anesthesia cannot be completed for all cases when the BP decrease is limited to 30\% and 40\%. In contrast, when the BP decrease is limited to 50\%,
Figure 5.5: The closed-loop responses of the patients with BP decrease limited to 50% (individualized safety system).

induction of anesthesia can be completed in almost all cases. Fig. 5.5 shows the closed-loop responses of the patients with the bound on BP decrease at 50%.

Closed-loop anesthesia with a model-invariant safety system

Previously, we showed that all patients could achieve a DoH index between 40 and 60 when a 50% decrease in blood pressure is allowed. In this section, we formalize
a model-invariant viability kernel for the same population based on the assumption discussed in section 5.0.1.

We assume the PK model of each patient is known [47]. We assume the patient’s BP PD model is unknown, however it is part of the model set identified in [7]. We calculate the model-invariant viability kernel according to this multi-model uncertainty description.

We limit blood pressure decrease to be less than 50% and use the constraints defined in (5.7) and (5.8) on the states of the PK model and the infusion rate. The calculated model-invariant viability kernel under the mentioned assumptions is shown in Fig. 5.2. The resulting model-invariant safety-preserving controller requires feedback from all states of the BP model. The patient PK model is assumed to be known and the corresponding states can be predicted. The PD model on the other hand is uncertain, unknown, and cannot be predicted for each patient.

Although we assume a noisy measure of blood pressure is available, the proposed output-feedback safety-preserving control cannot be used here due to the large measurement noise identified for the model set. We therefore employ the worst case prediction of blood pressure decrease as feedback to the model-invariant safety-preserving control. We define the worst case blood pressure decrease $BP_{wc}$ as follows:

$$BP_{wc}^i = \arg\max_{BP_i^j} \{|BP_i^j| \mid i = 1, \ldots, 9\}, \quad (5.10)$$

where $BP_i^j$ is the blood pressure decrease predicted using the $i^{th}$ member of the BP model set. The model-invariant safety-preserving control maintains the states of all members of the model set within the safe region. According to the definition of the worst case blood pressure decrease, the control input which keeps $BP_{wc}^i$ less than 50%, maintains all $BP_i^j$’s below 50%. Consequently, since we assume that the true blood pressure model of the patients exists in the model set, maintaining the worst case predicted blood pressure decrease below 50% guarantees that the decrease in blood pressure of each patient stays within the safe range. Fig. 5.6 illustrates the block diagram of the system implemented in simulation.

The bottom plot in Fig. 5.7 shows blood pressure decrease of all patients using
Figure 5.6: Block diagram of closed-loop anesthesia with the model-invariant safety system.

The model-invariant safety system. The safety system maintains the blood pressure of all patients within the safe bound. However, the safety system does not allow the closed-loop controller to complete induction of anesthesia for the majority of the population. The model-invariant safety system introduces significant conservatism compared to the individualized safety system (see Fig. 5.10).

Closed-loop anesthesia with a model-invariant safety system using falsification

The model-invariant viability kernel is calculated as the intersection of the viability kernels of all individual members of the model set. Accordingly, if the viability kernel of one of the models is restrictive, the model-invariant safety system restricts the drug infusion for all patients. In Section 5.1, we showed that clinical blood pressure data during the first minutes following the start of propofol infusion contain sufficient information to falsify outliers. By falsifying restrictive models in the model set, and recalculating the model-invariant viability kernel as the intersection of the remaining models, conservatism of model-invariant safety-preserving control can be reduced. If no model is falsified, the model-invariant viability kernel remains unchanged and safety is not compromised.

Here, we employ the model-invariant safety system we formalized in the previous part, but instead of calculating the model invariant viability kernel off-line, we calculate it online. At each sample time, we falsify a member of the model set if the difference between its simulated response and the measured BP is bigger than
Figure 5.7: The closed-loop responses of the patients with BP decrease limited to 50% (model-invariant safety system). The model-invariant safety system does not allow the induction completion in several cases.

a certain threshold. Then, we calculate the model-invariant viability kernel (the intersection) with the viability kernel of the falsified model removed. We specify the threshold according to (5.6). Furthermore, the worst case predicted blood pressure decrease, which is defined in (5.10), is selected from the predictions of the unfalsified models. Fig. 5.8 illustrates the block diagram of the implemented closed-loop anesthesia with the model-invariant safety system including falsification.
Fig. 5.8: Block diagram of closed-loop anesthesia with the model-invariant safety system including falsification.

Fig. 5.9 illustrates the closed-loop responses of the patients when the model-invariant safety system with falsification is in place. Accordingly, induction of anesthesia is completed in all cases and DoH reaches an index between 40-60 for all patients. Comparing Fig. 5.7 and Fig. 5.9 depicts a significant improvement in the performance of the model-invariant safety system when falsification is included. Fig. 5.10 compares conservatism of the safety systems discussed in this work. The individualized safety system maintains safety with minimum conservatism while the model-invariant safety system significantly increases conservatism. However, adding model falsification to the model-invariant safety system significantly decreases conservatism of the model-invariant safety system and brings it to the conservatism level of the individualized safety system.

5.3 Conclusion

We proposed a model-invariant safety system for closed-loop anesthesia that guarantees to maintain a patient’s blood pressure within a safety bound despite inter-patient variability. We suggested the use of model falsification to reduce conservatism introduced by the model-invariant safety-preserving control. The effectiveness of the proposed model-invariant safety-preserving control approach was demonstrated in simulation. Augmenting the model-invariant safety-preserving controller using feedback from noisy blood pressure measurements for model falsi-
Figure 5.9: The closed-loop responses of the patients with BP decrease limited to 50% (individualized safety system including falsification). Falsification reduces conservatism of the model-invariant safety system resulting in the induction completion in all cases.

fication significantly reduces conservatism introduced by model uncertainty. Model-falsification reduces this model uncertainty and subsequently decreases conservatism of the model-invariant safety system.

We presented the proposed safety system as a proof-of-concept of a model-invariant safety system for closed-loop anesthesia which includes model-falsification.
Figure 5.10: Comparing DoH of the patients with different BP safety systems, achieved after 20 min from the start of the closed-loop administration of propofol: (1) the individualized safety system; (2) the model-invariant safety system, (3) the model-invariant safety system including falsification. Results of the model-invariant safety system with falsification nearly matches results of the individualized safety system.

The proposed solution is limited by the assumption that a patient model exist in a known, finite model set, and that all outliers are removed from the data prior to falsification. Generalizing the results of this chapter to cases in which these assumptions are not met requires further research.
Chapter 6

Conclusion and Future Work

6.1 Conclusions

In this work, we proposed a mathematical framework for formal verification and safety-preserving control of systems with model uncertainty with a focus in closed-loop anesthesia. We introduced the model-invariant safety-preserving control technique which guarantees safety systems with a certain class of multiplicative model uncertainty. A safety-preserving controller synthesized using this technique maintains the states of all instances of an uncertain system within a set of viability constraints.

Furthermore, we introduced one of the first safety-preserving control schemes for output-feedback control systems. Called output-feedback safety-preserving, this method provides a guarantee of safety for systems in which safety-critical states are not measurable by maintaining the estimated states within a pre-calculated set of constraints. We also discussed the proposed output-feedback safety-preserving control scheme in the presence of model uncertainty.

Finally, we formalized safety systems for closed-loop anesthesia. First, we discussed and formalized the existing safety systems in the literature. These safety systems specify constraints on patient states predicted using population-based models in an open-loop fashion. These safety systems do not consider model uncertainty and inter-patient variability. Accordingly, maintaining the predicted states within a set of constraints does not guarantee that the actual patient states re-
main within the constraint set. In the last step, we employed the proposed model-invariant safety-preserving technique and formalized a safety system with guarantees that patient blood pressure remains within a safe bound during closed-loop anesthesia despite inter-patient variability. The model-invariant safety-preserving control scheme introduces conservatism as it inherently deals with model uncertainty. We proposed a novel approach to reduce conservatism using model falsification. In this framework, we use online measurements to decrease model uncertainty based on a falsification policy. We showed that this significantly reduces conservatism of the proposed model-invariant safety-preserving control technique.

In a nutshell, the main contributions of this work are:

- Introducing a model-invariant safety-preserving control technique which extends existing safety-preserving control techniques to cases with multiplicative model uncertainty.

- Introducing an output-feedback safety-preserving control technique which extends existing safety-preserving control techniques to uncertain output-feedback control systems.

- Formalizing existing safety systems for closed-loop anesthesia.

- Formalizing a model-invariant safety system for closed-loop anesthesia with reduced conservatism using model falsification.

### 6.2 Future Work

Our proposed safety-preserving control techniques deal with uncertainty in the state space matrices. In this work, no time delay is considered in the models. If there is no model uncertainty, the Smith predictor [97] can be employed to predict the states and safety-preserving controllers can be modified accordingly [2]. If the state space model matrices or time delay are uncertain, any prediction is perturbed by error. To the best of our knowledge, safety-preserving control in the presence of time delays has not been discussed in the literature. Because not all safety critical control systems are delay free, discussion of safety verification in the presence of time delays requires further research.
The model-invariant safety-preserving control technique relies on a multi-model description of model uncertainty. The existing models for closed-loop anesthesia are population-based and describe average responses of patients to anesthetics. Uncertainty is not specified for these models. To formalize a safety system for closed-loop anesthesia that guarantees safety in spite of inter-patient variability, a comprehensive set of patient models needs to be identified to describe uncertainty. This model set should include extreme patient responses.

In the last two chapters of this thesis, we formalized safety systems which limit patient blood pressure to a safety bound during closed-loop anesthesia. We showed that although patients achieve a desired level of anesthesia in the presence of blood pressure constraints, the induction time is increased to avoid hypotension. A prolonged induction of anesthesia is not desirable in operating rooms as it delays the start of operations and may result in patient discomfort. In manual anesthesia practice, hypotension treatment is provided to avoid low blood pressure. Similarly, this clinical intervention can be done in closed-loop to achieve fast induction of anesthesia while avoiding hypotension. To achieve this, further clinical and technical research is required.
Bibliography


94


95


98