

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Risk-Sensitive Safety Analysis and Control for Trustworthy Autonomy

Permalink

<https://escholarship.org/uc/item/7p5575zg>

Author

Chapman, Margaret P

Publication Date

2020

Peer reviewed|Thesis/dissertation

Risk-Sensitive Safety Analysis and Control for Trustworthy Autonomy

by

Margaret P. Chapman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Claire Tomlin, Chair

Professor Murat Arcak

Associate Professor Anil Aswani

Spring 2020

Risk-Sensitive Safety Analysis and Control for Trustworthy Autonomy

Copyright 2020
by
Margaret P. Chapman

Abstract

Risk-Sensitive Safety Analysis and Control for Trustworthy Autonomy

by

Margaret P. Chapman

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Claire Tomlin, Chair

Methods for managing dynamical systems typically invoke one of two perspectives. In the *worst-case* perspective, the system is assumed to behave in the most harmful way; this perspective is used to provide formal safety guarantees. In the *risk-neutral* perspective, the system is assumed to behave as expected; this perspective is invoked in reinforcement learning and stochastic optimal control. While the worst-case perspective is useful for safety analysis, it can lead to unnecessarily conservative decisions, especially in settings where uncertainties are non-adversarial. The risk-neutral perspective is less conservative and useful for optimizing the system's performance on average. However, optimizing average performance is not guaranteed to protect against harmful outcomes and thus is not appropriate for safety-critical applications.

This thesis consists of two parts. First, we present an analytical and computational toolkit for cancer modeling and management that we have developed with cancer biologists by invoking the worst-case perspective. In addition to providing biological insights about breast cancer and theoretical insights about switched systems, this work has motivated the need for new mathematical methods that facilitate less conservative but still protective control of dynamical systems.

Towards this aim, we have devised a risk-sensitive mathematical method for safety analysis that blends the worst-case and risk-neutral perspectives by leveraging the *Conditional Value-at-Risk* measure. The second part of this thesis presents the mathematical development of this risk-sensitive safety analysis method. We also show its practical application to evaluating the safety of urban water infrastructure, using a numerical example that has been developed in collaboration with water resources engineers.

To David Freyberg and Claire Tomlin

Contents

Contents	ii
List of Figures	iii
1 Introduction: Risk-Neutral vs. Worst-Case Decision Analysis Methods for Dynamical Systems	1
1.1 Overview	2
1.2 Notation	3
2 On Modeling and Managing a Cancer Subtype	5
2.1 Data-Driven Modeling of TNBC Dynamics	6
2.2 Controller Synthesis via the Worst-Case Perspective	20
3 Risk-Sensitive Safety Analysis via CVAR	34
3.1 Motivating Example	34
3.2 Mathematical Background	36
3.3 Mathematical Formulation	46
3.4 Numerical Example	60
3.5 Summary	62
4 Future Directions	64
4.1 Risk-Sensitive Control of Cancer Systems	64
4.2 Approximate Risk-Sensitive Safety Analysis	65
4.3 Potential Alternate Approaches to Risk-Sensitive Safety Analysis	66
4.4 Extensions to Risk-Sensitive Safety Analysis	68
4.5 Additional Exciting Future Directions	70
Bibliography	72
A Supplementary Results for Chapter 3	81
A.1 Lemmas 6 and 7	81
A.2 Lemmas 8-10	83

List of Figures

- 2.1 **Therapy-specific dynamical model** [19]. We have used time series data to identify a linear time-invariant model of a triple-negative breast cancer (TNBC) cell population in response to initial treatment. In the model, the live cells occupy four differentiation states and can transition, divide, or die. The dynamics parameters (ρ_{12} , ρ_{21} , ρ_3 , ρ_{4D} , etc.) are the average rates of these actions taken by the live cells in each differentiation state following initial treatment. 10
- 2.2 **Therapy-specific transition gains** [19]. For each treatment condition, values of the transition gains from the AM-optimized dynamics matrix are shown in units of $\frac{\# \text{ cells at } (k+1) \text{ multiples of 12 hours}}{\# \text{ cells at } k \text{ multiples of 12 hours}}$. Each transition gain from differentiation state i to differentiation state j of sufficient magnitude ($\rho_{ij} \geq 0.10$) is depicted as an arrow directed from i to j . Arrow style specifies gain magnitude. A dotted arrow means $\rho_{ij} \in [0.10, 0.30)$, a dashed arrow means $\rho_{ij} \in [0.30, 0.70)$, and a solid arrow means $\rho_{ij} \in [0.70, 1.00]$ 13
- 2.3 **95% confidence intervals for the dynamics parameters** [19]. These intervals indicate the variability of the dynamics parameters with respect to measurement error. Non-overlapping intervals of a given parameter specify a statistically significant difference. For example, a statistically significant reduction in K14^{hi}-to-VIM^{hi}K14^{low} transition has been detected under Trametinib versus DMSO because the ρ_{12} -interval for Trametinib is strictly below the ρ_{12} -interval for DMSO. Recall that ρ_{12} is the K14^{hi}-to-VIM^{hi}K14^{low} transition gain. 14
- 2.4 **Ensemble model predictions in comparison to test data** [19]. The test samples (black stars) and ensemble model predictions (gray bands) are shown for each treatment condition: DMSO (row 1), Trametinib (row 2), BEZ235 (row 3), and Trametinib+BEZ235 (row 4). The model ensemble is a collection of models that have been identified from the training data via the resampling residuals bootstrap algorithm [108] for each treatment condition. In each plot, we show a 95% confidence interval (gray band) around the median (black dotted line) of the ensemble model predictions. *Higher* p-values indicate *better* consistency between predictions and test data over the time horizon (12h, 24h, . . . , 60h). 16

- 2.5 **Single model predictions in comparison to test data**, where the differentiation states are defined by K14 only [19]. The test samples (black stars) and single model predictions (pink bands) are shown for each treatment condition: DMSO (row 1), Trametinib (row 2), BEZ235 (row 3), and Trametinib+BEZ235 (row 4). The single model was identified on the training data using K14^{hi} and K14^{low} as the differentiation states for each treatment condition. The pink band extends between the maximum prediction and the minimum prediction out of four predictions in total at each time point (0h, 12h, ..., 72h). The dotted line indicates the median of the predictions. *Higher* p-values indicate *better* consistency between predictions and test data over the time horizon (12h, 24h, ..., 60h). 17
- 2.6 **Further investigations of Trametinib-induced K14^{hi} enrichment hypothesis** [19]. Left: Trametinib K14^{hi} live cell predictions by the AM-optimized dynamics matrix (pink band) are shown in comparison to test data (black stars). Right: Trametinib K14^{hi} live cell predictions by a dynamics matrix identified with additional constraints (pink band) in comparison to test data (black stars). The additional constraints are $\rho_{12} \geq 0.59$, the value of ρ_{12} for DMSO, and $\rho_{31} \leq 0.19$, the value of ρ_{31} for DMSO. Recall that ρ_{12} is the K14^{hi}-to-VIM^{hi}K14^{low} transition gain, and ρ_{31} is the K19^{hi}VIM^{low}K14^{low}-to-K14^{hi} transition gain. In each plot, the pink band extends between the maximum prediction and the minimum prediction out of four predictions in total at each time point (0h, 12h, ..., 72h). The dotted line indicates the median of the predictions. Higher p-values indicate better consistency between predictions and test data over the time horizon (12h, 24h, ..., 60h). 19
- 2.7 **A proposed treatment strategy**. The images of the cancer cells are from [84]. 20
- 2.8 A cyclic schedule of three therapies, $\mathcal{D} := \{\diamond, \square, \triangle\}$, with the waiting times $(k_\diamond, k_\square, k_\triangle)$ is shown for two cycles [20]; see (2.9). For example, $l_{32} = \square$ means that therapy $\square$ is the third therapy applied in cycle 2, and the next therapy will be applied $k_\square$ time points later. Notice that each therapy in $\mathcal{D}$ is applied once per cycle, and the order of the therapies in each cycle may vary. 26
- 2.9 A simulation of the switched dynamical system (2.17) for a cancer cell population in response to a cyclic schedule of the therapies $\{P, B, T\}$ with the waiting times (k_P, k_B, k_T) for 40 cycles (160 days) [20]. Each vertical grid line denotes the start of a cycle. **Top**: Normalized population size $\frac{\|x(t)\|_1}{\|x_0\|_1}$ is shown at each time t of the application of a therapy. E.g., in the first cycle, if the order were T first, B second, and P third, then $\frac{\|x(t)\|_1}{\|x_0\|_1}$ would be plotted at $t \in \{0, k_T, k_T + k_B, k_T + k_B + k_P\}$. **Middle**: The error $\xi_{\delta_t}(t)$ is shown at each time point $t \in \{0, 1, \dots, 319\}$. There are 320 time points in total since $40 \text{ cycles} \times \frac{8 \text{ time points}}{\text{cycle}} = 320$. The duration of each discrete time interval $[t, t+1)$ is 0.5 days. **Bottom**: The error product $\prod_{t=T_{m-1}}^{T_m-1} \xi_{\delta_t}(t)$ per cycle $m \in \{1, 2, \dots, 40\}$ is shown. Note that each error product is less than or equal to $\frac{1}{\eta}$ 32

3.1	A stormwater catchment system in Lenexa, Kansas that consists of two ponds in series connected by a stream [21]. This system is currently operated by OptiRTC, Inc. (Boston, MA).	35
3.2	HJ safe sets for three different design choices when there is constant surface runoff of 2 cubic feet per second into each pond over a 4-hour time horizon [21]. In each plot, each point in the volume represents an initial condition from which the system is predicted to avoid overflows. The x_1 -axis is the water elevation of pond 1; the x_2 -axis is the water elevation of the stream; and the x_3 -axis is the water elevation of pond 2. A reference vector that points approximately in the (5 ft, 3 ft, 3.5 ft)-direction is shown with each safe set. 5 feet, 3 feet, and 3.5 feet are the maximum water elevations of pond 1, the stream, and pond 2, respectively. An estimate of the distance that each safe set extends along this vector is provided.	36
3.3	A standard surface runoff profile for pond 1 (r_1) and pond 2 (r_2) over a 4-hour time horizon [21]. Surface runoff into each pond has been estimated using the two-year rainfall event for Lenexa, Kansas with the USEPA Stormwater Management Model for various model parameter sets via PCSWMM (Computational Hydraulics International, Ontario) [85].	37
3.4	HJ safe sets for very large facilities using the standard surface runoff profiles (r_1 , r_2) over a 4-hour time horizon [21]. The surface runoff profiles are shown in Fig. 3.3. A reference vector that points approximately in the (5 ft, 3 ft, 3.5 ft)-direction is shown with each non-empty safe set. 5 feet, 3 feet, and 3.5 feet are the maximum water elevations of pond 1, the stream, and pond 2, respectively. An estimate of the distance that each non-empty safe set extends along this vector is provided.	38
3.5	Conditional Value-at-Risk (CVAR) is a measure of one-sided tail risk of a random variable Y . Intuitively, $\text{CVAR}_\alpha(Y)$ can be interpreted as the expectation of the $\alpha \cdot 100\%$ worst realizations of Y . More specifically, if Y is a continuous cost random variable, then $\text{CVAR}_\alpha(Y)$ is the expectation of Y conditioned on the event $\{Y \geq \text{VAR}_\alpha(Y)\}$, where $\text{VAR}_\alpha(Y)$ is the minimum cost in the $\alpha \cdot 100\%$ worst cases.	39
3.6	Probability distributions for the disturbance d_t to the TCL system (3.3). The horizontal axis shows the possible (discrete) realizations of the disturbance in degrees Celsius. Their associated probabilities are shown on the vertical axis. . .	42

- 3.7 Histograms of $C_{0:T} = \sum_{t=0}^T \max(x_t - 21, 20 - x_t)$ for the TCL system (3.3) initialized at $x_0 = 20.1$ °C are shown under different control policies. **Top row:** A control policy has been computed with respect to CVAR_α . **Bottom row:** A control policy has been computed with respect to $\rho_{e,\theta}$ (3.1). **Left column:** Nearly risk neutral $(\alpha, \theta) = (0.99, -0.01)$. **Center column:** Typical risk aversion $(\alpha, \theta) = (0.05, -0.95)$. **Right column:** Nearly worst case $(\alpha, \theta) = (0.01, -0.99)$. One million samples of $C_{0:T}$ are shown in each histogram, and relevant empirical statistics of these samples are displayed. The red circle marks the empirical mean, and each green circle marks the empirical mean plus/minus one/two standard deviations. The same pseudorandom sequence of disturbance realizations has been used for each histogram and chosen according to the “Original” distribution in Fig. 3.6. 43
- 3.8 Same setting as in Fig. 3.7 except that $x_0 = 20.5$ °C. 44
- 3.9 Histograms of $C_{0:T} = \sum_{t=0}^T \max(x_t - 21, 20 - x_t)$ are shown for the TCL system (3.3) initialized at $x_0 = 20.1$ °C under a perturbed probability distribution for the disturbance. For each histogram, the same pseudorandom sequence of disturbance realizations has been used to sample $C_{0:T}$ via the “Perturbed” distribution shown in Fig. 3.6. Control policies have been computed using the “Original” distribution for the disturbance shown in Fig. 3.6. **Left:** A control policy has been computed with respect to CVAR_α for $\alpha = 0.05$. **Right:** A control policy has been computed with respect to $\rho_{e,\theta}$ for $\theta = -0.95$ (3.1). One million samples of $C_{0:T}$ are shown in each histogram, and relevant empirical statistics of these samples are displayed. The red circle marks the empirical mean, and each green circle marks the empirical mean plus/minus one/two standard deviations. 45
- 3.10 A non-linear two-tank stormwater system with an automated valve, where water flows by gravity between the tanks and can flow in either direction. 62
- 3.11 Approximate contours of $\mathcal{U}_\alpha^r$ and $\mathcal{S}_\alpha^r$ (denoted by $\partial\hat{\mathcal{U}}_\alpha^r$ and $\partial\hat{\mathcal{S}}_\alpha^r$, respectively) are shown for $\alpha \in \{0.99, 0.05, 0.01\}$ and $r \in \{1.25, 1.5\}$ for the non-linear stormwater system model. These results demonstrate that risk-sensitive safety analysis can be applied to a non-linear dynamical system. $\hat{\mathcal{U}}_\alpha^r$ provides an under-approximation of $\hat{\mathcal{S}}_\alpha^r$ when $\beta e^{\gamma r}$ is sufficiently large, although Theorems 4 and 5 have been proven for linear dynamical systems. Approximations of $J_0(\cdot, \alpha)$ and $W_0^*(\cdot, \alpha)$ are also provided. $\mathcal{U}_\alpha^r$ is the $\beta e^{\gamma r}$ -sub-level set of $J_0^*(\cdot, \alpha) = J_0(\cdot, \alpha)$ with $(\beta, \gamma) = (2 \cdot 10^{-11}, 13)$. $\mathcal{S}_\alpha^r$ is the r -sub-level set of $W_0^*(\cdot, \alpha)$ 63

Acknowledgments

I greatly thank my thesis advisor Professor Claire Tomlin and my long-term mentor Professor David Freyberg for supporting the natural evolution of my research interests and for making a wonderful positive impact in my life. I will do my very best to pass this positive impact forward to my future students with the patience, wisdom, and grace that Professor Tomlin and Professor Freyberg have shown me.

I appreciate Professor Murat Arcaak for his guidance about framing the risk-sensitive safety analysis work, for his support of my research career, and for his review of this thesis. I would like to thank Thomas Low for his friendship, for our many wonderful conversations over the years, and for sharing his inspiring research ideas. I thank Professor David Sedlak for sharing his hydrology expertise, for welcoming me into the ReNUWIt NSF Engineering Research Center, and for his kind support of my diverse research interests. I am grateful to Professor Ole Hald for patiently teaching me real analysis, for his helpful advice about the importance of “playing around with math,” and for his mathematical insight that led to the simplification of a matrix norm in the drug scheduling analysis paper. I also thank Professor Joe Gray for his support of my cancer modeling efforts and for sharing his idea on an interesting future research direction.

I am especially grateful to my research collaborators. The mathematical modeling and control of triple-negative breast cancer is joint work with Dr. Tyler Risom, Professor Ellen Langer, Professor Rosalie Sears, Professor Anil Aswani, Eric Mazumdar, Professor Roel Dobbe, and Professor Claire Tomlin. I thank Dr. Tyler Risom for spending many hours on the collection and analysis of a novel biological time series data set, for his interest in biological systems modeling, and for his constructive feedback and biological insights. I give special thanks to Professor Ellen Langer and Professor Rosalie Sears for their generous support of my research trajectory, for sharing their systems biology expertise, for their important biological interpretations of our modeling predictions, and for their informative feedback on our paper. I thank Professor Anil Aswani for his guidance on statistical analysis and system identification that greatly facilitated our cancer modeling contributions, for the advice that he has kindly provided over the years, for his support of my research career, and for his review of this thesis. Moreover, I thank Eric Mazumdar for our discussions that led to the completion of the asymptotic stability proofs, Professor Roel Dobbe for his mentorship at the start of my research journey, and Professor Claire Tomlin for her patient guidance on the development of the linear time-invariant models, for being a wonderful teacher and role model, and for her review of this thesis.

The risk-sensitive safety analysis contribution and the related stormwater systems examples are joint work with Kevin Smith, Jonathan Lacotte, Professor Insoon Yang, Yuxi Han, Professor Aviv Tamar, Donggun Lee, Victoria Cheng, Professor Jaime Fisac, Dr. Susmit Jha, Professor Marco Pavone, Professor Claire Tomlin, and Professor David Freyberg. I would like to give special thanks to Kevin Smith, who is an absolute joy to work with. I thank Kevin Smith for teaching me about stormwater management, for his resilience when our projects stalled, for his patience with running code on a high-performance computing

cluster, for his software contributions, for his careful attention to detail, and for his willingness to explain and share his novel ideas. I appreciate Jonathan Lacotte for the many hours that we spent together deriving mathematical proofs and for his technical insights. Furthermore, I thank Professor Insoon Yang for sharing his expertise on measure theory, topology, and functional analysis, for his future direction idea, and especially for his clear and thoughtful explanations of challenging mathematical concepts. I thank Yuxi Han for her friendship and for patiently answering my questions about functional analysis. I would like to thank Professor Aviv Tamar and Dr. Sumeet Singh for helpful discussions about risk measures. I appreciate Donggun Lee for sharing mathematical insights and Victoria Cheng for her assistance on software implementations. I thank Professor Jaime Fisac for offering several practice preliminary exams and problem-solving strategies, for discussions about Hamilton-Jacobi reachability, and for his creative insight to combine reachability theory with risk measures. I would like to thank Dr. Susmit Jha for a valuable summer internship experience at SRI International and for his advice to study Conditional Value-at-Risk. I thank Professor David Freyberg for his idea to develop a new design methodology for stormwater systems that better quantifies precipitation uncertainty and risk and for his encouragement of my research journey over the years. I am grateful to Professor Marco Pavone for sharing his knowledge of risk measures, for his detailed feedback on our risk-sensitive safety analysis paper, and for his support of my career trajectory.

I gratefully acknowledge fruitful discussions with Dr. Walid Krichene, Dr. Max Balandat, Sylvia Herbert, Zhiyan Ding, Michael Lim, Palak Bhushan, Professor Laura Heiser, Professor James Korkola, Professor Mo Chen, and Professor Alessandro Abate.

In addition, I would like to thank the many professors who provided valuable insights about my current and future research directions in the winter and spring of 2020. In particular, I thank Professor Pamela Allen for teaching me about Chronic Lymphomatic Leukemia, Professor Tamer Başar for introducing me to Large Deviations Theory, and Professor Xunyu Zhou for suggesting potential alternate approaches to solve the CVAR-MDP problem. I give thanks to Professor John Tsitsiklis for sharing his ideas about the key open areas in the field of systems and control.

I am grateful for my funding sources, which include a National Science Foundation (NSF) Graduate Research Fellowship, a Berkeley Fellowship for Graduate Study, a Tau Beta Pi Engineering Honors Society fellowship (Williams No. 35), and NSF research grants (NSF PIRE UNIV59732 and NSF CPS 1740079). Moreover, I acknowledge support provided by the National Institutes of Health Center “Systems Biology of Collective Cell Decisions” through Stanford University and the National Cancer Institute (NCI) CSBC consortia “Model-Based Predictions of Responses to RTK Pathway Therapies” through OHSU NCI #U54CA112970 and “Measuring, Modeling and Controlling Heterogeneity” through OHSU NCI #1U54CA209988-01A1.

I give very special thanks to my dear friends, including Patricia Hidalgo-Gonzalez, Stefanie Garcia, Sylvia Herbert, Oladapo Afolabi, Yuxi Han, Jaimie Swartz, Emily Harris, Kathryn Vanderboll, David Fridovich-Keil, Vicenç Rubies Royo, Miriam Connor, Julie Tibshirani, Dr. Robert Matthew, Lea Barrow, and Quinn Price, for their emotional support,

kindness, and compassion and for bringing me incredible joy. Lastly, I would like to express great appreciation to Michael Nowak and to my family for providing wonderful care, nourishment, and love.

Chapter 1

Introduction: Risk-Neutral vs. Worst-Case Decision Analysis Methods for Dynamical Systems

Methods for managing dynamical systems typically invoke one of two perspectives: risk-neutral or worst-case. In the *risk-neutral* perspective, the dynamical system is assumed to behave as expected over time. Examples of risk-neutral methods include: stochastic optimal control [90] [58] [10] [46] [62] [80], stochastic reachability [1] [81], and reinforcement learning (stochastic adaptive control) [99] [98] [11].

On the other hand, in the *worst-case* perspective, the dynamical system is assumed to behave in the most harmful way over time. Examples of worst-case methods include: robust control [8] [14] [77] [116] [91] [74], Hamilton-Jacobi reachability [70] [33] [23], and minimax reachability [9] [12]. Additional worst-case methods are the analysis and computation of reachable or invariant sets via sums of squares programming [67], linear and second-order cone programming [66], geometric control [86] [87] [64], or linear matrix inequalities [75]. We have used the worst-case perspective to design and analyze cancer treatment schedules, which we present in Chapter 2 of this thesis.

The worst-case and risk-neutral decision analysis methods for dynamical systems have distinct advantages and disadvantages. Worst-case methods are used to provide formal safety guarantees for dynamical systems in the presence of uncertainty. While the worst-case perspective is useful for safety analysis, it can lead to unnecessarily conservative decisions, especially in settings where uncertainties are non-adversarial. The risk-neutral perspective is less conservative and useful for optimizing the performance of dynamical systems on average. However, optimizing average performance is not guaranteed to protect against harmful outcomes and thus is not appropriate for safety-critical applications.

There is a clear gap between the worst-case and risk-neutral perspectives, and some methods have been devised to help fill this gap. Examples include methods for the analysis and control of dynamical systems that are based on minimizing state-trajectory costs with respect to an assumed family of probability distributions [102] [110] or minimizing state-

trajectory costs that are assessed using risk measures [25] [44] [79] [89] [92] [93] [102].

In this thesis, we present a new safety analysis method for dynamical systems that further fills the gap between the worst-case and risk-neutral perspectives. This method, which is called *risk-sensitive safety analysis*, uses the *Conditional Value-at-Risk* measure to evaluate the safety of a given dynamical system with respect to the $\alpha \cdot 100\%$ worst-case outcomes of the state trajectory.¹ The development of risk-sensitive safety analysis is motivated by the conservativeness that arises by applying standard worst-case methods to safety-critical dynamical systems under large uncertainties. Examples of such systems include a cancer treatment system, which we model and analyze in Chapter 2, and a stormwater catchment system, which we model and analyze in Chapter 3. The development of risk-sensitive safety analysis is also motivated by the limitation of the standard *Exponential Utility* measure to quantify the $\alpha \cdot 100\%$ worst cases in settings with non-Gaussian state-trajectory costs. We examine the differences between using the Conditional Value-at-Risk measure and the Exponential Utility measure for optimal control of dynamical systems through a numerical example in Chapter 3.

This thesis has inspired many interesting future research directions, including risk-sensitive control of cancer systems, approximate risk-sensitive safety analysis for scaling to high-dimensional systems, and the generalization of risk-sensitive safety analysis to a larger class of risk measures. These and other promising ideas are presented in the final chapter of this thesis.

1.1 Overview

The structure of this thesis is as follows. The next section of the Introduction presents mathematical notation that is used throughout the thesis. Then, Chapter 2 presents a body of research on modeling and managing a cancer subtype called *triple-negative breast cancer*. Most of the content in Chapter 2 is from the previously published papers [20] and [19]. As practical control theorists, we are interested in improving how triple-negative breast cancer (TNBC) is managed, but there are no standard models to predict how this cancer subtype evolves over time. Thus, the first part of Chapter 2 is devoted to the study of new TNBC dynamical models that we have developed using biological time series data. The second part of Chapter 2 is focused on the synthesis of a class of cancer treatment schedules, which leverages the worst-case perspective for managing dynamical systems under uncertainty.

Motivated by the conservativeness of worst-case methods for dynamical systems with large uncertainties, such as cancer treatment systems, Chapter 3 is focused on the mathematical development of our new risk-sensitive safety analysis method. Prior to presenting the mathematical details, we further motivate risk-sensitive safety analysis via a numerical example of a stormwater catchment system. We also present background on the Conditional Value-at-Risk (CVAR) measure, the Exponential Utility risk measure, and well-established

¹ $\alpha \in [0, 1]$ is the *risk-sensitivity level*.

safety analysis methods to explain where our CVAR-based methodology fits within the literature. In the rest of Chapter 3, we define the notion of a *risk-sensitive safe set*, provide a provably correct value iteration algorithm to estimate risk-sensitive safe sets and control policies for a class of linear systems, and develop an illustrative numerical example. Most of the content in Chapter 3 is from the previously published conference papers [21] [18] and the manuscript under review [22]. The final chapter of this thesis, Chapter 4, presents exciting future research directions, which we look forward to pursuing.

1.2 Notation

Mathematical notation for this thesis is provided below. Recommended references on real analysis and probability theory include [5] and [34].

- $\forall$ means for all. $\exists$ means there exists. $\iff$ means if and only if or is equivalent to. $\in$ means is an element of. $\subseteq$ means is a subset of.
- $\mathbb{R}^n$ is the set of real-valued n -dimensional vectors. I.e., $x \in \mathbb{R}^n$ means that $x = [x_1, x_2, \dots, x_n]^T$, where $x_i \in \mathbb{R}$ for each i .
- $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ is the L^p norm of a vector $x \in \mathbb{R}^n$, where $|x_i|$ is the absolute value of $x_i \in \mathbb{R}$. Chapter 2 uses $\|\cdot\|_1$ and $\|\cdot\|_2$ in various settings.
- $\mathbb{R}_+^{p \times q}$ is the set of $p \times q$ matrices with real nonnegative entries. $\mathbb{R}_+^p$ is the set of p -dimensional vectors with real nonnegative entries.
- $\mathbb{N} := \{1, 2, \dots\}$ is the set of natural numbers.
- $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ is the set of natural numbers including zero.
- $\Pi_{i=1}^p A(i) := A(p)A(p-1) \cdots A(2)A(1)$ is a product of p matrices. Note that the order of matrix multiplication is important.
- $\lceil y \rceil$ is the ceil function, and $\lfloor y \rfloor$ is the floor function. E.g., $\lceil 3.6 \rceil = 4$, and $\lfloor 3.4 \rfloor = 3$.
- $(k_j)_{j \in \{1, 2, \dots, n\}} := (k_1, k_2, \dots, k_n)$ is a finite sequence of elements. Note that elements in a sequence are ordered. Individual elements need not be unique; i.e., we may have $k_i = k_j$ for $i \neq j$.
- $\{k_j\}_{j \in \mathcal{A}}$ is a set of elements. Note that elements in a set are *not* ordered. E.g., $\{\pi, 1, 4, 1, 3\} = \{\pi, 1, 4, 3\} = \{3, 1, \pi, 4\} \neq \{3, \pi, 4\}$.
- If X is any set and $f : X \rightarrow \mathbb{R}^n$ is bounded, then the uniform norm of f is given by $\|f\|_u := \sup_{x \in X} \|f(x)\|_2$, where $\|\cdot\|_2$ is the Euclidean distance on $\mathbb{R}^n$. E.g., if $y \in \mathbb{R}^n$, then $\|y\|_2 := (\sum_{i=1}^n |y_i|^2)^{1/2}$, where $|\cdot|$ is the absolute value.

- If X is a metric space, then $\mathcal{B}(X)$ is the Borel σ -algebra on X . If $A \in \mathcal{B}(X)$, then A is called a Borel set.
- If $1 \leq p \leq \infty$, then $L^p := L^p(\Omega, \mathcal{F}, \mathbb{P})$ is the collection of functions $f : \Omega \rightarrow \mathbb{R}$, measurable relative to $\mathcal{F}$ and $\mathcal{B}(\mathbb{R})$, such that $\|f\|_p := (\int_{\Omega} |f(\omega)|^p \mathbb{P}(d\omega))^{1/p} < \infty$, where $|f(\omega)|$ is the absolute value of $f(\omega) \in \mathbb{R}$. In other words, $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is the collection of random variables with finite p^{th} moment and set of events $\mathcal{F}$. The integral $\int_{\Omega} |f| d\mathbb{P}$ is shorthand for $\int_{\Omega} |f(\omega)| \mathbb{P}(d\omega)$.
- $f \in L^{\infty} := L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ indicates that f is a bounded random variable.
- If f is a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then $\sigma(f)$ is the σ -algebra generated by f , and $\mathbb{E}(f) := \int_{\Omega} f d\mathbb{P} := \int_{\Omega} f(\omega) \mathbb{P}(d\omega)$ is the expected value of f .
- “Measurable” is short for Borel measurable, and “a.e.” stands for almost everywhere or almost every. “USC” denotes upper semi-continuous, and “LSC” denotes lower semi-continuous.
- Bold text is used to signify a deterministic quantity in Chapter 3; e.g., $x_t = \mathbf{x}$ indicates that the random variable x_t takes on the value $\mathbf{x}$.

Chapter 2

On Modeling and Managing a Cancer Subtype

Triple-negative breast cancer (TNBC) is an especially aggressive and deadly form of breast cancer that disproportionately affects younger women or women of African descent. This cancer is difficult to treat for two key reasons in particular. First, TNBC lacks the three most common proteins in breast cancer (which is why it is called *triple-negative*), so it is hard to find treatments that specifically target proteins on the tumor surface. Second, TNBC is composed of many different types of cells, where the cells are different in their phenotypic traits.

This heterogeneity in TNBC and in other basal-like breast cancers has been characterized by Risom et al. through the use of *differentiation-state* biological markers [84]. The experimental data have shown that after treatment with certain targeted therapies, some TNBC cells are able to survive and their phenotypic traits change [84, Figure 3f]. These phenotypic traits, which are associated with specific cellular behaviors, have been measured by imaging the relative fluorescence produced by the differentiation-state biological markers. Thus, the term *differentiation state* refers to a specific collection of phenotypic traits that is related to specific cellular behaviors and can be observed via differentiation-state marker expression.¹

We have developed data-driven mathematical models of TNBC differentiation-state dynamics by applying statistical learning techniques to cancer *cell line* data [19].² These models explain how TNBC cells may escape from particular therapies through differentiation-state transitions [19], which is relevant for designing improved cancer treatment strategies. Build-

¹The terms *cell state* and *phenotypic state* are also used in the literature. A cell state is the most general term and refers to a cell type with particular genetic or phenotypic traits. A phenotypic state is a cell state with particular phenotypic traits. A differentiation state is a phenotypic state that is measured through the use of differentiation-state markers. I.e., differentiation states are phenotypic states, and phenotypic states are cell states. The term *differentiation-state marker expression* refers to the relative fluorescence produced by differentiation-state biological markers.

²A *cell line* is a collection of cells from a single origin that is cultured in a laboratory setting. The cancer cell line for our data set, which is called HCC1143, originated from a (human) patient many years ago.

ing from our modeling contributions, we have proposed a class of therapeutic treatment schedules with tolerability specifications and have derived sufficient mathematical conditions for the decay of cancer cell populations under this class by invoking the worst-case perspective [20]. As well as providing theoretical insights about switched dynamical systems, this work has motivated the need for mathematical methods that facilitate less conservative but still protective control of dynamical systems.

This chapter has two main sections. First, we present the data-driven mathematical models of TNBC dynamics, most of which has been published in [19]. Additional details about the mathematical and numerical methods used for the modeling are provided in [19, S1 Appendix]. In the second section of this chapter, we present the synthesis of a class of cancer treatment schedules, most of which has been published in [20].

While our primary goal is to improve how triple-negative breast cancer (TNBC) is managed, there are no standard models to predict how TNBC evolves over time. Thus, we first devoted significant research effort to the development and analysis of TNBC dynamical models, which is presented next.

2.1 Data-Driven Modeling of TNBC Dynamics

Computational models have been built to examine the phenotypic state dynamics of cancer cell populations and the potential role of these dynamics in the development of therapeutic resistance [36] [43] [84] [17] [113] [115] [114] [76] [3]. A Markov chain model has predicted that cancer stem-like cells can arise from non-stem-like cells using probabilities identified from observations at two time points [43]. Although parameter estimation error was not examined, the prediction has been validated in an experiment [43]. Another pivotal study has used ordinary differential equation (ODE) modeling to predict that cells expressing a transient drug-tolerant phenotype arise from non-stem-like cells [36]. While the model itself was not tested on independent data, the prediction has been validated empirically [36]. Further, an ODE model has been developed using the principles of biochemical reactions to represent cell-state birth, death, and transition [115] [114]. A dynamical model that generalized prior cell-state transition models [43] [115] [114] has been constructed using a Markov process with a finite number of cell divisions [76], and phenotypic state equilibria and stability properties have been studied [76]. In the related field of clonal tumor evolution, a stochastic genotypic state birth-death process model with mutations and a corresponding deterministic ODE model have been developed [113]. The models along with Monte Carlo sampling and observations at two time points have informed parameter sensitivity analysis, a treatment window approximation, and investigations of therapeutic scheduling [113]. Although our first modeling effort in the HCC1143 cell line of basal, mesenchymal, and non-basal/non-mesenchymal states included the estimation of parameter variability, the training data set was small for the number of parameters that required identification, and no statistically significant therapy-induced effects on differentiation-state transitions were detected [17]. Studies with cell-state dynamical models rarely include statistical analysis

of model parameters (refs. [17] and [113] are exceptions) because the available data often lacks sufficient quality and quantity at multiple time points. However, in this chapter, we leverage novel time series data sets from Risom et al. [84] to estimate parameter variability, infer statistically significant therapy-induced effects on differentiation-state transitions, and examine how well our models generalize to unseen data.

Next, we summarize the content of the data sets from Risom et al. [84]. Cell populations from a TNBC cell line called HCC1143 were cultured and treated with one of four different agents: $1\mu\text{M}$ Trametinib, $1\mu\text{M}$ BEZ235, $1\mu\text{M}$ Trametinib+ $1\mu\text{M}$ BEZ235 (equal-ratio combination), and DMSO. Trametinib and BEZ235 are therapies that inhibit the MEK and PI3K/mTOR cell signaling pathways, respectively.³ DMSO, which is short for Dimethylsulfoxide, is a baseline condition that represents the absence of therapy. The numbers of cells in four different differentiation states and the numbers of cells, where the dying cells were also specified, were measured every 12 hours over 6-7 time points in 4-15 replicate populations following initial treatment [84]. The four differentiation states are: 1) basal, 2) mesenchymal, 3) luminal, and 4) non-basal/non-mesenchymal/non-luminal. The first three states predominate “basal-like” triple-negative tumors, “claudin-low” triple-negative tumors, and “luminal” ER+ tumors respectively [60] [82] [39], and many triple-negative tumors harbor a heterogeneous mixture of cells occupying all four states [84] [38]. The four differentiation states are defined according to high or low expression of three different biological markers, Cytokeratin 14 (K14), Vimentin (VIM), and Cytokeratin 19 (K19) as follows: 1) basal (K14^{hi}), 2) mesenchymal ($\text{VIM}^{\text{hi}}\text{K14}^{\text{low}}$), 3) luminal ($\text{K19}^{\text{hi}}\text{VIM}^{\text{low}}\text{K14}^{\text{low}}$), and 4) non-basal/non-mesenchymal/non-luminal ($\text{K19}^{\text{low}}\text{VIM}^{\text{low}}\text{K14}^{\text{low}}$).

Using these data, we have developed mathematical models to examine the feasibility of transitions between any two of the four key differentiation states in triple-negative breast cancer cell populations under different treatment conditions. We have used these models to infer new biological insights: 1) how often HCC1143-derived cells may transition between any two of the four differentiation states following treatment with therapy or DMSO, 2) the statistical significance or insignificance of therapy-induced differences in the transition rates, and 3) how changes in transition rates may underlie certain differentiation-state aggregations of drug-tolerant cells reported by [84]. Taken together, these insights demonstrate the feasibility of transitions in the context of the four key differentiation states in triple-negative breast cancer and how different treatments can distinctly affect the behaviors of these transitions. The following sections of this thesis present the dynamical models, our data-driven model identification procedure, uncertainty analysis of the model parameters, model validation, and an important biological hypothesis that is suggested by these computational results.

³A cell signaling pathway is a series of chemical reactions involving a collection of molecules within a cell that work together to control a cell function [73].

Therapy-Specific Differentiation-State Dynamical Models

We have identified a dynamical model to predict the evolution of the four differentiation-state subpopulations in response to a particular treatment condition (Trametinib, BEZ235, Trametinib+BEZ235, or DMSO). Each therapy-specific model quantifies how the number of live cells in each differentiation state and how the number of dead or dying cells in total change over time following initial treatment. The key feature of each therapy-specific model is the dynamics matrix, which contains the average rates of cell division, cell death, and transition between the four differentiation states under a particular treatment condition. These *dynamics parameters* are defined as follows: ρ_i is the *division gain* of differentiation state i ; ρ_{iD} is the *death gain* of differentiation state i ; ρ_{ij} is the *transition gain* from differentiation state i to differentiation state j . For example, ρ_{12} is the transition gain from the basal state ($K14^{hi}$) to the mesenchymal state ($VIM^{hi}K14^{low}$), and ρ_3 is the division gain of the luminal state ($K19^{hi}VIM^{low}K14^{low}$). A *gain* is a proportional value that quantifies the relationship between the magnitude of an input and the magnitude of an output and is a discrete-time analog of a rate.

Specifically, we have used a *linear time-invariant* model to represent the evolution of a cancer cell population in response to initial treatment with therapy j :

$$x(t+1) = A_j \cdot x(t), \quad t = 0, 1, \dots, T-1, \quad (2.1)$$

where $A_j \in \mathbb{R}^{5 \times 5}$ is the dynamics matrix for therapy j and $x = [x_1, x_2, x_3, x_4, x_5]^T \in \mathbb{R}^5$ is the vector containing the numbers of cancer cells in different categories. Specifically, $x_1 \in \mathbb{R}$ is the number of live basal (cancer) cells; $x_2 \in \mathbb{R}$ is the number of live mesenchymal (cancer) cells; $x_3 \in \mathbb{R}$ is the number of live luminal (cancer) cells; $x_4 \in \mathbb{R}$ is the number of live non-basal/non-mesenchymal/non-luminal (cancer) cells; and $x_5 \in \mathbb{R}$ is the number of dead or dying (cancer) cells. We have utilized a fluid-like representation for cell populations, where x is not necessarily a vector of integers, to accommodate the limitations of the data which do not distinguish between the live cells and the dying cells occupying a particular differentiation state. The discrete-time interval $[t, t+1)$ is the duration between two consecutive observations, or 12 hours, and T represents a finite time horizon of either 60 or 72 hours. The dynamics matrix A_j contains the transition gains, division gains, and death gains in the following form:

$$A_j := \begin{bmatrix} \alpha_1 & \rho_{21} & \rho_{31} & \rho_{41} & 0 \\ \rho_{12} & \alpha_2 & \rho_{32} & \rho_{42} & 0 \\ \rho_{13} & \rho_{23} & \alpha_3 & \rho_{43} & 0 \\ \rho_{14} & \rho_{24} & \rho_{34} & \alpha_4 & 0 \\ \rho_{1D} & \rho_{2D} & \rho_{3D} & \rho_{4D} & 1 \end{bmatrix} \quad (2.2)$$

$$\alpha_i := \rho_i - \rho_{iD} - \sum_{s=1, s \neq i}^4 \rho_{is} \quad i = 1, 2, 3, 4.$$

For example, $\alpha_2 = \rho_2 - \rho_{2D} - \rho_{21} - \rho_{23} - \rho_{24}$. The form of A_j is derived by using principles of mass balance. For example, since we have assumed that live cells are able to transition, divide, or die, we have the following equation for the dynamics of the number of live cells in

differentiation state 1:

$$x_1(t+1) = \alpha_1 \cdot x_1(t) + \rho_{21} \cdot x_2(t) + \rho_{31} \cdot x_3(t) + \rho_{41} \cdot x_4(t) + 0 \cdot x_5(t),$$

where $\alpha_1 = \rho_1 - \rho_{1D} - \rho_{12} - \rho_{13} - \rho_{14}$. The term $\alpha_1 \cdot x_1(t)$ is the net contribution to the differentiation state 1 live subpopulation from itself through division, where the cells that die or transition to other differentiation states are subtracted out. The term $\rho_{j1} \cdot x_j(t)$ is the number of live cells that transition from differentiation state j to differentiation state 1 for $j = 2, 3, 4$. The reasoning to derive the other dynamics equations is similar. Fig. 2.1 provides a visual depiction of a therapy-specific dynamical model that we have identified by using time series data from a triple-negative breast cancer (TNBC) cell line.

We have made idealistic assumptions that the evolution of a TNBC cell population is linear, time-invariant, Markovian, and fully observable. These assumptions have been motivated by the limited number of samples (90-99 per treatment condition) that are available to identify a dynamical model of differentiation-state behaviors. The differentiation-state measurements have not been automated, which has restricted the quantity and quality of time series data that is available. These challenges are likely to be mitigated in the future through the advancement of measurement technologies for time series data collection. In the meantime, however, we have been able to estimate parameter variability, perform model validation, and propose interesting biological insights by using our dynamical models, which will be presented throughout the chapter. The next section details how we have identified the therapy-specific dynamical models using the available data.

Data-Driven Identification of Dynamical Models

The core numerical problem is to estimate a dynamics matrix A_j for each treatment condition $j \in \{\text{Trametinib}, \text{BEZ235}, \text{Trametinib+BEZ235}, \text{DMSO}\}$ that fits the available data sufficiently well under the form specified by the linear time-invariant model (2.1). This problem cannot be solved exactly because of two key limitations of the data: 1) the data do not distinguish between the live cells and the dying cells in a given differentiation state; and 2) not all measurements are available due to instrument errors.

To address the first challenge, we have combined the observed numbers of cells in each differentiation state and the observed death fractions into the form of the cell type vector x , where death has been distributed evenly across the differentiation states in view of our preliminary work [19, S1 Appendix].

To address the second challenge, we have inserted these data samples into an alternating minimization (AM) algorithm to obtain an estimate of A_j , which we refer to as the *AM-optimized dynamics matrix* ($\hat{A}_j$). Alternating minimization [26] is a local optimization method that reduces the value of a given cost function by alternating the role of the optimization variable between two variables.⁴ In our setting, these two variables are a data

⁴Expectation maximization is a special case of alternating minimization [30] [59] [15] [16] [106].

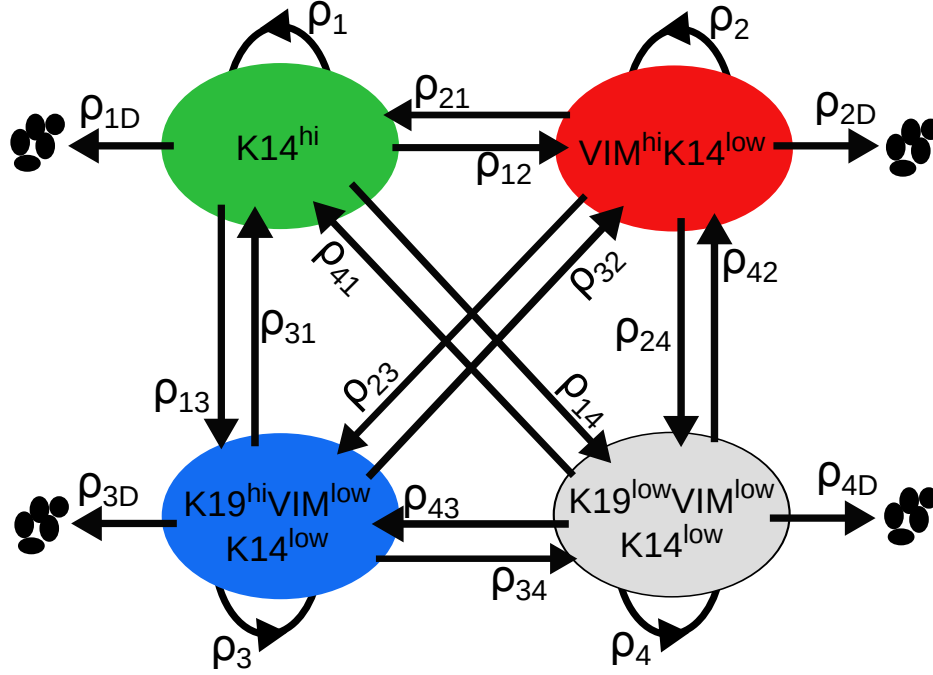


Figure 2.1: **Therapy-specific dynamical model** [19]. We have used time series data to identify a linear time-invariant model of a triple-negative breast cancer (TNBC) cell population in response to initial treatment. In the model, the live cells occupy four differentiation states and can transition, divide, or die. The dynamics parameters (ρ_{12} , ρ_{21} , ρ_3 , ρ_{4D} , etc.) are the average rates of these actions taken by the live cells in each differentiation state following initial treatment.

variable X and a dynamics matrix variable A . We have used *initialization for local optimization* [13] to help mitigate the possibility of converging to a local minimum that poorly represents the cancer dynamics. Specifically, we have initialized our alternating minimization algorithm with a dynamics matrix that solves a particular convex problem (within numerical accuracy), where the convex problem approximates our original non-convex problem. This convex problem is the minimization of our cost function (to be described in the next paragraph) in which the data variable has been set to an appropriate estimate $\hat{X}$ of its true value. Each column of $\hat{X}$ is a training data sample for a particular (time point, population) pair, or the sample mean of the available training data for the time point when training data for the (time point, population) pair was not available.⁵ We have shown that the values of the dynamics parameters converge within numerical accuracy during the iterative process of the alternating minimization algorithm [19, S3 Appendix]. In addition, we have assessed the sensitivity of the dynamics matrix returned by the algorithm with respect to the initialization

⁵Time series measurements from 15 replicate populations were taken, but some measurements could not be used since the associated images were out of focus.

of the data variable [19, S4 Appendix].

We have designed the cost function for the alternating minimization algorithm to penalize measurement error, process error, and estimation error measured in the L^2 norm [19, S1 Appendix]. This norm has been chosen because, as a general measure of length, it is well-suited to identify networks without known structural characteristics, such as sparsity. We have penalized measurement error and process error with equal weight in view of our preliminary analysis [19, S1 Appendix]. The cost function utilizes L^2 -regularization to induce element-wise shrinkage of the dynamics matrix to zero in order to reduce estimation error of the dynamics parameters [45] [50].

In the identification procedure, we have imposed several constraints on the dynamics matrix A_j (2.2), including nonnegativity, equal division gains, and equal death gains. The constraints on A_j are provided and justified in Table 2.1. For each treatment condition, we have constrained the division gains to be equal since the HCC1143 cell line has shown similar percentages of EdU-positive cells (cells that are synthesizing DNA) for the different differentiation-state marker expression levels over time [84] [19, S1 Figure]. For each treatment condition, we have constrained the death gains to be equal in view of our data processing choice to allocate observed death evenly across the differentiation-state subpopulations.

The values of the transition gains that we have identified using the above procedure are shown in Fig. 2.2. The values of the division and death gains are provided in Table 2.2. The next section presents how we have estimated the variability of the dynamics parameters using a well-established resampling algorithm.

Table 2.1: Constraints on A_j (2.2)

Constraint	Rationale
Each entry of A_j is nonnegative.	$x(t)$ is nonnegative for all time t since $x(t)$ contains numbers of cells.
$\rho_i \geq 1$ for $i = 1, 2, 3, 4$	$\rho_i \cdot x_i(t) - x_i(t)$ is the increase in the number of live cells in differentiation state i due to cell division during $[t, t + 1)$.
$\rho_{ij} \leq 1$ for $(i, j) \in \{1, 2, 3, 4\}^2, i \neq j$	Only a portion of live cells in differentiation state i at time t can transition to differentiation state j by time $t + 1$.
$\rho_{iD} \leq 1$ for $i = 1, 2, 3, 4$	Only a portion of live cells in differentiation state i at time t can die, or begin to die, by time $t + 1$.
Last column of A_j is $[0, 0, 0, 0, 1]^T$	Dead or dying cells accumulate over time and cannot come back to life.
$\rho_i = \rho_1$ for $i = 2, 3, 4$	EdU-positivity experiments suggest similar rates of cell division across the differentiation states [84].
$\rho_{iD} = \rho_{1D}$ for $i = 2, 3, 4$	Motivated by preliminary analysis [19, S1 Appendix, Section 5: System Identification].

Table 2.2: Therapy-specific division gains and death gains [19].

Treatment Condition	Division gain ρ_i	Death gain ρ_{iD}
DMSO	1.34	0.0057
Trametinib	1.07	0.019
BEZ235	1.09	0.0083
Trametinib+BEZ235	1.00	0.068

The value of the division gains ρ_i and the value of the death gains ρ_{iD} from each therapy-specific AM-optimized dynamics matrix are provided in units of $\frac{\# \text{ cells at } (k+1) \text{ multiples of 12 hours}}{\# \text{ cells at } k \text{ multiples of 12 hours}}$, where $i = 1, 2, 3, 4$ is a differentiation-state index. $i = 1$ is basal ($K14^{\text{hi}}$); $i = 2$ is mesenchymal ($VIM^{\text{hi}}K14^{\text{low}}$); $i = 3$ is luminal ($K19^{\text{hi}}VIM^{\text{low}}K14^{\text{low}}$); and $i = 4$ is non-basal/non-mesenchymal/non-luminal ($K19^{\text{low}}VIM^{\text{low}}K14^{\text{low}}$). We have imposed the following constraints: $\rho_1 = \rho_2 = \rho_3 = \rho_4$ and $\rho_{1D} = \rho_{2D} = \rho_{3D} = \rho_{4D}$ (see text for justification). Higher values indicate more frequent division or death on average over time compared to lower values.

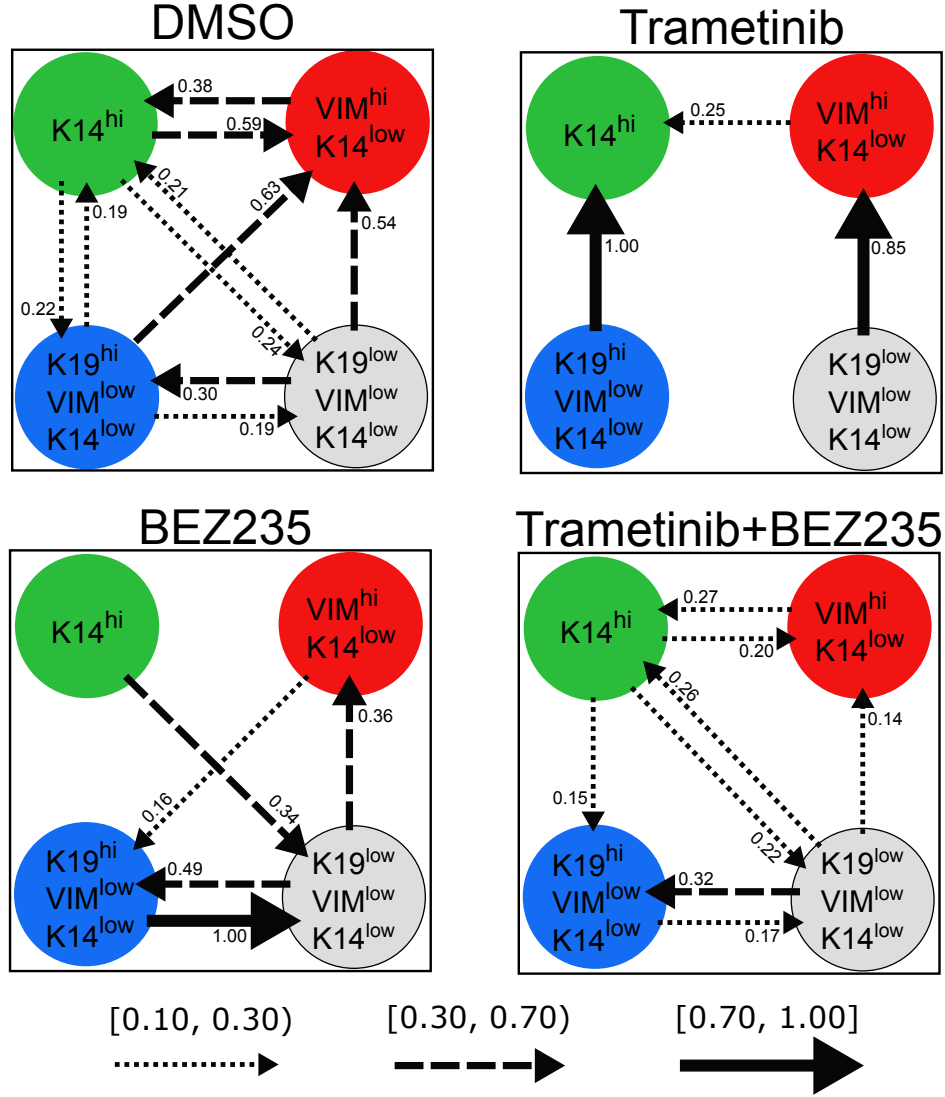


Figure 2.2: **Therapy-specific transition gains** [19]. For each treatment condition, values of the transition gains from the AM-optimized dynamics matrix are shown in units of $\frac{\# \text{ cells at } (k+1) \text{ multiples of 12 hours}}{\# \text{ cells at } k \text{ multiples of 12 hours}}$. Each transition gain from differentiation state i to differentiation state j of sufficient magnitude ($\rho_{ij} \geq 0.10$) is depicted as an arrow directed from i to j . Arrow style specifies gain magnitude. A dotted arrow means $\rho_{ij} \in [0.10, 0.30)$, a dashed arrow means $\rho_{ij} \in [0.30, 0.70)$, and a solid arrow means $\rho_{ij} \in [0.70, 1.00]$.

Uncertainty Analysis of the Dynamics Parameters

We have estimated the variability of the dynamics parameters with respect to measurement error using the resampling residuals bootstrap algorithm proposed by Wu [108].⁶ We have used the resampling method proposed by Davidson and Flachaire [28]. Our implementation assumes that measurement errors are homoskedastic (have constant variance) and independent across cell types conditioned on the (time point, population) pair in the data generating process. For each treatment condition, we have generated 120 bootstrapped dynamics matrices using the resampled data and the model identification procedure described in the previous section. From these 120 bootstrapped matrices, 120 samples of each dynamics parameter have been obtained, and a 95% confidence interval of each parameter has been computed by discarding the 3 largest samples and the 3 smallest samples (Fig. 2.3). For each treatment, we have also conducted a two-sided one-sample sign test for each dynamics parameter using the corresponding 120 bootstrapped samples [19, S2 Appendix].

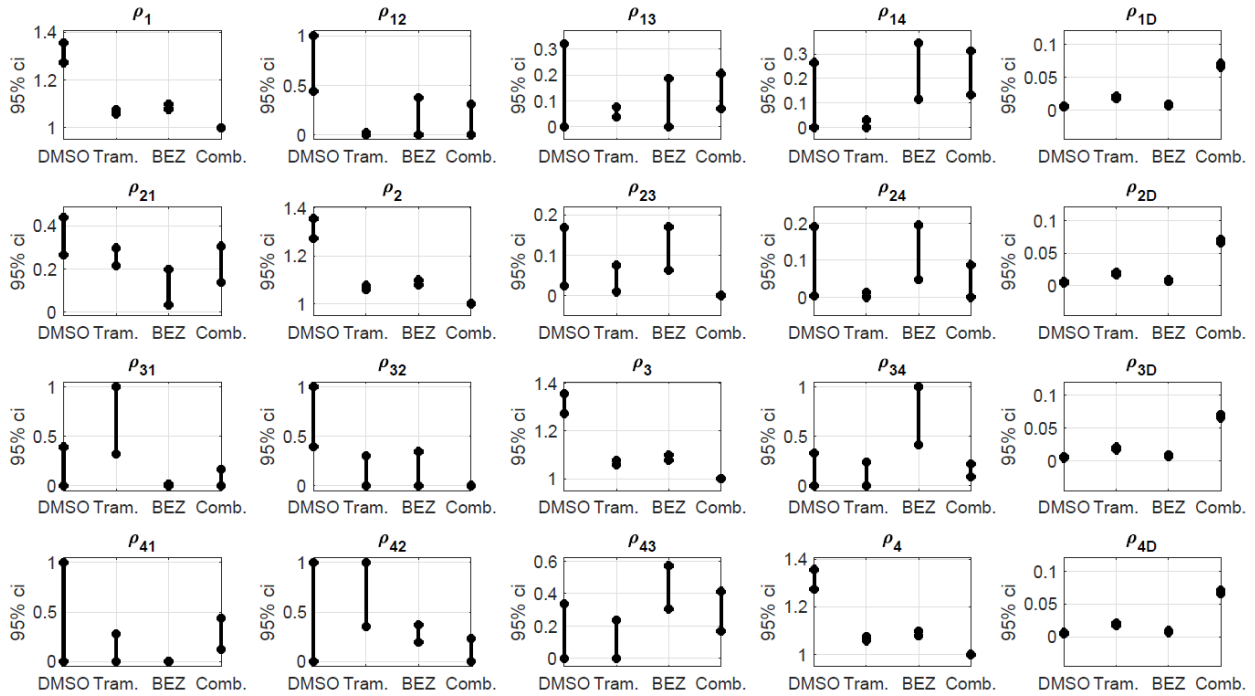


Figure 2.3: **95% confidence intervals for the dynamics parameters** [19]. These intervals indicate the variability of the dynamics parameters with respect to measurement error. Non-overlapping intervals of a given parameter specify a statistically significant difference. For example, a statistically significant reduction in $K14^{\text{hi}}$ -to- $VIM^{\text{hi}}K14^{\text{low}}$ transition has been detected under Trametinib versus DMSO because the ρ_{12} -interval for Trametinib is strictly below the ρ_{12} -interval for DMSO. Recall that ρ_{12} is the $K14^{\text{hi}}$ -to- $VIM^{\text{hi}}K14^{\text{low}}$ transition gain.

⁶This algorithm is also called *wild bootstrap* in the literature.

Model Validation

We have compared our dynamical models to a separate data set that was collected about a year before the training data set. The separate data set (which we call the *test data set*) includes 6 time points of observations from 4 replicate populations that were imaged every 12 hours following initial treatment. This model validation approach is particularly stringent; it is common to test models using data that was collected around the same time to avoid inter-experimental variability.

Predictions using a family of bootstrapped models (which we call a *model ensemble*) in comparison to the test data are shown in Fig. 2.4 for each treatment condition. The ensemble model predictions and the test data demonstrate consistency in the number of $K14^{hi}$ live cells under DMSO, the number of $K14^{hi}$ live cells under Trametinib, and the number of $VIM^{hi}K14^{low}$ live cells under Trametinib+BEZ235, evident by comparable trends and lack of significant differences (Fig. 2.4). There is also qualitative agreement between the predictions and the test data in the number of dead/dying cells for each treatment condition (Fig. 2.4). In certain cases, the predictions and the test data both increase overall, although their respective rates of change differ; e.g., see $VIM^{hi}K14^{low}$ and $K19^{low}K14^{low}VIM^{low}$ for DMSO (Fig. 2.4). The most severe discrepancies involve the differentiation states defined by VIM or K19 (Fig. 2.4), which can be explained partly by biological knowledge.

Experiments have shown that Vimentin (VIM) and Cytokeratin 19 (K19) display a continuum of low expression to high expression in HCC1143 cells, which makes the low and high cutoffs more variable across replicate experiments and introduces noise into the subpopulation fractions [19, S2 Figure]. Cytokeratin 14, however, is strongly expressed by a subset of cells and is weakly expressed, or lacks expression, in the other subset of cells [19, S2 Figure]. This biphasic expression pattern forms distinct high and low subpopulations, so the fraction of cells in each subpopulation is more similar across replicate experiments.⁷

Driven by these findings, for each treatment condition, we have identified a lower-dimensional dynamics matrix on the training data using $K14^{hi}$ and $K14^{low}$ as the differentiation-state definitions, and we have evaluated how well this matrix predicts the test data. As shown in Fig. 2.5, the predictions and the test data in this setting demonstrate qualitative consistency (comparable trends) and quantitative consistency (sufficiently large p-values, $p > 0.05$) for most cell types ($K14^{hi}$ live, $K14^{low}$ live, dead/dying) and treatment conditions.

The next section describes an important biological hypothesis that is suggested by the dynamical models.

⁷These experimental results suggest that the differentiation states in the HCC1143 cell line defined by the expression of Vimentin and Cytokeratin 19 may be modeled as continuous rather than discrete entities, which indicates the utilization of a hybrid dynamical system model in future work. The framework of hybrid dynamical systems is utilized for systems that have both discrete and continuous dynamical components. A fundamental reference for this framework is Dr. Claire Tomlin's dissertation [101].

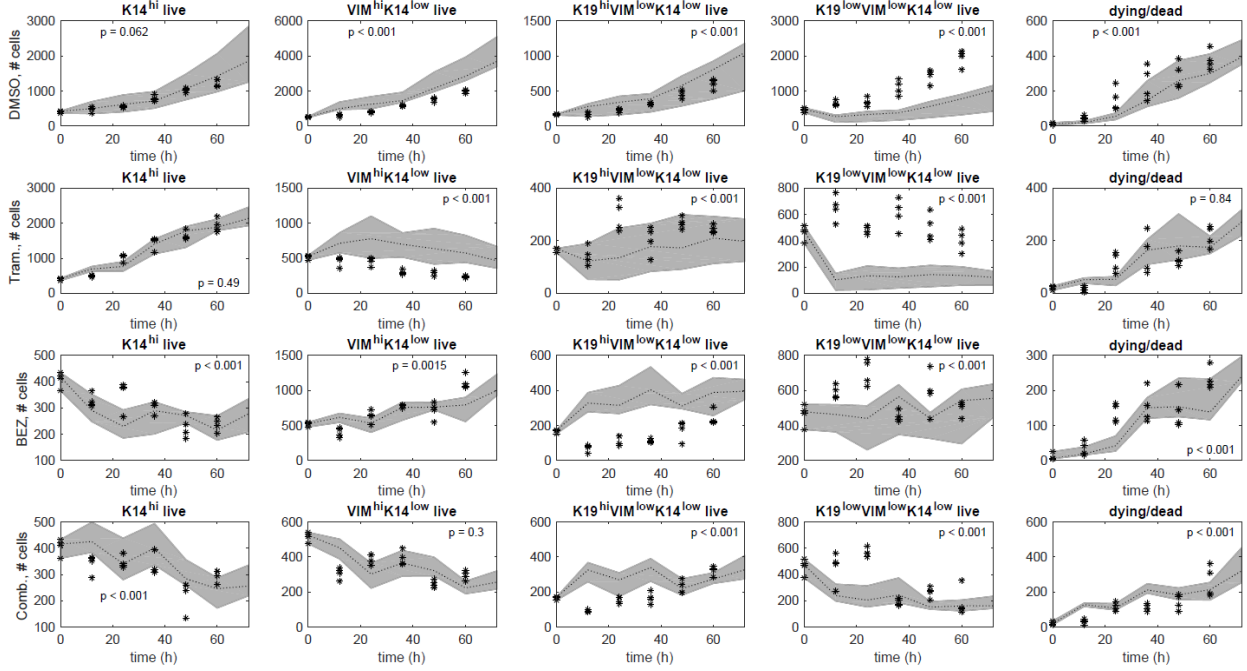


Figure 2.4: **Ensemble model predictions in comparison to test data** [19]. The test samples (black stars) and ensemble model predictions (gray bands) are shown for each treatment condition: DMSO (row 1), Trametinib (row 2), BEZ235 (row 3), and Trametinib+BEZ235 (row 4). The model ensemble is a collection of models that have been identified from the training data via the resampling residuals bootstrap algorithm [108] for each treatment condition. In each plot, we show a 95% confidence interval (gray band) around the median (black dotted line) of the ensemble model predictions. *Higher* p-values indicate *better* consistency between predictions and test data over the time horizon (12h, 24h, ..., 60h).

Biological Hypothesis

An important biological and clinical question is to understand why the $K14^{hi}$ cells are able to survive following treatment with Trametinib therapy, where this survival is evident through the experimental data provided by Risom et al. [84, Figure 3f]. Additional experiments and mathematical modeling have indicated that the survival of $K14^{hi}$ cells is not a consequence of increased death of $K14^{low}$ cells [84, Figures 3ef and Figure 3h-left] but may instead be a consequence of changes in differentiation-state transitions [84, Figure 3h-right and Figures 5ghi].

Specifically, our modeling results suggest that less frequent transitions from $K14^{hi}$ to $VIM^{hi}K14^{low}$ may be critical to the $K14^{hi}$ enrichment that has been observed following Trametinib treatment. The $K14^{hi}$ -to- $VIM^{hi}K14^{low}$ transition gain ρ_{12} is significantly reduced under Trametinib versus DMSO because the ρ_{12} -confidence interval for Trametinib is strictly below that for DMSO (Fig. 2.3). No significant difference in the reverse direction, $VIM^{hi}K14^{low}$ to

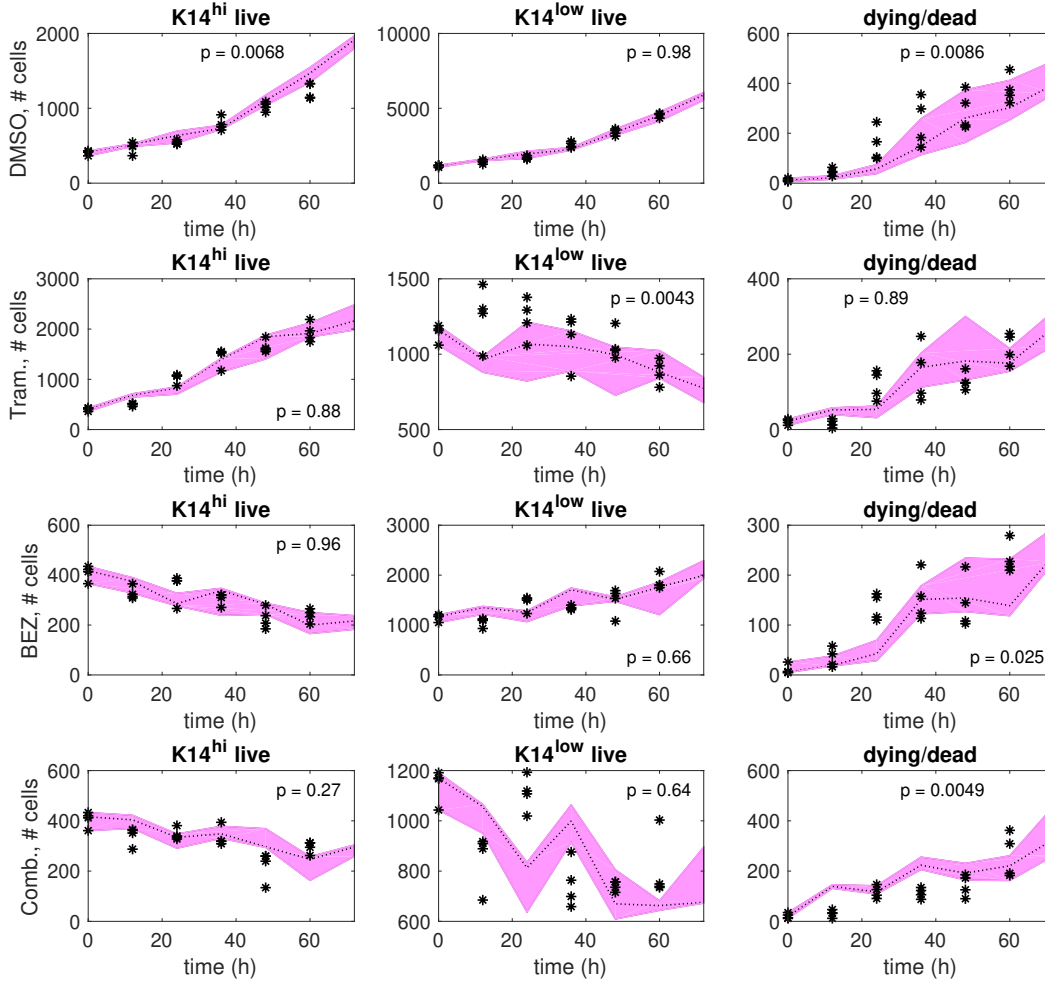


Figure 2.5: **Single model predictions in comparison to test data**, where the differentiation states are defined by K14 only [19]. The test samples (black stars) and single model predictions (pink bands) are shown for each treatment condition: DMSO (row 1), Trametinib (row 2), BEZ235 (row 3), and Trametinib+BEZ235 (row 4). The single model was identified on the training data using K14^{hi} and K14^{low} as the differentiation states for each treatment condition. The pink band extends between the maximum prediction and the minimum prediction out of four predictions in total at each time point (0h, 12h, ..., 72h). The dotted line indicates the median of the predictions. *Higher* p-values indicate *better* consistency between predictions and test data over the time horizon (12h, 24h, ..., 60h).

K14^{hi}, has been detected under Trametinib versus DMSO since the ρ_{21} -confidence intervals for Trametinib and DMSO overlap (Fig. 2.3).

More frequent transitions from K19^{hi}VIM^{low}K14^{low} to K14^{hi} may also underlie the K14^{hi} enrichment that has been observed following Trametinib treatment. In particular, the K19^{hi}VIM^{low}K14^{low}-to-K14^{hi} transition gain equals its upper bound for Trametinib, $\rho_{31} = 1$, and is approximately five times smaller for DMSO, $\rho_{31} = 0.19$ (Fig. 2.2).⁸ No significant increase has been detected because the ρ_{31} -confidence intervals for Trametinib and DMSO overlap, but the amount of overlap is small compared to the length of either interval. The ρ_{31} -confidence interval for Trametinib is $[0.32, 1]$, and the ρ_{31} -confidence interval for DMSO is $[0, 0.39]$ (Fig. 2.3).

To further examine the predictions above, we have trained another dynamics matrix for Trametinib with two additional constraints: 1) $\rho_{12} \geq 0.59$, which is the value of the DMSO K14^{hi}-to-VIM^{hi}K14^{low} transition gain, and 2) $\rho_{31} \leq 0.19$, which is the value of the DMSO K19^{hi}VIM^{low}K14^{low}-to-K14^{hi} transition gain (Fig. 2.2). Fig. 2.6 shows the K14^{hi} live cell trajectories predicted by the further constrained dynamics matrix and those predicted by the (Trametinib) AM-optimized dynamics matrix in comparison to test data. The AM-optimized dynamics matrix provides trajectories that demonstrate qualitative and quantitative consistency with the test data, whereas the further constrained dynamics matrix fails in these regards. This simulation result supports our prediction that decreased K14^{hi}-to-VIM^{hi}K14^{low} transition or increased K19^{hi}VIM^{low}K14^{low}-to-K14^{hi} transition underlie the K14^{hi} enrichment that follows Trametinib treatment in comparison to DMSO.

Discussion

The dynamical models that we have developed provide improved understanding of the nature of differentiation-state heterogeneity in triple-negative breast cancer and more specifically, a means to predict how therapy can affect differentiation-state transitions. However, it is important to note that empirical validation of our predictions poses particular challenges. Current antibody-based techniques for assessing intracellular protein expression in cells grown in two dimensions require cell fixation (i.e., preservation of cellular structures for further analysis). So, observing the numbers of cells in each differentiation state can be performed only in fixed cells, and transitions cannot be observed in real time. Nonetheless, if cell-surface markers were found to correlate well with the four differentiation states in our study, then existing methods could be used to validate our hypotheses. A given state could be isolated via Fluorescence-Activated Cell Sorting [43], and then the homogeneous cell population could be treated and observed for changes in cell-surface marker expression.

The accuracy and the predictive power of the differentiation-state dynamical models will improve as experimental methods improve. Since dying cells show false positivity for all markers, our instruments could not simultaneously detect the differentiation-state marker expression of a single cell and whether that cell was alive or dying. To manage this limitation,

⁸The transition gains represent proportions of cells and so are bounded between 0 and 1 (Table 2.1).

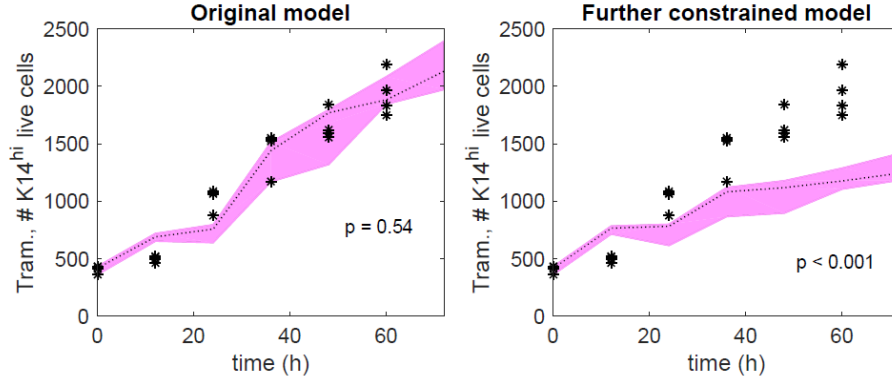


Figure 2.6: **Further investigations of Trametinib-induced K14^{hi} enrichment hypothesis** [19]. Left: Trametinib K14^{hi} live cell predictions by the AM-optimized dynamics matrix (pink band) are shown in comparison to test data (black stars). Right: Trametinib K14^{hi} live cell predictions by a dynamics matrix identified with additional constraints (pink band) in comparison to test data (black stars). The additional constraints are $\rho_{12} \geq 0.59$, the value of ρ_{12} for DMSO, and $\rho_{31} \leq 0.19$, the value of ρ_{31} for DMSO. Recall that ρ_{12} is the K14^{hi}-to-VIM^{hi}K14^{low} transition gain, and ρ_{31} is the K19^{hi}VIM^{low}K14^{low}-to-K14^{hi} transition gain. In each plot, the pink band extends between the maximum prediction and the minimum prediction out of four predictions in total at each time point (0h, 12h, ..., 72h). The dotted line indicates the median of the predictions. Higher p-values indicate better consistency between predictions and test data over the time horizon (12h, 24h, ..., 60h).

we distributed the observed death fractions evenly across the observed numbers of cells occupying each differentiation state to estimate the data samples required for modeling and subsequent analyses [19, S1 Appendix]. Moreover, our instruments can only detect cells with intact nuclei, so dying cells can fade from view. This is one reason why the number of dead or dying cells in the data may decrease. While empirical observations indicate time-varying rates of cell division and death, our models are restricted to encoding these rates on average (see [84, Figure 4c] for cell division data; see [19, Figure 6] for death data; Table 2.2 provides division and death gains). There will be potential to relax the time-invariance assumption when more time series data is available to help mitigate overfitting [52].

Our models have predicted that treating HCC1143 cells with a MEK inhibitor, a PI3K/mTOR inhibitor, or a combination of these inhibitors can alter rates of transitions between basal, mesenchymal, luminal, and non-basal/non-mesenchymal/non-luminal states relative to DMSO. These predictions provide new biological insights into how changes in transition rates may underlie certain differentiation-state aggregations of drug-tolerant persister cells that have been reported by [84]. In particular, our findings support differentiation-state transition as the major mechanism underlying resistance to MEK and PI3K/mTOR inhibitors. Our modeling work demonstrates the feasibility of this mechanism by predicting—with statistical rigor—the directionality of state transition in the absence of, and in the presence

of, therapeutic pressure. Improved understanding of the directionality of state transition may inform the design of mechanistic studies that promote the development of improved treatment strategies for heterogeneous plastic cancers.

2.2 Controller Synthesis via the Worst-Case Perspective

One important experimental finding discussed in the previous section is that the application of Trametinib therapy to the triple-negative breast cancer cell line HCC1143 facilitates aggregation of the basal ($K14^{hi}$) differentiation state [84, Figure 3f]. That is, Trametinib is able to drive a heterogeneous cancer cell population to a homogeneous state. This outcome suggests the following treatment strategy. Why not first apply Trametinib to drive the cancer cell population to the $K14^{hi}$ state, and then administer a different therapy that specifically targets the residual $K14^{hi}$ cells? This proposed strategy is illustrated in Fig. 2.7.

The challenge is that it is difficult to identify a single therapy that specifically targets one differentiation state without causing other types of interactions in the cell population as well. More generally, it is difficult to balance the requirement to kill the cancer cells in a timely manner without also causing undue harm to the non-cancer (i.e., healthy) cells. While we do not yet have the data to quantify how healthy cells respond to cancer treatments, we do know that cancer therapies can be quite harmful to healthy cells, which is evident by adverse secondary reactions [42] [104]. These critical challenges indicate the need for new computational tools to inform the development of *more tolerable treatment strategies that are designed to control the dynamics of cancer*. This section presents one approach towards this goal through the use of a switched dynamical system model.

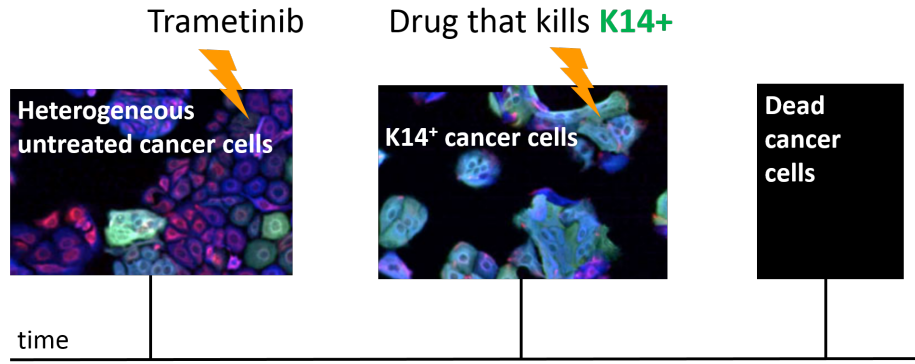


Figure 2.7: **A proposed treatment strategy.** The images of the cancer cells are from [84].

Problem Formulation

We assume that linear time-invariant (LTI) dynamical models for the responses of cancer cell populations to various therapies at a fixed dose are available (such models have been identified in Section 2.1). We suppose that some of these therapies can shrink the live cancer cell population but are very toxic to healthy cells. In addition, there are other therapies that only slow the growth of the live cancer cell population but are less toxic to healthy cells. While applying any drug in the first group repeatedly should eradicate the cancer, we assume that this approach is impractical due to toxicity concerns, negative secondary reactions, or the development of drug resistance. The less toxic drugs, on the other hand, cannot kill the cancer alone, but they do reduce the cancer growth rate and are better tolerated. We further assume that longer waiting times between treatments are preferable since longer waiting times imply a smaller total therapeutic concentration in the cells over a fixed time period.

In the setting described above, we seek to address a controller synthesis problem for cancer treatment. The dynamical system is a population of cancer cells subject to therapeutic intervention, and the control input is how the available therapies are administered (e.g., which therapy is applied at what time, the order in which the therapies are applied, and the waiting time between treatments). Our goal is to *stabilize* the live cancer cell population, which means that we aim to drive the number of live cancer cells to zero as time becomes sufficiently large. Overall, we analyze *schedules composed of therapies with varying toxicities and long waiting times between treatments* by employing a switched dynamical system model.

Stability and controller synthesis of switched dynamical systems have been well-studied in the literature (e.g., see [53], [27], [97], [63], and [65]). Further, the synthesis of therapeutic schedules for HIV treatment has been posed as the optimization of control laws for switched systems [49] [47] [48]. Inspired by these works, we consider the following problem at the intersection of switched systems and cancer treatment.

We propose the use of a cyclic schedule of $d \in \mathbb{N}$ therapies, where each therapy is applied once per cycle in any order. Some of the d therapies can shrink the live cancer cell population but are extremely toxic to healthy cells ($\mathcal{I}$ denotes the set of these therapies). Other less toxic therapies can only slow the growth of the live cancer cell population ($\mathcal{J}$ denotes the set of these therapies). We provide an upper bound on the cancer growth rate in response to a single therapy by using the matrix norm induced by the vector L^1 norm and the matrix structure from Section 2.1 (Lemma 1). We derive a set of maximal waiting times between therapies under the assumption that waiting time bounds representing a measure of toxicity to normal cells or the onset of resistance are available for each therapy (Lemma 2). This assumption is justified in part by the limited efficacy of using one therapy to treat certain cancers. In the absence of modeling error, we show that a cyclic schedule with a set of maximal waiting times stabilizes the live cancer cell population exponentially (Theorem 1). Further, we prove that if the modeling error is bounded and if the product of the errors in each cycle is sufficiently small, then a cyclic schedule with a set of maximal waiting times also exponentially stabilizes the population (Theorem 2). Using this last result, we derive a

conservative upper bound on the amount of time required for the population to settle to a small size (Corollary 1).

Switched Dynamical System Model

Recall from Section 2.1 that we have identified and validated a linear time-invariant model to represent the evolution of a cancer cell population in response to initial treatment with a single therapy. In particular, we have modeled how the numbers of live cancer cells in various differentiation states and how the number of dead or dying cancer cells evolve over time following treatment.

Now, we are interested in stabilizing the live cancer cell population. I.e., we would like to design a (tolerable) sequence of therapies that will drive the live cancer cell population to the origin, as time becomes sufficiently large. Hence, we consider the following switched dynamical system model:

$$x(t+1) = A_{\delta_t} \cdot x(t); \quad t \in \mathbb{N}_0, \delta_t \in \mathcal{D}, x(0) = x_0, \quad (2.3)$$

where $x = [x_1, x_2, \dots, x_p]^T \in \mathbb{R}^p$ is the live cancer cell population with x_i being the number of live cells in differentiation state i , and there are p differentiation states.⁹ δ_t is the therapy that is most active in the cancer cell population at time t (i.e., the therapy that has been applied most recently), and $\mathcal{D} := \{1, 2, \dots, d\}$ is the set of available therapies. A_{δ_t} takes the form of the first p rows and p columns of (2.2) as shown below:

$$A_{\delta_t} = \begin{bmatrix} \alpha_1 & \rho_{21} & \cdots & \rho_{(p-1)1} & \rho_{p1} \\ \rho_{12} & \alpha_2 & \cdots & \rho_{(p-1)2} & \rho_{p2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{1(p-1)} & \rho_{2(p-1)} & \cdots & \alpha_{p-1} & \rho_{p(p-1)} \\ \rho_{1p} & \rho_{2p} & \cdots & \rho_{(p-1)p} & \alpha_p \end{bmatrix} \quad (2.4)$$

$$\alpha_i = \rho_i - \rho_{iD} - \sum_{s=1, s \neq i}^p \rho_{is}, \quad i = 1, 2, \dots, p,$$

where the parameters $\rho_{ij} = \rho_{ij}(\delta_t)$, $\rho_i = \rho_i(\delta_t)$, and $\rho_{iD} = \rho_{iD}(\delta_t)$ depend on the therapy δ_t . We assume that the initial live cell population x_0 is non-zero. Recall from the previous section that we have imposed the following (well-justified) constraints: each entry of A_{δ_t} is nonnegative, $\rho_1(\delta_t) = \rho_2(\delta_t) = \dots = \rho_p(\delta_t)$, and $\rho_{1D}(\delta_t) = \rho_{2D}(\delta_t) = \dots = \rho_{pD}(\delta_t)$ for each therapy δ_t . We denote the difference between the division gain and death gain for therapy δ_t as follows:

$$\mu_{\delta_t} := \rho_1(\delta_t) - \rho_{1D}(\delta_t) = \rho_2(\delta_t) - \rho_{2D}(\delta_t) = \dots = \rho_p(\delta_t) - \rho_{pD}(\delta_t). \quad (2.5)$$

The switched dynamical system (2.3) assumes that the response to a therapy applied at time t does not depend on the therapies applied previously. This assumption is not true

⁹With slight abuse of notation, the vector x has p dimensions, and the dynamics matrix A_{δ_t} has p rows and p columns. In Section 2.1, $p = 4$.

generally. Our theoretical results involving modeling error (Theorem 2, Corollary 1) start to address the possibility of interactions between therapies to relax this assumption. Two additional important assumptions that the switched dynamical system (2.3) implies are the following: 1) the only external factor affecting the system is the therapeutic input; and 2) the system is fully observable. While these assumptions are not true in practice, they are motivated by the limitations of current knowledge, and they may be relaxed with improved measurement techniques in the future.

The switched dynamical system (2.3) is called a *positive system* because each element of x is nonnegative [31]. Positive systems may make computational tasks, such as estimation, more challenging due to the presence of additional constraints. On the other hand, positive switched systems may be useful for the optimization of therapeutic schedules; see [49], [47], [48], and the next section.

Preliminary Results

In the following lemma, we derive an upper bound on the cancer growth rate by using the structure of the dynamics matrix (2.4).

Lemma 1. *Let A_{δ_t} be given by (2.4). Then, the matrix norm of A_{δ_t} induced by the vector L^1 norm is μ_{δ_t} , which is defined by (2.5).*

Proof.

$$\begin{aligned} \|A_{\delta_t}\|_1 &= \max_{i=1,2,\dots,p} |\alpha_i(\delta_t)| + \sum_{s=1,s \neq i}^p |\rho_{is}(\delta_t)| \\ &= \max_{i=1,2,\dots,p} |\mu_{\delta_t} - \sum_{s=1,s \neq i}^p \rho_{is}(\delta_t)| + \sum_{s=1,s \neq i}^p |\rho_{is}(\delta_t)| \\ &= \max_{i=1,2,\dots,p} \mu_{\delta_t} - \sum_{s=1,s \neq i}^p \rho_{is}(\delta_t) + \sum_{s=1,s \neq i}^p \rho_{is}(\delta_t) \\ &= \mu_{\delta_t}, \end{aligned}$$

where the third line holds because each entry of A_{δ_t} is nonnegative. $\square$

Remark 1. If the division gains are not assumed to be equal and the death gains are not assumed to be equal, then $\|A_{\delta_t}\|_1 = \max_{i=1,2,\dots,p} \rho_i(\delta_t) - \rho_{iD}(\delta_t)$, which is the largest net growth rate of the p differentiation-state subpopulations.

Definition 1. $k_j \in \mathbb{N}$ is the waiting time between the application of therapy $j \in \mathcal{D}$ and the application of the next therapy.

The next lemma provides a set of maximal waiting times under the assumption that waiting time bounds related to toxicity (or to the onset of resistance) are available for each therapy. The waiting times together with a therapeutic sequence determine which therapy is applied and when that therapy is applied. The waiting times are designed so that

the treatment regimen shrinks the live cancer cell population over time, while limiting the toxicity to normal cells or avoiding the onset of drug resistance implicitly.¹⁰

Lemma 2. Suppose $(L_j, U_j) \in \mathbb{N}^2$ with $L_j \leq U_j$ for $j \in \mathcal{D}$. Let $\mathcal{I} := \{i \in \mathcal{D} : \mu_i \in (0, 1)\}$ and $\mathcal{J} := \{j \in \mathcal{D} : \mu_j \geq 1\}$ be non-empty. Assume $\beta := \prod_{i \in \mathcal{I}} \mu_i^{U_i} \prod_{j \in \mathcal{J}} \mu_j^{L_j} \in (0, 1)$. Choose any $\epsilon \in [\beta, 1)$. Consider the following optimization program:

$$\begin{aligned} & \underset{(k_j)_{j \in \mathcal{D}}}{\text{maximize}} && \sum_{j \in \mathcal{D}} k_j \\ & \text{subject to} && \prod_{j \in \mathcal{D}} (\mu_j)^{k_j} \leq \epsilon \\ & && k_j \in [L_j, U_j] \cap \mathbb{N} \text{ for } j \in \mathcal{D}. \end{aligned} \tag{2.6}$$

Then, $k_i^* = U_i$ for all $i \in \mathcal{I}$, and $(k_j^*)_{j \in \mathcal{J}}$ can be found via Algorithm 1. Further, if $\mathcal{I} = \{1\}$ and $\mathcal{J} = \{2\}$, then $k_2^* = \min \left(U_2, \lfloor \frac{\log \epsilon - U_1 \log \mu_1}{\log \mu_2} \rfloor \right)$.

Data: $n_l \in \mathcal{J}$ s.t. $\mu_{n_1} \leq \mu_{n_2} \leq \dots \leq \mu_{n_J}$, $J = |\mathcal{J}|$;

Result: $(k_j^*)_{j \in \mathcal{J}}$;

initialize $q = J$, $k_{n_l} = U_{n_l}$ for $l = 1, \dots, J$;

while true do

if $\prod_{i \in \mathcal{I}} (\mu_i)^{U_i} \prod_{l=1}^J (\mu_{n_l})^{k_{n_l}} \leq \epsilon$ **then**

$k_{n_l}^* = k_{n_l}$ for $l = 1, \dots, J$;

break;

else

if $k_{n_q} = L_{n_q}$ **then**

$q = q - 1$;

end

$k_{n_q} = k_{n_q} - 1$;

end

end

Algorithm 1: For the optimization program (2.6) in Lemma 2.

Proof. Choose any $(k_j)_{j \in \mathcal{D}}$ satisfying the constraints of (2.6). Because $\mu_i \in (0, 1)$ and $k_i \leq U_i$ for all $i \in \mathcal{I}$,

$$\epsilon \geq \prod_{i \in \mathcal{I}} \mu_i^{k_i} \prod_{j \in \mathcal{J}} \mu_j^{k_j} \geq \prod_{i \in \mathcal{I}} \mu_i^{U_i} \prod_{j \in \mathcal{J}} \mu_j^{k_j}. \tag{2.7}$$

¹⁰An important area for future research is to explicitly measure the death of normal cells under a given therapy and use these measurements to quantify toxicity for that therapy explicitly.

Thus, $k_i^* = U_i \forall i \in \mathcal{I}$. Algorithm 1 initializes each k_j ($j \in \mathcal{J}$) to be as large as possible. If this choice satisfies the inequality constraint of (2.6), then the algorithm terminates. If not, the algorithm decreases the waiting time associated with the largest μ_j (which is denoted by k_j) by 1. The algorithm starts with the largest μ_j to obtain the largest reduction, $\mu_j^{k_j} \geq \mu_j^{k_j-1}$, possible. If the k_j reaches its minimum, the algorithm moves onto the waiting time associated with the second largest μ_j . The algorithm is guaranteed to terminate because $\beta \leq \epsilon$.

If $\mathcal{I} = \{1\}$ and $\mathcal{J} = \{2\}$, then the inequalities $\mu_1^{U_1} \mu_2^{L_2} \leq \mu_1^{U_1} \mu_2^{k_2} \leq \epsilon$ are equivalent to

$$L_2 \leq k_2 \leq \frac{\log \epsilon - U_1 \log \mu_1}{\log \mu_2}. \quad (2.8)$$

Choose k_2^* as the largest value satisfying (2.8) and $k_2 \in [L_2, U_2] \cap \mathbb{N}$. $\square$

Remark 2. In Lemma 2, β is the fastest possible decay rate of the live cancer cell population per treatment cycle, if each of the d drugs is applied once per cycle and the waiting times are provided by (2.6).

Main Results

Now we present the mathematical analysis results that utilize the lemmas in the previous subsection. We define a *cyclic schedule* of $\mathcal{D}$ as a sequence of therapies:

$$(l_{1m}, l_{2m}, \dots, l_{dm})_{m=1}^{\infty}, \quad (2.9)$$

such that $l_{im} \in \mathcal{D}$ is the i^{th} therapy applied in cycle m , and $\cup_{i=1}^d \{l_{im}\} = \mathcal{D}$ for each m . This means that each therapy in $\mathcal{D}$ is applied once per cycle, and the order of the therapies in each cycle may vary.

Fig. 2.8 illustrates a cyclic schedule of $\mathcal{D} := \{\diamond, \square, \triangle\}$ with the waiting times $(k_{l_{im}})_{i=1}^3$ for two cycles. Note that $l_{im} \in \mathcal{D}$ and $\cup_{i=1}^3 \{l_{im}\} = \mathcal{D}$ for $m = 1, 2$.

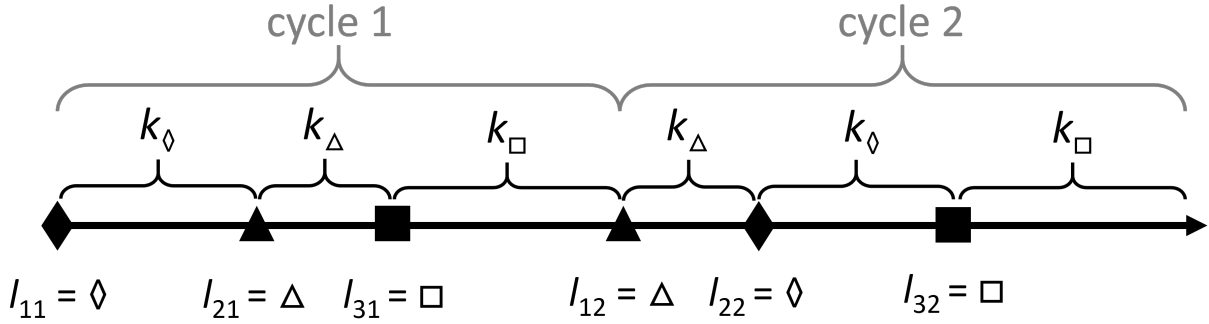
Analysis in the Absence of Modeling Error

The next theorem provides the following conclusion. In the absence of modeling error, a cyclic schedule that uses the maximal waiting times derived in Lemma 2 and an arbitrary ordering of the therapies in each cycle will exponentially stabilize the live cancer cell population modeled via the switched dynamical system (2.3).

Theorem 1. Assume the conditions of Lemma 2, choose any $\epsilon \in [\beta, 1)$, and let $\mathcal{K} := (k_j)_{j \in \mathcal{D}}$ be a solution to (2.6). Then, a cyclic schedule (2.9) with the waiting times $\mathcal{K}$ stabilizes the system (2.3) exponentially at the decay rate per cycle ϵ .

Proof. Define $T_p := p \sum_{j \in \mathcal{D}} k_j$ for $p \in \mathbb{N}_0$. Let $m \in \mathbb{N}$. By induction on (2.3) and (2.9),

$$x(T_m) = \prod_{c=1}^m \left((A_{l_{dc}})^{k_{l_{dc}}} \dots (A_{l_{2c}})^{k_{l_{2c}}} (A_{l_{1c}})^{k_{l_{1c}}} \right) x_0, \quad (2.10)$$



© 2018 IEEE

Figure 2.8: A cyclic schedule of three therapies, $\mathcal{D} := \{\diamond, \square, \Delta\}$, with the waiting times $(k_{\diamond}, k_{\square}, k_{\Delta})$ is shown for two cycles [20]; see (2.9). For example, $l_{32} = \square$ means that therapy $\square$ is the third therapy applied in cycle 2, and the next therapy will be applied $k_{\square}$ time points later. Notice that each therapy in $\mathcal{D}$ is applied once per cycle, and the order of the therapies in each cycle may vary.

such that $l_{ic} \in \mathcal{D}$ is the i^{th} drug applied in cycle c . Take the L^1 norm of (2.10) and use $\|A_{\delta_t}\|_1 = \mu_{\delta_t}$ from Lemma 1 to derive the following:

$$\begin{aligned} \|x(T_m)\|_1 &\leq \prod_{c=1}^m \left((\mu_{l_{dc}})^{k_{l_{dc}}} \cdots (\mu_{l_{2c}})^{k_{l_{2c}}} (\mu_{l_{1c}})^{k_{l_{1c}}} \right) \|x_0\|_1 \\ &= \left(\prod_{i \in \mathcal{D}} \mu_i^{k_i} \right)^m \|x_0\|_1, \end{aligned} \quad (2.11)$$

which holds because each therapy in $\mathcal{D}$ is applied once per cycle. Since $\prod_{i \in \mathcal{D}} (\mu_i)^{k_i} \leq \epsilon$ by Lemma 2, we have

$$\|x(T_m)\|_1 \leq \epsilon^m \|x_0\|_1. \quad (2.12)$$

Define $K_m \in \operatorname{argmax} \{\|x(t)\|_1 : t \in (T_{m-1}, T_m] \cap \mathbb{N}\}$. Because local maxima (or minima) only occur at times of therapy application, we have

$$x(K_m) = (A_{l_{im}})^{k_{l_{im}}} \cdots (A_{l_{2m}})^{k_{l_{2m}}} (A_{l_{1m}})^{k_{l_{1m}}} x(T_{m-1}), \quad (2.13)$$

where $l_{im} \in \mathcal{D}$ is the i^{th} drug applied in cycle m and $i \in \{1, \dots, d\}$ by (2.3) and (2.9). Define $U < \infty$ such that

$$U \geq \prod_{j \in \mathcal{D}'} \mu_j^{k_j} \quad \forall \mathcal{D}' \subseteq \mathcal{D}. \quad (2.14)$$

Take the L^1 norm of (2.13) and use $\|A_{\delta_t}\|_1 = \mu_{\delta_t}$ from Lemma 1 to obtain

$$\begin{aligned} \|x(K_m)\|_1 &\leq (\mu_{l_{im}})^{k_{l_{im}}} \cdots (\mu_{l_{2m}})^{k_{l_{2m}}} (\mu_{l_{1m}})^{k_{l_{1m}}} \|x(T_{m-1})\|_1 \\ &= \prod_{j \in \mathcal{D}'} (\mu_j)^{k_j} \|x(T_{m-1})\|_1 \\ &\leq U \|x(T_{m-1})\|_1, \end{aligned} \quad (2.15)$$

where $\mathcal{D}' = \{l_{1m}, l_{2m}, \dots, l_{im}\}$ in the second line. Finally, use (2.12) with (2.15) to see that

$$\|x(K_m)\|_1 \leq U\|x(T_{m-1})\|_1 \leq U\epsilon^{m-1}\|x_0\|_1. \quad (2.16)$$

Because $0 < \epsilon < 1$, $\|x(K_m)\|_1 \rightarrow 0$ exponentially with decay rate ϵ , as $m \rightarrow \infty$. Since $\|x(K_m)\|_1$ is a maximum on cycle m , the proof is complete. $\square$

Analysis in the Presence of Modeling Error

In this section, we provide mathematical analysis results in the presence of modeling error. An important source of modeling error arises because *how the cancer responds at the current time to the most recent therapy likely depends on the prior treatment regimen*, but the therapy-specific dynamical models are typically identified separately. A therapy j is applied to a cancer cell population that has not been treated before, and the therapy-specific dynamical model A_j is identified using the time series data that has been collected in response to therapy j . Since the experiments are laborious and expensive, one would like to conduct the least number of experiments, for example one experiment per therapy, and then use the dynamical models that have been identified from these experiments to synthesize therapeutic schedules that could be more effective than standard regimens. Using the mathematical machinery of switched systems, we sequence the dynamical models together, knowing that when the models are used in this way error arises, and we ask, what are sufficient conditions on the error under which stability is attained.

More formally, suppose that the therapy-treated live cancer cell population can be modeled using the following linear time-varying switched dynamical system:

$$x(t+1) = A_{\delta_t}(t) \cdot x(t); \quad t \in \mathbb{N}_0, \delta_t \in \mathcal{D}, x(0) = x_0, \quad (2.17a)$$

where the initial state $x_0 \in \mathbb{R}_+^p$ is non-zero. The dynamics matrix $A_{\delta_t}(t)$ quantifies the effect of applying the therapy $\delta_t \in \mathcal{D}$ at time t (or before time t), in addition to the effect of the *prior treatment regimen*: the ordering and the timing of the therapies that have been applied prior to the therapy δ_t . Suppose that the dynamics matrix in (2.17a) $A_{\delta_t}(t)$ is related to the dynamics matrix in (2.3) A_{δ_t} through the bounded multiplicative error term $\xi_{\delta_t}(t) \in (0, E]$ as follows:

$$\begin{aligned} \|A_{\delta_t}(t)\|_1 &= \|A_{\delta_t}\|_1 \cdot \xi_{\delta_t}(t) \\ &= \mu_{\delta_t} \cdot \xi_{\delta_t}(t), \end{aligned} \quad (2.17b)$$

where the second line holds by Lemma 1. Multiplicative error is mathematically convenient for quantifying the distance from the origin because the state vector $x(t+1)$ is determined by matrix multiplication; see (2.17a). While the dynamics matrix $A_{\delta_t}(t)$ and the multiplicative error $\xi_{\delta_t}(t)$ are functions of the prior treatment regimen, we do not write these dependencies explicitly to simplify notation.

Remark 3. The switched dynamical system (2.17) reduces to the error-free switched dynamical system (2.3) if $\xi_{\delta_t}(t) = 1 \forall t \in \mathbb{N}_0$.

Remark 4. If $\xi_{\delta_t}(t) > 1$, then the error has a de-stabilizing effect. If $\xi_{\delta_t}(t) < 1$, then the error has a stabilizing effect. If $\xi_{\delta_t}(t) = 1$, then the error has no effect.

We would like to derive sufficient conditions on the modeling error under which a cyclic schedule with the waiting times given by Lemma 2 stabilizes the live cancer cell population. Next, we show that exponential stability is attained, if the errors are bounded and if the error product in each cycle is sufficiently small.

Theorem 2. Assume the conditions of Lemma 2, choose any $\epsilon \in [\beta, 1)$, and let $\mathcal{K} := (k_j)_{j \in \mathcal{D}}$ be a solution to (2.6). Define $T_p := p \sum_{j \in \mathcal{D}} k_j$ for $p \in \mathbb{N}_0$. If $\xi_{\delta_t}(t) \in (0, E] \forall t \in \mathbb{N}_0$ for some $E < \infty$ and if $\exists \eta \in (\epsilon, \infty)$ such that

$$\frac{1}{\eta} \geq \prod_{t=T_{m-1}}^{T_m-1} \xi_{\delta_t}(t) \quad \forall m \in \mathbb{N}, \quad (2.18)$$

then a cyclic schedule (2.9) with the waiting times $\mathcal{K}$ stabilizes the system (2.17) exponentially at the decay rate per cycle $\frac{\epsilon}{\eta}$.

Proof. Let $m \in \mathbb{N}$. Use (2.17a) to derive the following equation:

$$x(T_m) = \prod_{t=0}^{T_m-1} A_{\delta_t}(t) \cdot x_0. \quad (2.19)$$

Take the L^1 norm of (2.19) to obtain

$$\begin{aligned} \|x(T_m)\|_1 &\leq \prod_{t=0}^{T_m-1} \|A_{\delta_t}(t)\|_1 \cdot \|x_0\|_1 \\ &= \prod_{t=0}^{T_m-1} \mu_{\delta_t} \cdot \xi_{\delta_t}(t) \cdot \|x_0\|_1 \\ &= \prod_{c=1}^m \prod_{t=T_{c-1}}^{T_c-1} \mu_{\delta_t} \cdot \xi_{\delta_t}(t) \cdot \|x_0\|_1. \end{aligned} \quad (2.20)$$

Because each therapy $j \in \mathcal{D}$ is active for k_j time points in every cycle, we have

$$\prod_{t=T_{c-1}}^{T_c-1} \mu_{\delta_t} = \prod_{j \in \mathcal{D}} (\mu_j)^{k_j} \quad \forall c \in \{1, 2, \dots, m\}. \quad (2.21)$$

Thus, (2.20) and (2.21) imply

$$\|x(T_m)\|_1 \leq \left(\prod_{j \in \mathcal{D}} (\mu_j)^{k_j} \right)^m \prod_{c=1}^m \left(\prod_{t=T_{c-1}}^{T_c-1} \xi_{\delta_t}(t) \right) \|x_0\|_1. \quad (2.22)$$

Since $\prod_{j \in \mathcal{D}} (\mu_j)^{k_j} \leq \epsilon$ by Lemma 2, we have

$$\begin{aligned} \|x(T_m)\|_1 &\leq \epsilon^m \prod_{c=1}^m \left(\prod_{t=T_{c-1}}^{T_c-1} \xi_{\delta_t}(t) \right) \|x_0\|_1 \\ &\leq \epsilon^m \left(\frac{1}{\eta} \right)^m \|x_0\|_1, \end{aligned} \quad (2.23)$$

where the second line holds by (2.18).

Let $K_m \in \operatorname{argmax} \{ \|x(t)\|_1 : t \in (T_{m-1}, T_m] \cap \mathbb{N} \}$. Because $\|x(K_m)\|_1$ is a maximum on cycle m , it suffices to show that $\|x(K_m)\|_1 \rightarrow 0$ exponentially with decay rate ϵ/η , as $m \rightarrow \infty$.

Because the errors are bounded, $\exists B < \infty$ such that

$$B \geq \prod_{t=T_{m-1}}^{\tau} \mu_{\delta_t} \xi_{\delta_t}(t) \quad \forall \tau \in (T_{m-1}, T_m - 1] \cap \mathbb{N}, \quad \forall m \in \mathbb{N}. \quad (2.24)$$

For example, if $E \geq 1$, then $B = E^L \cdot \prod_{j \in \mathcal{J}} (\mu_j)^{k_j}$ satisfies (2.24), where $L = \sum_{j \in \mathcal{D}} k_j$ and $\mathcal{J} = \{j \in \mathcal{D} : \mu_j \geq 1\}$.

Eq. (2.17a) implies that

$$x(K_m) = \prod_{t=T_{m-1}}^{K_m-1} A_{\delta_t}(t) \cdot x(T_{m-1}). \quad (2.25)$$

Take the L^1 norm of (2.25) to derive the following inequality:

$$\begin{aligned} \|x(K_m)\|_1 &\leq \prod_{t=T_{m-1}}^{K_m-1} \mu_{\delta_t} \cdot \xi_{\delta_t}(t) \cdot \|x(T_{m-1})\|_1 \\ &\leq B \cdot \|x(T_{m-1})\|_1 \\ &\leq B \left(\frac{\epsilon}{\eta} \right)^{m-1} \|x_0\|_1, \end{aligned} \quad (2.26)$$

where the last line holds by (2.23). □

Lastly, we provide a conservative settling time result for the switched dynamical system (2.17) subject to a cyclic schedule with the waiting times given by Lemma 2.

Definition 2. For any $\gamma \in (0, 1)$, T is a γ -settling time for the system (2.17) if $\|x(t)\|_1 \leq \gamma \|x_0\|_1$ for all $t \geq T$.

Corollary 1. Assume the conditions of Lemma 2, choose any $\epsilon \in [\beta, 1)$, and let $\mathcal{K} := (k_j)_{j \in \mathcal{D}}$ be a solution to (2.6). Let the system (2.17) evolve under a cyclic schedule (2.9) with the waiting times $\mathcal{K}$. Let $B < \infty$ satisfy (2.24). Suppose $\exists \eta \in (\epsilon, \infty)$ that satisfies (2.18). Then, for any $\gamma \in (0, 1)$, a γ -settling time is K_{m_γ} , which is defined as follows:

$$\begin{aligned} K_{m_\gamma} &\in \operatorname{argmax} \{ \|x(t)\|_1 : t \in (T_{m_\gamma-1}, T_{m_\gamma}] \cap \mathbb{N} \} \\ T_{m_\gamma} &= m_\gamma \sum_{j \in \mathcal{D}} k_j \\ m_\gamma &= \left\lceil \frac{\log \gamma - \log B}{\log \epsilon - \log \eta} + 1 \right\rceil, \quad B > \gamma. \end{aligned} \quad (2.27)$$

Proof. Take $m \in \mathbb{N}$. By (2.26), we have

$$\|x(K_m)\|_1 \leq B \left(\frac{\epsilon}{\eta} \right)^{m-1} \|x_0\|_1. \quad (2.28)$$

Let $S_m := B \left(\frac{\epsilon}{\eta} \right)^{m-1} \|x_0\|_1$, and notice that $(S_m)_{m \in \mathbb{N}}$ is a decreasing sequence. Let $\gamma \in (0, 1)$. It suffices to find $m_\gamma \in \mathbb{N}$ such that

$$\|x(K_{m_\gamma})\|_1 \leq B \left(\frac{\epsilon}{\eta} \right)^{m_\gamma-1} \|x_0\|_1 \leq \gamma \|x_0\|_1 \quad (2.29)$$

because K_{m_γ} is a γ -settling time; see that $\|x(t)\|_1 \leq \|x(K_{m_\gamma})\|_1 \leq \gamma \|x_0\|_1$ for all $t \geq K_{m_\gamma}$. Choose $B > \gamma$; we can do this because B is an upper bound. Use (2.29) and recall that $\|x_0\|_1 > 0$, $0 < \frac{\gamma}{B} < 1$, and $0 < \frac{\epsilon}{\eta} < 1$ to find the following:

$$\begin{aligned} B \left(\frac{\epsilon}{\eta} \right)^{m_\gamma-1} &\leq \gamma && \iff \\ (m_\gamma - 1) \log \left(\frac{\epsilon}{\eta} \right) &\leq \log \left(\frac{\gamma}{B} \right) && \iff \\ m_\gamma &\geq \frac{\log \gamma - \log B}{\log \epsilon - \log \eta} + 1. \end{aligned} \quad (2.30)$$

We choose m_γ to equal the smallest natural number that satisfies the last line above. $\square$

The next section provides a numerical example to illustrate the mathematical analysis results.

Numerical Example

We have estimated therapy-specific dynamical models for two differentiation states (basal, non-basal) using time series measurements of triple-negative breast cancer cell populations that were collected by Risom et al. [84]. These therapy-specific dynamical models, which take the form of (2.4) with $p = 2$ differentiation states, are provided below:

$$A_P = \begin{bmatrix} 0.755 & 0.081 \\ 0.169 & 0.843 \end{bmatrix} \quad (2.31)$$

$$A_B = \begin{bmatrix} 0.896 & 0 \\ 0.186 & 1.083 \end{bmatrix} \quad (2.32)$$

$$A_T = \begin{bmatrix} 1.030 & 0.231 \\ 0.022 & 0.821 \end{bmatrix}, \quad (2.33)$$

where therapy P is Trametinib+BEZ235, therapy B is BEZ235, therapy T is Trametinib, and the discrete time interval $[t, t+1)$ is 12 hours long. Using (2.31)-(2.33), we find that $\|A_P\|_1 = \mu_P = 0.924$, $\|A_B\|_1 = \mu_B = 1.083$, and $\|A_T\|_1 = \mu_T = 1.052$. Thus, $\mathcal{I} := \{i \in \mathcal{D} : \mu_i \in (0, 1)\} = \{P\}$ and $\mathcal{J} := \{j \in \mathcal{D} : \mu_j \geq 1\} = \{B, T\}$.

In this numerical example, we have set $L_i = 2$ (1 day) $\forall i \in \mathcal{D}$ since receiving treatment several times per day is inconvenient. We have chosen $U_P = 4$ (2 days), $U_B = 8$ (4 days), and

$U_T = 6$ (3 days) to illustrate a conservative scenario since $\mu_P < 1$ and $1 \leq \mu_T < \mu_B$. So, $\beta = \mu_P^{U_P} \mu_B^{L_B} \mu_T^{L_T} = 0.947$. We have set $\epsilon = 0.95$ to ensure that $\epsilon \in [\beta, 1)$. We have computed the following waiting times using Lemma 2: $k_P = 4$, $k_B = 2$, and $k_T = 2$. We have set $\eta = 0.96$ to ensure that $\eta \in (\epsilon, \infty)$. We have chosen the errors $\xi_{\delta_t}(t) \in (0, E]$ to be pseudorandom values drawn from the uniform distribution on the interval $[0.9, 1.5]$ conditioned on the satisfaction of (2.18); hence, $E = 1.5$. The upper bound $B = E^L (\mu_B)^{k_B} (\mu_T)^{k_T} = 33.3$, where $L = k_P + k_B + k_T = 8$ (4 days).

Fig. 2.9 shows an example simulation of the cancer system (2.17) in response to a cyclic schedule of the three therapies $\{P, B, T\}$ with the waiting times (k_P, k_B, k_T) for 40 cycles (160 days).¹¹ For example, if therapy P is applied at time zero, then $x(k_P)$ is given by

$$\begin{aligned} x(k_P) &= \left(\prod_{t=0}^{k_P-1} A_P(t) \right) x_0 = \left(\prod_{t=0}^{k_P-1} \xi_P(t) A_P \right) x_0 \\ &= \left(\prod_{t=0}^{k_P-1} \xi_P(t) \right) \left(\prod_{t=0}^{k_P-1} A_P \right) x_0 \\ &= \left(\prod_{t=0}^{k_P-1} \xi_P(t) \right) (A_P)^{k_P} x_0, \end{aligned} \quad (2.34)$$

where the second line holds since the errors are scalars. Further, if therapy T is applied next, then $x(k_P + k_T)$ is given by

$$x(k_P + k_T) = \left(\prod_{t=k_P}^{k_P+k_T-1} \xi_T(t) \right) (A_T)^{k_T} x(k_P). \quad (2.35)$$

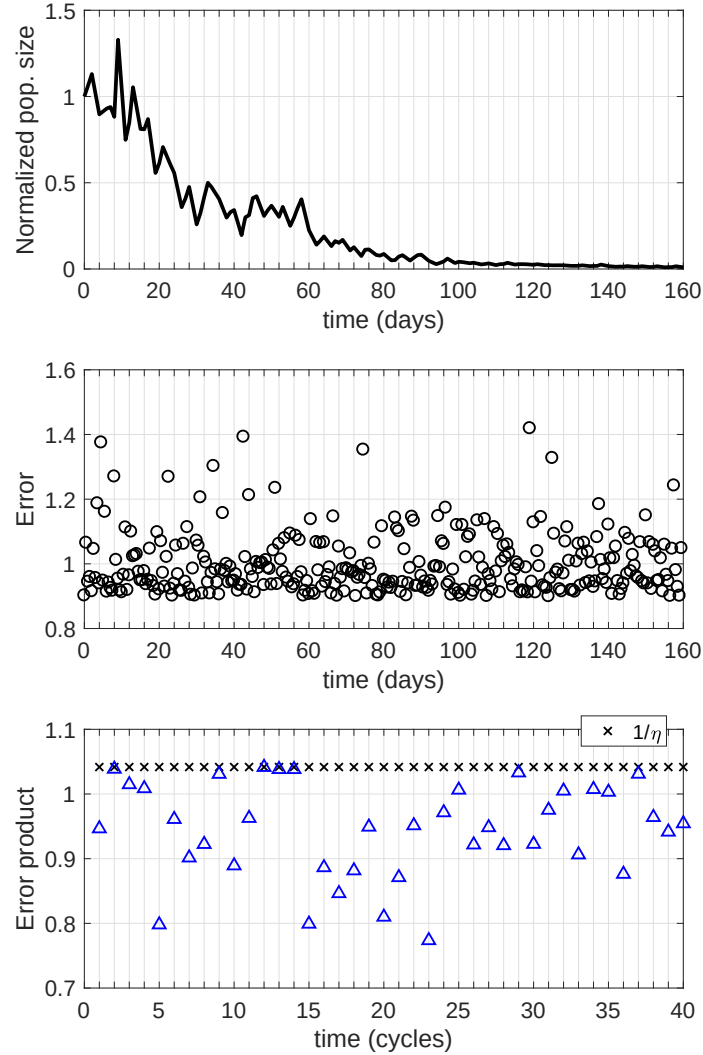
We have randomly chosen an ordering of the three therapies for each cycle. If (2.18) was not satisfied for a cycle, then the errors for that cycle were regenerated. The initial condition $x_0 = [220, 612]^T$ is the estimated number of live cells in each differentiation state (basal, non-basal) at time zero averaged over fifteen populations, where the data comes from [84].

In Fig. 2.9, the saw-tooth behavior arises mainly because the cancer cell population shrinks following treatment with Trametinib+BEZ235 (therapy P) but grows following treatment with the other two therapies. The simulation shows that the settling time provided by Corollary 1 is conservative, where this conservativeness arises partly because B is a worst-case upper bound. For example, if $\gamma = \frac{1}{10}$, then $m_\gamma = 556$ cycles according to Corollary 1. However, after about 19.5 cycles, the cancer cell population stays below one-tenth its original size in Fig. 2.9. Further, the simulation shows that although the errors are similar in magnitude throughout the time horizon, the size of the live cancer cell population decays exponentially.

Discussion

Here we have used the machinery of switched dynamical systems to synthesize and analyze the effects of cyclic therapeutic schedules with toxicity specifications on cancer cell

¹¹The code has been written in MATLAB R2016b (The MathWorks, Inc., Natick, MA) and is available here: https://github.com/chapmanmp/CDC_2018_Github/tree/master/CDC_2018_MatlabCode.



© 2018 IEEE

Figure 2.9: A simulation of the switched dynamical system (2.17) for a cancer cell population in response to a cyclic schedule of the therapies $\{P, B, T\}$ with the waiting times (k_P, k_B, k_T) for 40 cycles (160 days) [20]. Each vertical grid line denotes the start of a cycle. **Top:** Normalized population size $\frac{\|x(t)\|_1}{\|x_0\|_1}$ is shown at each time t of the application of a therapy. E.g., in the first cycle, if the order were T first, B second, and P third, then $\frac{\|x(t)\|_1}{\|x_0\|_1}$ would be plotted at $t \in \{0, k_T, k_T + k_B, k_T + k_B + k_P\}$. **Middle:** The error $\xi_{\delta_t}(t)$ is shown at each time point $t \in \{0, 1, \dots, 319\}$. There are 320 time points in total since $40 \text{ cycles} \times \frac{8 \text{ time points}}{\text{cycle}} = 320$. The duration of each discrete time interval $[t, t+1)$ is 0.5 days. **Bottom:** The error product $\prod_{t=T_{m-1}}^{T_m-1} \xi_{\delta_t}(t)$ per cycle $m \in \{1, 2, \dots, 40\}$ is shown. Note that each error product is less than or equal to $\frac{1}{\eta}$.

populations in theory. Our utilization of switched dynamical systems has been motivated by our computational modeling contributions in Section 2.1.

The theoretical contributions in the current section have inspired the collection of new data on the response of cancer cell populations to schedules with two distinct therapies. These data are being used to develop improved computational models and cancer treatment strategies for finite time horizons, where distributions of the errors are being estimated from the data. This ongoing research has been submitted for review with the lead author Marius Wiggert. In addition, experiments are being conducted by our collaborators at the Oregon Health and Science University to observe the death of healthy cells in response to various cancer therapies, which will provide quantitative toxicity measures.

In this chapter, we have presented a computational and analytical toolkit to improve the understanding and management of a cancer subtype called triple-negative breast cancer (TNBC). Using time series data from a TNBC cell line, we have identified dynamical models of cancer cell populations in response to different therapies. Subsequent computational analyses on these models have provided biological insights about how differentiation-state behaviors may contribute to the aggregations of therapy-tolerant cells that have been observed following treatment. Moreover, we have proposed a class of therapeutic treatment schedules with toxicity specifications and have derived sufficient mathematical conditions for the decay of cancer cell populations under this class. This is a particularly important contribution since treatment decisions for triple-negative breast cancer are typically made by trial and error.

However, the benefits that we anticipate from the above research will only be realized if our dynamical models are sufficiently predictive and if our worst-case uncertainty conditions are always satisfied. The predictive power of the models will improve as more high-quality data becomes available, which will happen over time as measurement techniques advance. Hence, this thesis will focus on the second challenge of requiring worst-case uncertainty conditions to always be satisfied. More broadly, similar requirements are inherent to standard worst-case decision analysis methods for dynamical systems, such as: robust control [8] [14] [77] [116] [91] [74], Hamilton-Jacobi reachability [70] [33] [23], and minimax reachability [9] [12]. Unfortunately, assuming that worst-case uncertainty conditions are always satisfied may lead to excessive conservativeness, which is not appropriate for safety-critical systems that have large uncertainty bounds or have unbounded uncertainties. Examples of such systems are those that are influenced by natural phenomena, including disease systems (within an individual or transmission through society) and infrastructure systems that are affected by uncertain natural resources (e.g., food, water, and energy systems).

The limitations of standard worst-case decision analysis methods for highly uncertain systems motivate the development of new mathematical methods that can balance the worst-case and risk-neutral perspectives. The next chapter will present a new safety analysis method for dynamical systems that is able to attain this balance by leveraging the *Conditional Value-at-Risk* measure.

Chapter 3

Risk-Sensitive Safety Analysis via CVAR

In the previous chapter, we have argued that standard worst-case decision analysis is not appropriate for systems with large uncertainties, such as cancer treatment systems. In the current chapter, we present a new mathematical method for safety analysis that quantifies the notion of worst case in a more flexible way by using the *Conditional Value-at-Risk* (CVAR) measure. We call this method *risk-sensitive safety analysis via CVAR* or *risk-sensitive safety analysis* for brevity. Most of the material presented here is from a manuscript that is under review [22] and from two published conference papers [18] [21].

This chapter has four parts. First, we show a numerical example of a stormwater catchment system that is based on a real system in Lenexa, Kansas for which worst-case safety analysis is too conservative to provide practical design guidance. This example in addition to the cancer treatment system from the previous chapter motivate the need for new mathematical methods that facilitate less conservative but still protective control of dynamical systems under uncertainty. Following this motivating example, we provide background on risk measures and background on safety analysis, which are needed to understand our contribution. The third part of the chapter focuses on our contribution, which is the mathematical development of risk-sensitive safety analysis. Fourth, we demonstrate the practical utility of risk-sensitive safety analysis as a method to evaluate the risk of overflows for a stormwater catchment system under precipitation uncertainty.

3.1 Motivating Example

We have applied a standard worst-case safety analysis method called *Hamilton-Jacobi reachability* [70] [23] [2] to evaluate various designs of a stormwater catchment system in Lenexa, Kansas (Fig. 3.1). Given a dynamical system model in continuous time, Hamilton-Jacobi reachability can be used to compute the set of initial states from which the state trajectory is guaranteed to avoid constraint violation by assuming that the disturbance

(uncertainty) is bounded and behaves in the most adversarial way.¹ In the current example, the dynamical system model is the stormwater catchment system depicted in Fig. 3.1 with the state vector $x = [x_1, x_2, x_3]^T \in \mathbb{R}^3$, where x_1 is the water elevation of pond 1, x_2 is the water elevation of the stream, and x_3 is the water elevation of pond 2. The approach is to estimate the set of initial water elevations from which the system can avoid overflows under a particular design by using Hamilton-Jacobi reachability (called: *HJ safe set*). Then, we evaluate how different design choices, such as adding an automated valve or increasing the surface area of a pond, affect the size and shape of the HJ safe set in comparison to a baseline design.

When using a small constant surface runoff rate (2 cubic feet per second) over a 4-hour time horizon, HJ safe sets provide interesting insights about how different design choices influence the ability of the system to avoid overflows (Fig. 3.2). By considering a wide range of realistic initial conditions and by utilizing a dynamical systems perspective, HJ safe sets provide benefits for design in comparison to standard practice, which only considers one initial condition and one pond in isolation.

However, when using standard surface runoff profiles (Fig. 3.3), HJ safe sets turn out to be empty and so are limited in their usefulness as a design tool. Additional experiments have demonstrated that very large facilities would be needed to provide non-empty HJ safe sets under standard surface runoff profiles (Fig. 3.4). Very large facilities are not financially feasible in general since stormwater systems are typically funded by taxpayer money.

This stormwater catchment system example motivates the need for a new safety analysis method that permits a more flexible definition of conservativeness than a standard worst-case safety analysis approach.

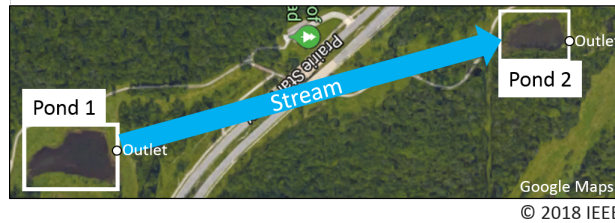


Figure 3.1: A stormwater catchment system in Lenexa, Kansas that consists of two ponds in series connected by a stream [21]. This system is currently operated by OptiRTC, Inc. (Boston, MA).

¹A particularly clear presentation of Hamilton-Jacobi reachability can be found in Dr. Anayo Akametalu's dissertation [2].

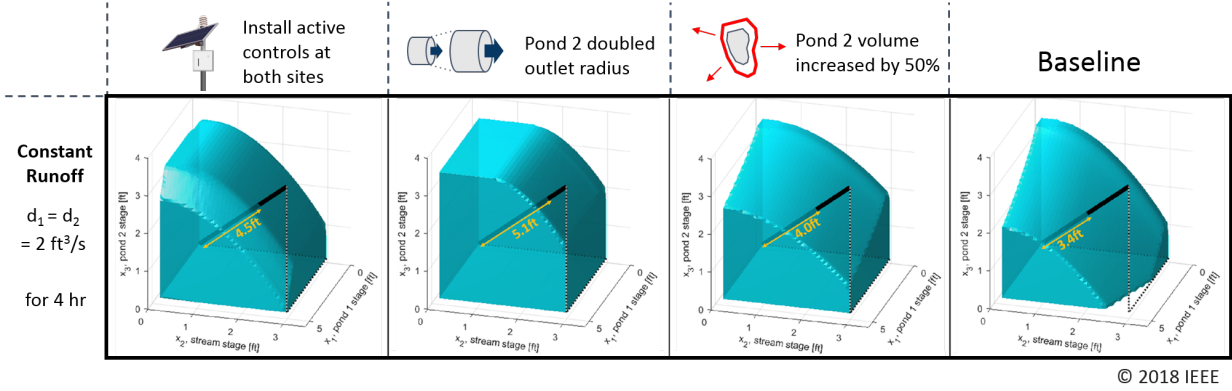


Figure 3.2: HJ safe sets for three different design choices when there is constant surface runoff of 2 cubic feet per second into each pond over a 4-hour time horizon [21]. In each plot, each point in the volume represents an initial condition from which the system is predicted to avoid overflows. The x_1 -axis is the water elevation of pond 1; the x_2 -axis is the water elevation of the stream; and the x_3 -axis is the water elevation of pond 2. A reference vector that points approximately in the (5 ft, 3 ft, 3.5 ft)-direction is shown with each safe set. 5 feet, 3 feet, and 3.5 feet are the maximum water elevations of pond 1, the stream, and pond 2, respectively. An estimate of the distance that each safe set extends along this vector is provided.

3.2 Mathematical Background

Now that we have presented the above motivating example, we provide mathematical background that is needed to understand our new risk-sensitive safety analysis method.

Conditional Value-at-Risk (CVAR) vs. Exponential Utility (Mean-Variance)

The purpose of this subsection is to explain Conditional Value-at-Risk (CVAR) and demonstrate how CVAR is different from the more commonly used Mean-Variance measure using a numerical example. CVAR is a type of *risk measure*, which is a mapping on a space of random variables to the extended real line. In this thesis, the random variables are random costs; i.e., smaller values are better, and larger values are worse. There are many different types of risk measures, and each risk measure quantifies the deviation from the expectation of a random variable in a distinct way.² Here we focus on two risk measures: Exponential Utility (Mean-Variance) and Conditional Value-at-Risk (CVAR).

²Shapiro et al. [92, Chapter 6.3] provides a detailed overview of risk measures, and Kisiala [55, Chapter 2] provides an intuitive presentation of risk measures with emphasis on CVAR.

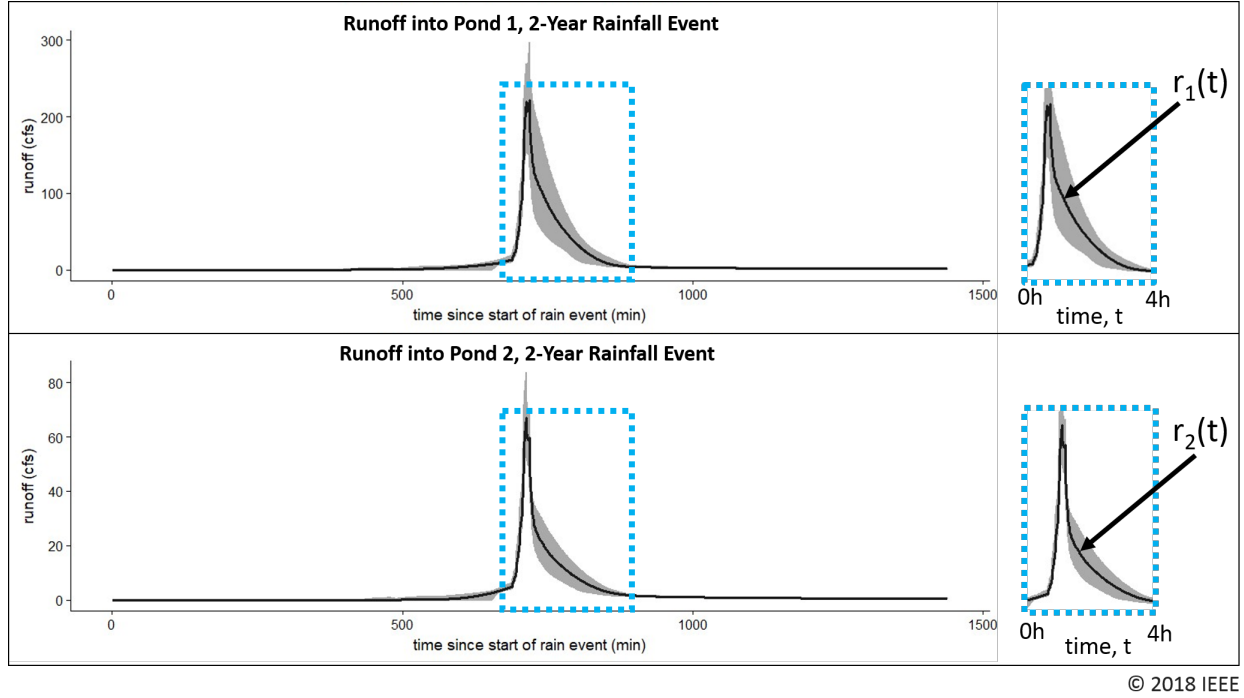


Figure 3.3: A standard surface runoff profile for pond 1 (r_1) and pond 2 (r_2) over a 4-hour time horizon [21]. Surface runoff into each pond has been estimated using the two-year rainfall event for Lenexa, Kansas with the USEPA Stormwater Management Model for various model parameter sets via PCSWMM (Computational Hydraulics International, Ontario) [85].

The *Exponential Utility* of a random cost variable Y is defined as follows:

$$\rho_{e,\theta}(Y) := \frac{-2}{\theta} \log \mathbb{E}(e^{\frac{-\theta}{2}Y}) = \mathbb{E}(Y) - \frac{\theta}{4}\mathbb{V}(Y) + O(\theta^2), \quad (3.1)$$

where $\theta \in (-1, 0) \cup (0, 1)$ is the *risk-sensitivity level*, $\mathbb{E}(Y)$ is the expectation of Y , and $\mathbb{V}(Y)$ is the variance of Y [107, Eqn. 1.10, Eqn. 1.11]. The second equality holds since $0 < |\theta| < 1$, which can be shown using two Taylor expansions by grouping together the higher-order terms in $O(\theta^2)$. The value of θ encodes the user's preference for being more risk-averse or being more risk-seeking, where the notion of risk is encoded in terms of variance. Specifically, “high variance is disadvantageous” is a risk-averse perspective, and “high variance is advantageous” is a risk-seeking perspective. If θ is near -1 , then $\rho_{e,\theta}$ is more risk-averse and *penalizes* high variance. On the other hand, if θ is near 1 , then $\rho_{e,\theta}$ is more risk-seeking and *values* high variance. If θ is near 0 , then $\rho_{e,\theta}$ is nearly risk-neutral and minimal weight is placed upon variance. Exponential Utility encodes the notion of risk in terms of variance, whereas Conditional Value-at-Risk encodes the notion of risk in terms of one tail of a cost distribution, as we shall see next.

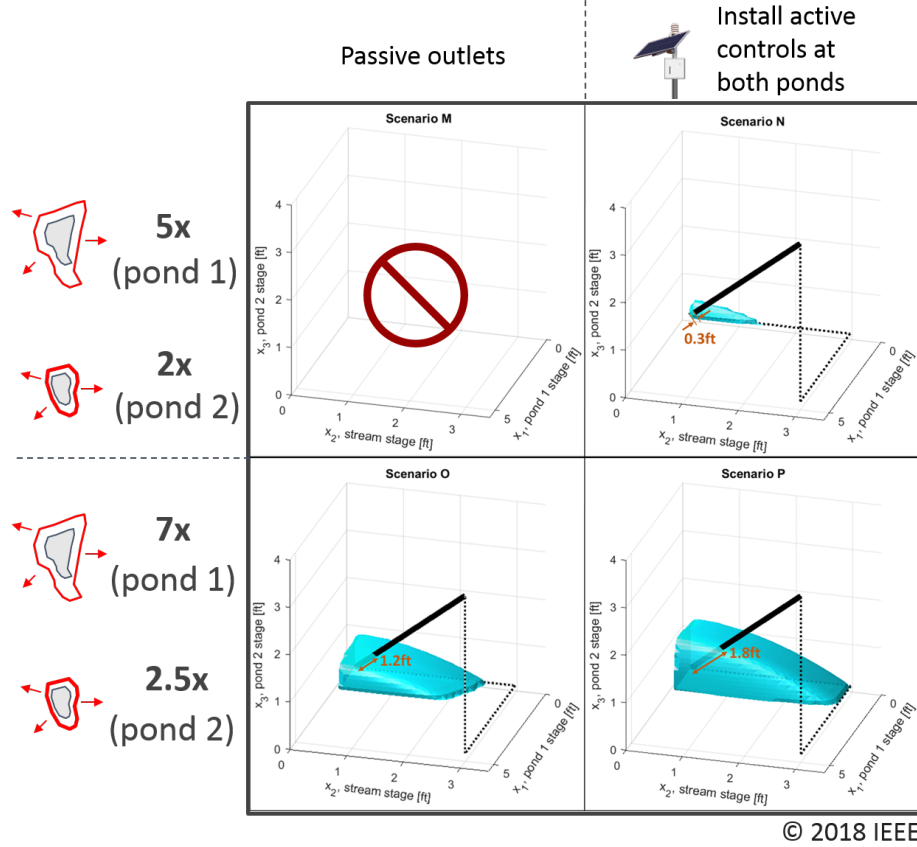


Figure 3.4: HJ safe sets for very large facilities using the standard surface runoff profiles (r_1, r_2) over a 4-hour time horizon [21]. The surface runoff profiles are shown in Fig. 3.3. A reference vector that points approximately in the (5 ft, 3 ft, 3.5 ft)-direction is shown with each non-empty safe set. 5 feet, 3 feet, and 3.5 feet are the maximum water elevations of pond 1, the stream, and pond 2, respectively. An estimate of the distance that each non-empty safe set extends along this vector is provided.

There are various nearly equivalent definitions for Conditional Value-at-Risk (also called Average Value-at-Risk), and we use the following definition in this thesis. The *Conditional Value-at-Risk* of a bounded random cost variable Y at *risk-sensitivity level* $\alpha \in [0, 1]$ is defined as:³

$$\text{CVAR}_\alpha(Y) := \begin{cases} \frac{1}{\alpha} \int_{1-\alpha}^1 \text{VAR}_{1-\tau}(Y) d\tau & \text{if } 0 < \alpha \leq 1 \\ \|Y\|_\infty & \text{if } \alpha = 0, \end{cases} \quad (3.2a)$$

where

$$\text{VAR}_\alpha(Y) := \inf\{y \in \mathbb{R} \mid \Pr(Y \leq y) \geq 1 - \alpha\} \quad (3.2b)$$

is the *Value-at-Risk* at level α , and $\|Y\|_\infty$ is the essential supremum (essential least upper bound) of Y [92, Thm. 6.2] [78]. The above definition (3.2) assumes a risk-averse setting, where α is associated with the upper (worse) tail of the probability distribution of Y . The alternate name “Average Value-at-Risk” comes from the first case in (3.2a), where CVAR is an average of Value-at-Risk. After a few steps, one can show that $\text{CVAR}_1(Y) = \mathbb{E}(Y)$. If the cumulative distribution function of Y , $H_Y(y) := \Pr(Y \leq y)$, is continuous at $y = \text{VAR}_\alpha(Y)$, then $\text{CVAR}_\alpha(Y) = \mathbb{E}(Y|Y \geq \text{VAR}_\alpha(Y))$, which is illustrated in Fig. 3.5, and which explains the name “Conditional Value-at-Risk” [92, Thm. 6.2]. Intuitively, $\text{CVAR}_\alpha(Y)$ can be interpreted as the expectation of the $\alpha \cdot 100\%$ worst realizations of Y . Hence, $\text{CVAR}_\alpha(Y)$ is able to balance the worst-case and risk-neutral perspectives by focusing on the upper tail of the cost distribution via the parameter α .

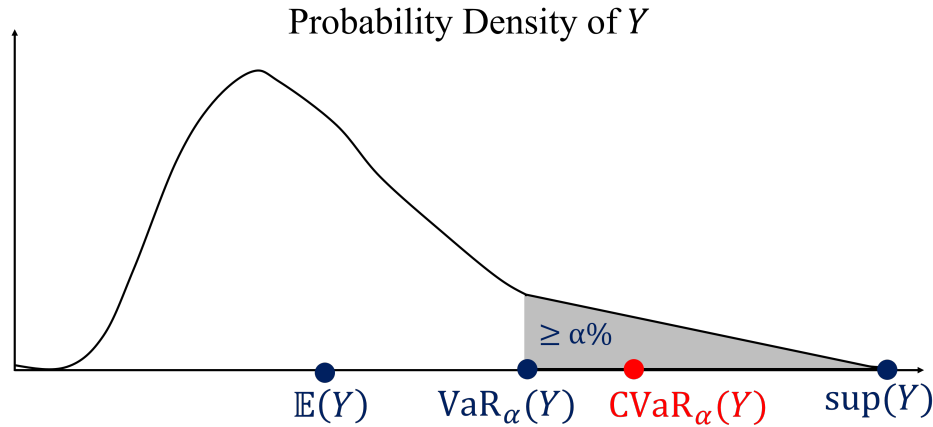


Figure 3.5: Conditional Value-at-Risk (CVAR) is a measure of one-sided tail risk of a random variable Y . Intuitively, $\text{CVAR}_\alpha(Y)$ can be interpreted as the expectation of the $\alpha \cdot 100\%$ worst realizations of Y . More specifically, if Y is a continuous cost random variable, then $\text{CVAR}_\alpha(Y)$ is the expectation of Y conditioned on the event $\{Y \geq \text{VAR}_\alpha(Y)\}$, where $\text{VAR}_\alpha(Y)$ is the minimum cost in the $\alpha \cdot 100\%$ worst cases.

³Alternatively, we could consider random variables with finite expectation and $\alpha \in (0, 1]$ to ensure that CVAR is bounded.

For $\alpha \in [0, 1]$, CVAR_α is a *coherent* risk measure on the space of bounded random variables. This means that for fixed $\alpha \in [0, 1]$, $\text{CVAR}_\alpha : L^\infty \rightarrow \mathbb{R}$ is *convex*, *monotonic*, *translation equivariant* ($\text{CVAR}_\alpha(Y + a) = \text{CVAR}_\alpha(Y) + a$ if $a \in \mathbb{R}$), and *positively homogeneous* ($\text{CVAR}_\alpha(\lambda Y) = \lambda \text{CVAR}_\alpha(Y)$ if $\lambda \geq 0$) [78] [79]. Moreover, since CVAR_α is a coherent risk measure, it is equivalent to an expectation maximized over a specific set of probability density functions [92, Thm. 6.4, Eqn. 6.40, Eqn. 6.70]. The above properties are mathematically useful and have intuitive interpretations. For example, the convexity of coherent risk measures is consistent with the notion that the diversification of assets decreases risk. Moreover, CVAR_α may be considered a “robustified” expectation, since it is equivalent to a worst-case expectation with respect to certain perturbations in the assumed probability distribution [25]. We will use the above properties in our proofs later in this chapter.

Next, we provide a numerical comparison between simulated outcomes that result from control policies that have been computed with respect to a CVAR criterion or an Exponential Utility criterion at different levels of risk aversion. For this example, we use a thermostatically controlled load (TCL) system with a heavy-tailed finite disturbance distribution, where the state $x_t \in \mathbb{R}$ is the temperature of a thermal mass in degrees Celsius. An outcome is a realization of the following random cost of the state trajectory, $C_{0:T} = \sum_{t=0}^T \max(x_t - 21, 20 - x_t)$, under a control policy that we have computed offline. $C_{0:T}$ is the cumulative deviation outside a safe temperature range of 20 to 21 °C over a one-hour time horizon. We use the following model for a TCL dynamical system:

$$x_{t+1} = ax_t + (1 - a)(b - \eta R P u_t) + d_t, \quad t = 0, 1, \dots, T - 1, \quad (3.3)$$

which we have adopted from [110], and which was first developed by [72]. In the above model, $x_t \in \mathbb{R}$ is the temperature (°C) of a thermal mass, $u_t \in [0, 1]$ is a continuous control input from no power to full power, and $d_t \in \mathbb{R}$ is a random disturbance due to environmental uncertainty (Table 3.1).

From two distinct initial conditions, we have sampled $C_{0:T} = \sum_{t=0}^T \max(x_t - 21, 20 - x_t)$ one million times using control policies that have been computed with respect to a CVAR criterion or an Exponential Utility criterion at different levels of risk aversion.⁴ We have utilized the heavy-tailed finite probability distribution for d_t called “Original” in Fig. 3.6 to train the control policies and to generate the histograms of $C_{0:T}$ shown in Fig. 3.7 and Fig. 3.8. Further, we have used the distribution for d_t called “Perturbed” in Fig. 3.6 to

⁴We have computed a control policy with respect to a CVAR criterion by using an algorithm that we will present in Sec. 3.3. The CVAR criterion for a fixed α takes the following form: minimize $\text{CVAR}_\alpha(C_{0:T})$ over a set of history-dependent control policies, subject to a given dynamical system model with a pre-specified probability distribution for the disturbance. The set of history-dependent control policies will be specified in Sec. 3.3. We have computed a control policy with respect to an Exponential Utility criterion by using standard value iteration [22, Section VII]. The Exponential Utility criterion for a fixed θ takes the following form: minimize $\rho_{e,\theta}(C_{0:T})$ over a set of Markov control policies, subject to a given dynamical system model with a pre-specified probability distribution for the disturbance. The control policies that we have computed are not optimal due to discretization that is typical for value iteration on a continuous state space. Also, we have restricted ourselves to *pre-commitment* policies when using the CVAR criterion, which will be explained in Sec. 3.3.

Table 3.1: Thermostatically Controlled Load Model Parameters

Symbol	Description	Value
a	time delay	$e^{\frac{-\Delta\tau}{CR}}$ (no units)
b	temperature shift	32 °C
C	thermal capacitance	2 $\frac{\text{kW hr}}{^\circ\text{C}}$
η	control efficiency	0.7 (no units)
K	constraint set	[20, 21] °C
P	range of energy transfer to or from the thermal mass	14 kW
R	thermal resistance	2 $\frac{^\circ\text{C}}{\text{kW}}$
$\Delta\tau$	duration of $[t, t + 1)$	$\frac{5}{60}$ hr
T	length of discrete time horizon	12 (= 1 hr)
U	action space	[0, 1] (no units)
X	state space	[18, 23] °C

hr = hours, kW = kilowatts, °C = degrees Celsius.

generate the histograms of $C_{0:T}$ in Fig. 3.9 to assess robustness to distribution estimation error.

Our results show that the CVAR criterion minimizes the mean of higher-consequence outcomes, evident by reduced weight on the upper tail of the empirical distribution of $C_{0:T}$ as the degree of risk aversion increases (Fig. 3.7, top row, left to right). The Exponential Utility criterion penalizes the mean and variance of $C_{0:T}$, and the empirical variance becomes smaller as the degree of risk aversion increases (Fig. 3.7, bottom row, left to right). While the Exponential Utility criterion penalizes variance, this criterion is not guaranteed to minimize the mean of higher-consequence outcomes in the setting of asymmetric cost distributions (see estimated values of $\text{CVAR}_{0.01}$, top vs. bottom row, Fig. 3.7). Notably, at a typical level of risk aversion, the empirical variance of $C_{0:T}$ is smaller under the CVAR criterion compared to the Exponential Utility criterion (Fig. 3.7, center column).

Our results indicate that the initial condition of the system impacts the relative advantage of using the CVAR criterion versus the Exponential Utility criterion for policy synthesis. The probability distribution of the disturbance in the TCL example represents a setting where the thermal mass is exposed to random temperature perturbations with a positive bias (Fig. 3.6), and the control input u_t is only able to provide heat. Thus, if the system is initialized at the center of the constraint set ($K = [20, 21]$ °C), the controller has limited authority to avoid high temperatures regardless of the chosen criterion (CVAR or Exponential Utility), as shown in Fig. 3.8. However, if the system is initialized at a cooler temperature where the controller has more authority, there is a substantial advantage to employing the CVAR criterion rather than the Exponential Utility criterion to reduce the mean of high-consequence outcomes (Fig. 3.7). In addition, the CVAR criterion continues to offer advantages over the Exponential Utility criterion when the cumulative cost $C_{0:T}$ is sampled under a perturbed probability distribution for the disturbance (Fig. 3.9).

Overall, using a CVAR criterion is quite important for safety-critical situations in which cost distributions cannot be well-approximated by symmetric distributions. Also, CVAR has the substantial advantage of improved *interpretability* compared to Exponential Utility. The CVAR criterion at the risk-sensitivity level α is guaranteed to minimize the mean of the α -fraction of worst-case outcomes, if the outcome distribution is continuous. However, the risk-sensitivity level θ of the Exponential Utility criterion does not map to a precise reduction in variance.

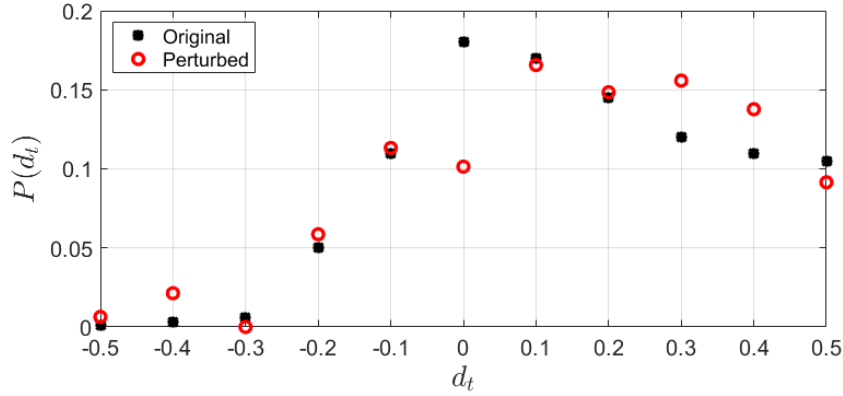


Figure 3.6: Probability distributions for the disturbance d_t to the TCL system (3.3). The horizontal axis shows the possible (discrete) realizations of the disturbance in degrees Celsius. Their associated probabilities are shown on the vertical axis.

Reachability Analysis for Safety-Critical Systems

In the previous subsection, we have discussed two key benefits of CVAR (ability to quantify asymmetric cost distributions and improved interpretability). In the current subsection, we provide background on reachability analysis to explain where our new method *risk-sensitive safety analysis via CVAR* fits within the existing literature.

Reachability analysis is a framework for formal verification of dynamical systems that is used to compute subsets of the state space X and control policies with certain desirable properties. In this thesis, we focus on *safety-critical* settings, where a dynamical system is required to satisfy particular safety constraints or is required to reduce the extent of constraint violation to a particular threshold. We are interested in computing *safe sets* for a given dynamical system model, which are sets of initial states from which the state trajectory is guaranteed to satisfy the safety constraints or reduce the extent of constraint violation to a particular threshold. Safe sets are defined in terms of level sets of an optimal control problem, where the optimal control problem is designed to quantify the extents of constraint violations when using different control policies. In addition to computing safe sets, we are interested in computing control policies that optimize the control problem and

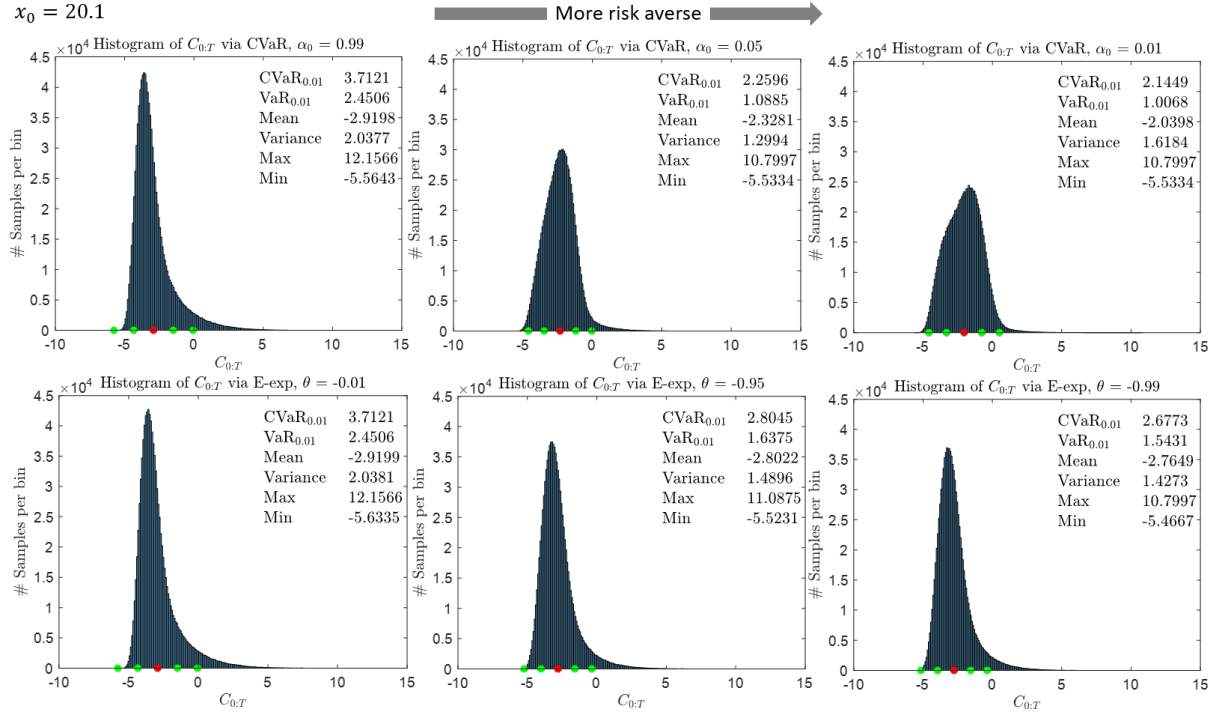
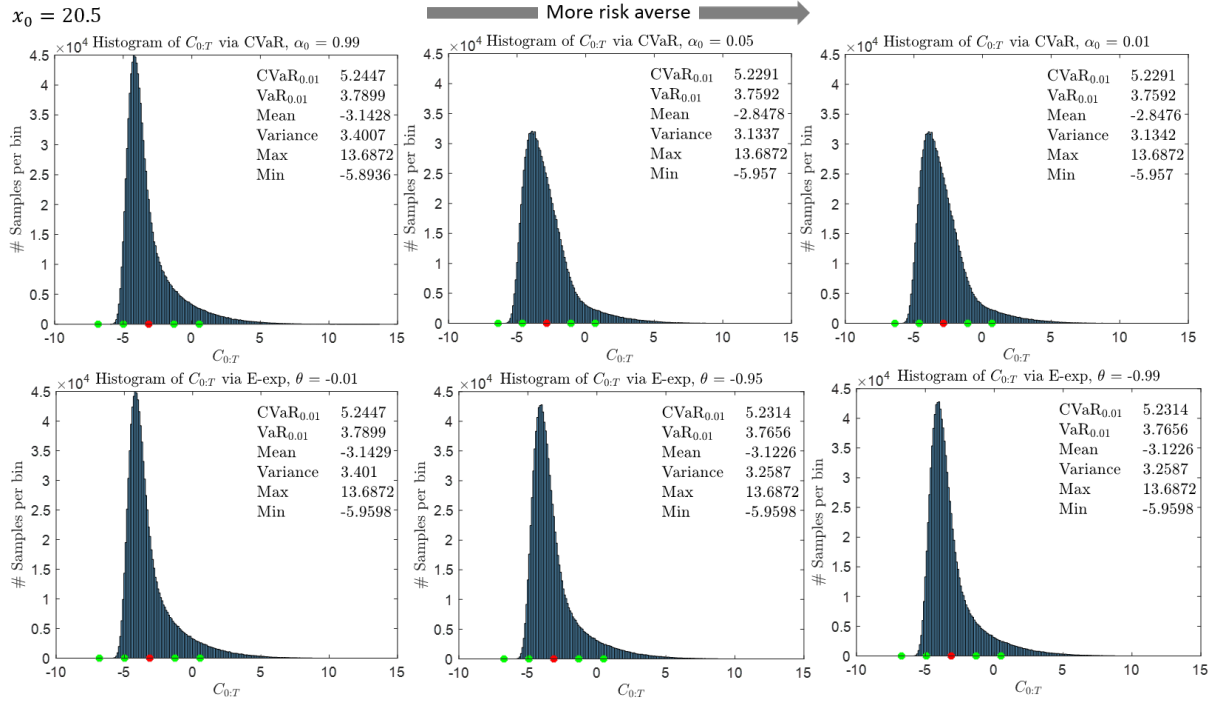


Figure 3.7: Histograms of $C_{0:T} = \sum_{t=0}^T \max(x_t - 21, 20 - x_t)$ for the TCL system (3.3) initialized at $x_0 = 20.1$ °C are shown under different control policies. **Top row:** A control policy has been computed with respect to CVAR_α . **Bottom row:** A control policy has been computed with respect to $\rho_{e,\theta}$ (3.1). **Left column:** Nearly risk neutral $(\alpha, \theta) = (0.99, -0.01)$. **Center column:** Typical risk aversion $(\alpha, \theta) = (0.05, -0.95)$. **Right column:** Nearly worst case $(\alpha, \theta) = (0.01, -0.99)$. One million samples of $C_{0:T}$ are shown in each histogram, and relevant empirical statistics of these samples are displayed. The red circle marks the empirical mean, and each green circle marks the empirical mean plus/minus one/two standard deviations. The same pseudorandom sequence of disturbance realizations has been used for each histogram and chosen according to the “Original” distribution in Fig. 3.6.

Figure 3.8: Same setting as in Fig. 3.7 except that $x_0 = 20.5$ °C.

can be implemented on real dynamical systems. We call the development of algorithms that facilitate such computations as accurately as possible and the study of the results from these computations *reachability-based safety analysis*.⁵ We assume that a safety constraint is given in the following form: $x_t \in K$ for $t = 0, 1, \dots, T$, where $K \subseteq X$ is a constraint set in the state space X that is specified by a domain expert.⁶ It is common to use a function $g : X \rightarrow \mathbb{R}$ to quantify the extent of constraint violation or satisfaction of a state x_t in a trajectory of the dynamical system.

There are different types of reachability-based safety analysis methods, and the type that is used depends on problem-specific assumptions, domain knowledge, and safety requirements. Using *worst-case safety analysis*, one can estimate the set of initial states from which the amount of constraint violation of any state trajectory is zero by assuming that the uncertainty is bounded and takes on its most harmful values over time. Worst-case safety analysis for dynamical systems in discrete time, which is called minimax reachability, was developed by Bertsekas and Rhodes in the early 1970s [12] [9]. Worst-case safety analysis for dynamical systems in continuous time, which is called Hamilton-Jacobi reachability, was

⁵A very important research direction, which will be discussed in Chapter 4, is the inexact computation of safe sets and control policies with quality-of-approximation guarantees to facilitate scalability to high-dimensional systems that may not have analytical models.

⁶This is a standard safety constraint. Using signal temporal logic to specify more descriptive safety constraints is an exciting area of research.

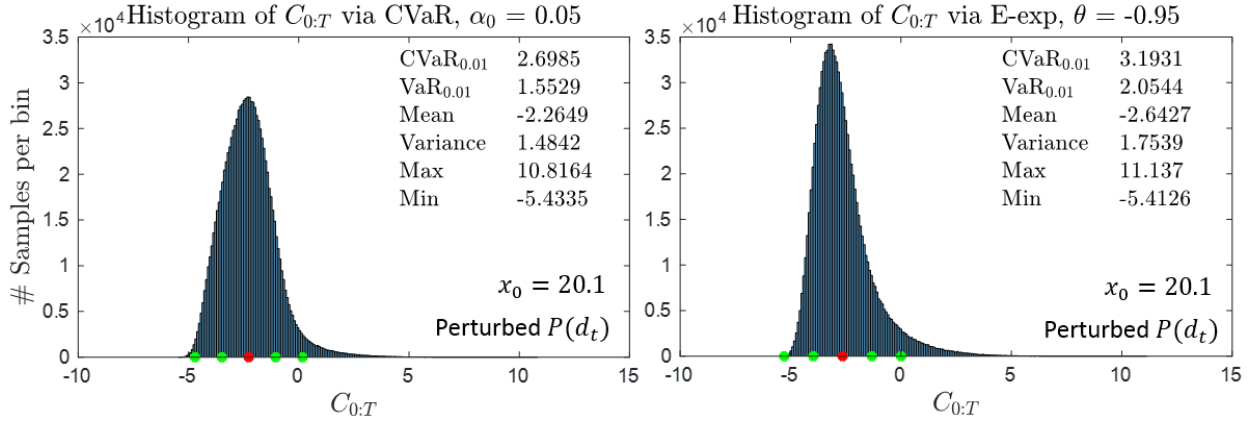


Figure 3.9: Histograms of $C_{0:T} = \sum_{t=0}^T \max(x_t - 21, 20 - x_t)$ are shown for the TCL system (3.3) initialized at $x_0 = 20.1$ °C under a perturbed probability distribution for the disturbance. For each histogram, the same pseudorandom sequence of disturbance realizations has been used to sample $C_{0:T}$ via the “Perturbed” distribution shown in Fig. 3.6. Control policies have been computed using the “Original” distribution for the disturbance shown in Fig. 3.6. **Left:** A control policy has been computed with respect to CVaR_α for $\alpha = 0.05$. **Right:** A control policy has been computed with respect to $\rho_{e,\theta}$ for $\theta = -0.95$ (3.1). One million samples of $C_{0:T}$ are shown in each histogram, and relevant empirical statistics of these samples are displayed. The red circle marks the empirical mean, and each green circle marks the empirical mean plus/minus one/two standard deviations.

developed by Mitchell et al. in the early 2000s [70].

A less conservative method called *stochastic safety analysis* was developed by Abate et al. in 2008 [1]. Using stochastic safety analysis, one can estimate the set of initial states from which the probability of constraint violation can be reduced to a required threshold by assuming that the uncertainty is non-adversarial and stochastic with a given probability distribution [1]. Stochastic safety analysis has been extended to the problem of reaching a target set within finite time [96], a dynamic games setting in which the uncertainty is adversarial [54] [29], and a distributionally robust setting where the probability distribution for the uncertainty is assumed to reside within a given family of distributions [110].

Our new method *risk-sensitive safety analysis via CVAR* is a generalization of stochastic safety analysis [1] in terms of how we define the optimal control problem that characterizes the safe sets. Using risk-sensitive safety analysis, one can estimate the set of initial states from which the expected amount of constraint violation in rare worst-case circumstances can be reduced to a required threshold. The mathematical formulation for our new method will be presented next.

3.3 Mathematical Formulation

This section provides four key contributions. First, we introduce the notion of a *risk-sensitive safe set* to specify the degree of safety for a given stochastic dynamical system, which leverages the Conditional Value-at-Risk (CVAR) measure at a given risk-sensitivity level α . Second, we show that risk-sensitive safe sets can be under-approximated by the solution to a Markov Decision Process (MDP), where the cost is assessed according to CVAR. This CVAR-MDP problem is *time-inconsistent*, meaning that a policy that is optimal when viewed at time zero is not necessarily optimal when viewed at a later time, so standard value iteration cannot be used as a solution method (see discussion in [69]). However, a non-standard value iteration algorithm on the state space augmented by the (one-dimensional) risk-sensitivity level space has been proposed to approximately solve the CVAR-MDP problem [25]. Third, we prove that this algorithm solves the CVAR-MDP problem for a class of linear systems. Fourth, for this class of systems, we show the existence of optimal history-dependent policies under a *pre-commitment* to certain risk-sensitivity level dynamics. These history-dependent policies are *tractable* since the history is summarized in two parameters, the current state (which may be multi-dimensional) and the current risk-sensitivity level (which is one-dimensional).

Risk-Sensitive Safe Sets

Suppose that a model for a stochastic dynamical system is given in the following form:

$$x_{t+1} = f_t(x_t, u_t, d_t), \quad t = 0, 1, \dots, T-1, \quad (3.4)$$

where $x_t \in X \subseteq \mathbb{R}^n$ is the state, $u_t \in U \subseteq \mathbb{R}^m$ is the control input, $d_t \in D \subseteq \mathbb{R}^a$ is the random disturbance, $f_t : X \times U \times D \rightarrow X$ is Borel measurable, and X , U , and D are non-empty Borel sets. The disturbances $d_0, d_1, \dots, d_{T-1}$ are random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, so each d_t is a function from Ω to D that is measurable relative to $\mathcal{F}$ and $\mathcal{B}(D)$. Assume that $d_0, d_1, \dots, d_{T-1}$ are independent and identically distributed, where their common distribution is defined by the probability measure $\mathbb{P}$ and is independent of all states and actions. Suppose that $K \subseteq X$ is a set of safety constraints, and $g_K : \mathbb{R}^n \rightarrow \mathbb{R}$ is a bounded function that quantifies the extent of constraint violation (e.g., a clipped signed distance function from the boundary of K [2, p. 8]). We specify the set of admissible control policies for the dynamical system model (3.4) as being *non-anticipatory*, *deterministic*, and *history-dependent*.

Definition 3 (Set of Admissible Control Policies). *The set of admissible control policies for time $t = 0, 1, \dots, T-1$ is defined as follows:*

$$\Pi_t := \left\{ \pi_t := (\mu_t, \mu_{t+1}, \dots, \mu_{T-1}) \mid \text{each } \mu_k : H_{t:k} \rightarrow U \text{ is Borel measurable} \right\}, \quad (3.5)$$

where $H_{t:k} := (X \times U)^{k-t} \times X$ is the set of histories from time t to time k , whose elements take the form $h_{t:k} := (x_t, u_t, \dots, x_{k-1}, u_{k-1}, x_k)$.

Next, we define a risk-sensitive safe set in terms of a level set of a non-standard optimal control problem over policies in Π_0 , where a maximum cost of the state trajectory is assessed using CVAR at a given risk-sensitivity level.

Definition 4 (Risk-Sensitive Safe Set). *The risk-sensitive safe set at $(\alpha, r) \in [0, 1] \times \mathbb{R}$ is defined as:*

$$\mathcal{S}_\alpha^r := \{\mathbf{x} \in X \mid W_0^*(\mathbf{x}, \alpha) \leq r\}, \quad (3.6a)$$

where

$$W_0^*(\mathbf{x}, \alpha) := \inf_{\pi \in \Pi_0} \text{CVAR}_\alpha^\pi \left(\max_{t=0,1,\dots,T} g_K(x_t) \mid x_0 = \mathbf{x} \right), \quad (3.6b)$$

and the state trajectory $(x_0, x_1, \dots, x_T)$ evolves under the policy $\pi \in \Pi_0$ according to (3.4) initialized at $x_0 = \mathbf{x}$.

The risk-sensitive safe set definition is well-motivated for several reasons. First, the definition incorporates different risk-sensitivity levels and non-binary distance to the constraint set, and thereby generalizes the classic maximal probabilistic safe set from [1]. Specifically, let $\epsilon \in [0, 1]$ be a maximum tolerable probability of constraint violation, and choose $\alpha = 1$, $r = \epsilon$, and $g_K(x) = \mathbb{1}_{\bar{K}}(x)$, where $\mathbb{1}_{\bar{K}}(x) = 1$ if $x \notin K$ and $\mathbb{1}_{\bar{K}}(x) = 0$ if $x \in K$. Then, the risk-sensitive safe set at $(\alpha, r) = (1, \epsilon)$ is equal to

$$\left\{ \mathbf{x} \in X \mid \inf_{\pi \in \Pi_0} \mathbb{E}^\pi \left(\max_{t=0,1,\dots,T} \mathbb{1}_{\bar{K}}(x_t) \mid x_0 = \mathbf{x} \right) \leq \epsilon \right\}, \quad (3.7)$$

which is the *maximal probabilistic safe set* at the ϵ -safety level [1] for a given stochastic system that evolves under policies in Π_0 .⁷ Moreover, $\mathcal{S}_\alpha^r$ encodes higher degrees of safety as α or r decrease: $\alpha_1 \geq \alpha_2$ and $r_1 \geq r_2 \implies \mathcal{S}_{\alpha_2}^{r_2} \subseteq \mathcal{S}_{\alpha_1}^{r_1}$. Also, $\mathcal{S}_\alpha^r$ specifies that the CVAR_α of the worst constraint violation of the state trajectory must be below a required threshold, whereas the safe set in [89] specifies that the CVAR_α of the constraint violation of x_t must be below a required threshold for each t . So, $\mathcal{S}_\alpha^r$ provides a risk-sensitive safety specification for the entire state trajectory.

Reduction to CVAR-Markov Decision Process

$\mathcal{S}_\alpha^r$ is a well-motivated safety specification but is difficult to estimate in a tractable way due to the CVAR and the maximum in the definition of W_0^* .⁸ To facilitate the estimation of risk-sensitive safe sets, we use the following method.

⁷Abate et al. [1] used the formalism of stochastic hybrid systems, which we do not utilize here.

⁸To compute $\inf_{\pi \in \Pi_0} \mathbb{E}^\pi(\max_{t=0,1,\dots,T} g_K(x_t) \mid x_0 = \mathbf{x})$ subject to a given stochastic dynamical system model, where Π_0 is an admissible set of control policies, one can define $z_t := \max_{k=0,1,\dots,t} g_K(x_k) = \max(z_{t-1}, g_K(x_t))$ and perform value iteration on the augmented state (x_t, z_t) . Adopting this approach to facilitate the estimation of risk-sensitive safe sets is an area for future work (Chapter 4).

Theorem 3 (Reduction to CVAR-MDP). *Fix $\beta > 0$ and $\gamma > 0$. For any $(\alpha, r) \in [0, 1] \times \mathbb{R}$, define the set $\mathcal{U}_\alpha^r \subseteq X$ as follows:*

$$\mathcal{U}_\alpha^r := \{\mathbf{x} \in X \mid J_0^*(\mathbf{x}, \alpha) \leq \beta e^{\gamma r}\}, \quad (3.8a)$$

where

$$\begin{aligned} J_0^*(\mathbf{x}, \alpha) &:= \inf_{\pi \in \Pi_0} \text{CVAR}_\alpha^\pi(C_{0:T} | x_0 = \mathbf{x}) \\ C_{0:T} &:= c_T(x_T) + \sum_{t=0}^{T-1} c_t(x_t, u_t), \end{aligned} \quad (3.8b)$$

such that the state trajectory evolves under the policy π according to the dynamical system model (3.4) initialized at $x_0 = \mathbf{x}$, $c_t : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a stage cost, and $c_T : \mathbb{R}^n \rightarrow \mathbb{R}$ is a terminal cost. If $c_t := \beta e^{\gamma g_K}$ for all t , where $g_K : \mathbb{R}^n \rightarrow \mathbb{R}$ is bounded, then $\mathcal{U}_\alpha^r \subseteq \mathcal{S}_\alpha^r$. Moreover, the gap between $\mathcal{U}_\alpha^r$ and $\mathcal{S}_\alpha^r$ can be reduced by increasing γ .

Proof. Use the log-sum-exp relation in [13, Sec. 3.1.5] to show: for any $y \in \mathbb{R}^p$ and $\gamma > 0$,

$$\max_{i=1,2,\dots,p} y_i \leq \frac{1}{\gamma} \log(\sum_{i=1}^p e^{\gamma y_i}) \leq \max_{i=1,2,\dots,p} y_i + \frac{\log p}{\gamma}. \quad (3.9)$$

So, as $\gamma \rightarrow \infty$, $\frac{1}{\gamma} \log(\sum_{i=1}^p e^{\gamma y_i}) \rightarrow \max_{i=1,2,\dots,p} y_i$.

Further, for any $\alpha \in [0, 1]$ and any bounded positive random variable Z , the following relation holds:

$$\text{CVAR}_\alpha(\log(Z)) \leq \log(\text{CVAR}_\alpha(Z)), \quad (3.10)$$

since $\text{CVAR}_\alpha(Z) = \sup_{\xi \in \mathcal{A}} \mathbb{E}_\xi(Z)$, where $\mathcal{A}$ is a specific set of probability density functions by [92, Thm. 6.4, Eqn. 6.40, Eqn. 6.70], and since $\mathbb{E}_\xi(\log(Z)) \leq \log(\mathbb{E}_\xi(Z))$ for any $\xi \in \mathcal{A}$ by Jensen's Inequality. For any $\alpha \in [0, 1]$, $\mathbf{x} \in X$, and $\pi \in \Pi_0$, the following inequalities hold:

$$\begin{aligned} \text{CVAR}_\alpha^\pi \left(\max_{t=0,1,\dots,T} g_K(x_t) \mid x_0 = \mathbf{x} \right) &\stackrel{(3.9)}{\leq} \frac{1}{\gamma} \text{CVAR}_\alpha^\pi \left(\log \left(\sum_{t=0}^T e^{\gamma g_K(x_t)} \right) \mid x_0 = \mathbf{x} \right) \\ &\stackrel{(3.10)}{\leq} \frac{1}{\gamma} \log \left(\text{CVAR}_\alpha^\pi \left(\sum_{t=0}^T e^{\gamma g_K(x_t)} \mid x_0 = \mathbf{x} \right) \right), \end{aligned} \quad (3.11)$$

by the relations cited above each inequality symbol, and since CVAR_α^π is monotonic and positively homogeneous. Take $\mathbf{x} \in \mathcal{U}_\alpha^r$. For any $\epsilon > 0$, there exists $\pi_\epsilon \in \Pi_0$ such that

$$0 \stackrel{(i)}{<} \text{CVAR}_{\alpha^\epsilon}^{\pi_\epsilon} \left(\sum_{t=0}^T e^{\gamma g_K(x_t)} \mid x_0 = \mathbf{x} \right) \stackrel{(ii)}{\leq} \epsilon + \inf_{\pi \in \Pi_0} \text{CVAR}_\alpha^\pi \left(\sum_{t=0}^T e^{\gamma g_K(x_t)} \mid x_0 = \mathbf{x} \right) \stackrel{(iii)}{\leq} \epsilon + e^{\gamma r}$$

for the following reasons: (i) $\forall \mathbf{y} \in X$, $e^{\gamma g_K(\mathbf{y})} > 0$, (ii) definition of infimum and g_K is bounded, (iii) $\mathbf{x} \in \mathcal{U}_\alpha^r$, CVAR_α^π is positively homogeneous, and $\beta > 0$. Take the logarithm of both sides and divide by $\gamma > 0$ to obtain

$$\frac{1}{\gamma} \log \left(\text{CVAR}_{\alpha^\epsilon}^{\pi_\epsilon} \left(\sum_{t=0}^T e^{\gamma g_K(x_t)} \mid x_0 = \mathbf{x} \right) \right) \leq \frac{1}{\gamma} \log(\epsilon + e^{\gamma r}). \quad (3.12)$$

Use the definition of W_0^* (3.6b) as an infimum to find

$$\begin{aligned} W_0^*(\mathbf{x}, \alpha) &\leq \text{CVAR}_\alpha^{\pi_\epsilon} \left(\max_{t=0,1,\dots,T} g_K(x_t) \middle| x_0 = \mathbf{x} \right) \\ &\stackrel{(3.11)}{\leq} \frac{1}{\gamma} \log \left(\text{CVAR}_\alpha^{\pi_\epsilon} \left(\sum_{t=0}^T e^{\gamma g_K(x_t)} \middle| x_0 = \mathbf{x} \right) \right) \\ &\stackrel{(3.12)}{\leq} \frac{1}{\gamma} \log (\epsilon + e^{\gamma r}), \end{aligned}$$

where the appropriate relation is cited above each inequality symbol. Finally, since $W_0^*(\mathbf{x}, \alpha) \leq \frac{1}{\gamma} \log (\epsilon + e^{\gamma r})$ for any $\epsilon > 0$, take $\epsilon \rightarrow 0$ and apply continuity of the logarithm to obtain $W_0^*(\mathbf{x}, \alpha) \leq r$. So, $\mathbf{x} \in \mathcal{S}_\alpha^r$. $\square$

Remark 5. In Theorem 3, the parameter β is included to counter numerical issues that arise when γ is large.

Theorem 3 indicates that $\mathcal{U}_\alpha^r$ is an under-approximation of the risk-sensitive safe set $\mathcal{S}_\alpha^r$. A purpose for Theorem 3 is to approximate the maximum in the definition of $\mathcal{S}_\alpha^r$ in terms of a summation since the latter may be more amenable to computation due to translation equivariance of CVAR.⁹ However, computing the function J_0^* , as defined in Theorem 3, is still difficult because the Conditional Value-at-Risk measure is *time-inconsistent* [4]. Intuitively, time-inconsistency implies that a policy that is optimal when viewed at time zero is not necessarily optimal when viewed at a later time [69]. Thus, a standard dynamic programming value iteration algorithm, such as [10, Sec. 1.3], cannot be used to compute J_0^* . In the next section, taking inspiration from [25] and [79], we prove that a value iteration algorithm on the augmented state space $\mathbb{X} := X \times [0, 1]$ computes J_0^* for a class of linear systems.

An Algorithm to Solve the CVAR-MDP Problem for a Class of Linear Systems

We consider a value iteration algorithm on $\mathbb{X} := X \times [0, 1]$ that induces dynamics on the risk-sensitivity level. The algorithm involves optimizing over a set of probability densities (i.e., weights) to determine the risk-sensitivity level at a given time, so a desired risk-sensitivity level is attained over the entire (finite) time horizon. We specify this set of probability densities next and use α_t to denote the risk-sensitivity level at time t .

⁹Translation equivariance of CVAR is the following property: $\text{CVAR}_\alpha(Y + a) = \text{CVAR}_\alpha(Y) + a$ if $a \in \mathbb{R}$. Please refer to Sec. 3.2 to review important properties of CVAR. Estimating risk-sensitive safe sets without the reduction to the CVAR-MDP problem is a future research direction, which will be discussed in Chapter 4. However, it should be noted that solving the CVAR-MDP problem (which we do in Sec. 3.3 for a class of linear systems) is an interesting research contribution itself.

Definition 5 (Risk Envelope). *Fix t , $x_t = \mathbf{x} \in X$, $u_t = \mathbf{u} \in U$, and $\alpha_t = \alpha \in [0, 1]$. The risk envelope for time t is defined as follows:*

$$\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) := \left\{ Z : \Omega \rightarrow \mathbb{R} \left| \begin{array}{l} \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t))\text{-measurable, } \int_{\Omega} Z d\mathbb{P} = 1, \\ 0 \leq \alpha Z(\omega) \leq 1 \text{ for a.e. } \omega \in \Omega \end{array} \right. \right\}, \quad (3.13)$$

where $\int_{\Omega} Z d\mathbb{P} := \int_{\Omega} Z(\omega) \mathbb{P}(d\omega)$ is the expectation of Z with respect to the probability measure $\mathbb{P}$, $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined, and $\sigma(f_t(\mathbf{x}, \mathbf{u}, d_t))$ is a sub- σ -algebra of $\mathcal{F}$.

The following algorithm, which was originally proposed by [25], defines the value function at time t in terms of a stage cost c_t and a worst-case weighted expected value function at time $t + 1$. Here the notion of worst case is specified by optimizing over the risk envelope for time t .

Algorithm 1 (CVAR Value Iteration). *Define the functions $J_{T-1}, \dots, J_1, J_0$ recursively as follows: for all $(\mathbf{x}, \alpha) \in \mathbb{X}$ and $t = T - 1, \dots, 1, 0$,*

$$J_t(\mathbf{x}, \alpha) := \inf_{\mathbf{u} \in U} \left\{ c_t(\mathbf{x}, \mathbf{u}) + \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} \int_{\Omega} Z(\omega) \cdot J_{t+1}(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega) \right\}, \quad (3.14)$$

with the terminal condition $J_T(\mathbf{x}, \alpha) := c_T(\mathbf{x})$.

In this section, we will prove that J_0^* , as defined by (3.8b), is equivalent to J_0 , as defined by Algorithm 1, by specifying additional assumptions, including linear time-varying (LTV) dynamics.

Assumption 1 (Conditions on Control System). *The following conditions hold:*

1. $x_{t+1} = f_t(x_t, u_t, d_t) := A_t x_t + B_t u_t + E_t d_t$ for $t = 0, 1, \dots, T - 1$. The matrices $A_t \in \mathbb{R}^{n \times n}$, $B_t \in \mathbb{R}^{n \times m}$, and $E_t \in \mathbb{R}^{n \times q}$ are given for each t . The initial condition $x_0 = \mathbf{x}$ is deterministic.
2. The stage cost functions $c_t : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ for $t = 0, 1, \dots, T - 1$ and the terminal cost function $c_T : \mathbb{R}^n \rightarrow \mathbb{R}$ are bounded, convex, and continuous.
3. The action space $U \subset \mathbb{R}^m$ is non-empty, compact, and convex.

Many practical control systems can be modeled by linear time-varying dynamics with compact convex action spaces and convex continuous stage (or terminal) costs. A concrete example is the thermostatically controlled load system model that we have presented in Sec. 3.2. Assumption 1 ensures that a minimax equality holds, and this minimax equality (which was not shown by [25]) is required to prove the validity of Algorithm 1.

The next definition specifies the CVAR-optimal cost-to-go at time t in terms of the set of admissible control policies Π_t (3.5).

Definition 6 (CVAR-Optimal Cost-to-Go). *The CVAR-optimal cost-to-go function $J_t^* : \mathbb{X} \rightarrow \mathbb{R}$ at $t = 0, 1, \dots, T-1$ is defined as follows:*

$$J_t^*(\mathbf{x}, \alpha) := \inf_{\pi_t \in \Pi_t} \text{CVAR}_{\alpha}^{\pi_t}(C_{t:T} | x_t = \mathbf{x}), \quad (3.15a)$$

where $C_{t:T}$ is the random cumulative cost of the state trajectory for time t ,

$$C_{t:T} := c_T(x_T) + \sum_{k=t}^{T-1} c_k(x_k, \mu_k(h_{t:k})), \quad (3.15b)$$

such that the state trajectory evolves under $\pi_t := (\mu_t, \mu_{t+1}, \dots, \mu_{T-1}) \in \Pi_t$ via the LTV dynamics initialized at $x_t = \mathbf{x}$. Moreover, we define the optimal terminal cost $J_T^*(\cdot, \alpha) := c_T$ for all $\alpha \in [0, 1]$.

Since we have specified Assumption 1 and the CVAR-optimal cost-to-go function J_t^* , we are ready to state the main result of this subsection.

Theorem 4 (Validity of CVAR Value Iteration). *The value function J_t , as defined recursively by Algorithm 1, is equivalent to the CVAR-optimal cost-to-go function J_t^* , as defined by (3.15), for $t = 0, 1, \dots, T$.*

Theorem 4 is significant because it allows us to solve the CVAR-MDP problem (3.15) numerically and hence estimate the risk-sensitive safe set under-approximation $\mathcal{U}_{\alpha}^r$ (3.8). Our strategy to prove Theorem 4 leverages a representation theorem for the Conditional Value-at-Risk measure from [79] that is amenable to a value iteration approach. After showing how the representation theorem simplifies in the setting of a Markov Decision Process (Lemma 3), we will prove a critical minimax equality (Lemma 4) and then use these two lemmas to prove Theorem 4.

Lemma 3 (Transfer Thm. 6 (iii) in [79] to MDP). *Fix t , $x_t = \mathbf{x} \in X$, $u_t = \mathbf{u} \in U$, and $\alpha_t = \alpha \in [0, 1]$. For any $\pi \in \Pi_{t+1}$, the following equality holds:*

$$\text{CVAR}_{\alpha}^{\pi}(C_{t+1:T} | x_t = \mathbf{x}, u_t = \mathbf{u}) = \sup_{Z \in \mathcal{R}_t^{\alpha}(\mathbf{x}, \mathbf{u})} \int_{\Omega} Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^{\pi}(C_{t+1:T} | x_{t+1}) \mathbb{P}(d\omega), \quad (3.16)$$

where $C_{t+1:T}$ is conditioned on $x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega))$ in the integral above, $C_{t+1:T}$ is the random cumulative cost of the state trajectory for time $t+1$ (3.15b), $\mathcal{R}_t^{\alpha}(\mathbf{x}, \mathbf{u})$ is the risk envelope for time t (3.13), and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined.

Proof. Let $\mathcal{F}_t := \sigma(h_{0:t})$, where $h_{0:t} := (x_0, u_0, \dots, x_{t-1}, u_{t-1}, x_t)$ is the random history from time 0 to time t . Fix $u_t \in U$ and $\pi \in \Pi_{t+1}$. Apply [79, Thm. 6 (iii)] to the random cumulative cost $C_{t+1:T}$ of the state trajectory conditioned on $(u_t, \pi) \in U \times \Pi_{t+1}$ to obtain

$$\text{CVAR}_{\alpha_t}^{\pi}(C_{t+1:T} | \mathcal{F}_t, u_t) = \text{ess sup}_Z \mathbb{E}(Z \cdot \text{CVAR}_{\alpha_t Z}^{\pi}(C_{t+1:T} | \mathcal{F}_{t+1}) | \mathcal{F}_t, u_t), \quad (3.17)$$

where the essential supremum is taken over the set of $\mathcal{F}_{t+1}$ -measurable random variables Z that satisfy $0 \leq \alpha_t Z \leq 1$ and $\mathbb{E}(Z|\mathcal{F}_t, u_t) = 1$ almost everywhere. We condition on $u_t \in U$ and $\pi \in \Pi_{t+1}$ above to fix the tree of possible outcomes, so [79, Thm. 6] applies, although the original theorem has no explicit notion of control. Further, since $(x_t, u_t, \alpha_t) = (\mathbf{x}, \mathbf{u}, \alpha)$ is given, $C_{t+1:T}$ is initialized at time $t + 1$, and the system is Markov, we can simplify (3.17) by removing the dependency on the history prior to time t as follows:

$$\text{CVAR}_\alpha^\pi(C_{t+1:T}|x_t = \mathbf{x}, u_t = \mathbf{u}) = \sup_Z \mathbb{E}(Z(x_{t+1})\text{CVAR}_{\alpha Z(x_{t+1})}^\pi(C_{t+1:T}|x_{t+1})|x_t = \mathbf{x}, u_t = \mathbf{u}),$$

where the supremum is taken over the set of Borel measurable functions $Z : X \rightarrow \mathbb{R}$ with $0 \leq \alpha Z(x_{t+1}) \leq 1$ for almost every x_{t+1} and $\mathbb{E}(Z(x_{t+1})|x_t = \mathbf{x}, u_t = \mathbf{u}) = 1$. The expectations above are taken with respect to the probability distribution of x_{t+1} conditioned on $(x_t, u_t) = (\mathbf{x}, \mathbf{u})$. Note that x_{t+1} is a function from Ω to X through d_t , and the disturbances are independent with a common distribution defined by the probability measure $\mathbb{P}$. Thus, for any Borel measurable function $g : X \rightarrow \mathbb{R}$, we have $\mathbb{E}(g(x_{t+1})|x_t = \mathbf{x}, u_t = \mathbf{u}) = \mathbb{E}(g(f_t(\mathbf{x}, \mathbf{u}, d_t))) := \int_\Omega g(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)))\mathbb{P}(d\omega) = \int_\Omega \hat{g}(\omega)\mathbb{P}(d\omega)$, where $\hat{g} : \Omega \rightarrow \mathbb{R}$ is $\sigma(f_t(\mathbf{x}, \mathbf{u}, d_t))$ -measurable; see [5, Thm. 6.4.2 (c), p. 251] and [46, Eqn. 3.4.2, p. 31]. So, the desired result holds. $\square$

Having shown Lemma 3, we proceed to Lemma 4. In Lemma 4, we will show a critical minimax equality that uses Sion's Minimax Theorem and two intermediary results (Lemmas 6 and 7, which are provided in Appendix A at the end of this thesis). The intermediary results specify properties of certain functions using Assumption 1 as well as properties of the risk envelope to allow the application of Sion's Minimax Theorem. Next, we state and prove Lemma 4.

Lemma 4 (Minimax Equality). *Fix t , $x_t = \mathbf{x} \in X$, $u_t = \mathbf{u} \in U$, and $\alpha_t = \alpha \in [0, 1]$. Define $H : \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \times \Pi_{t+1} \rightarrow \mathbb{R}$ as follows:*

$$H(Z, \pi) := \int_\Omega Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^\pi(C_{t+1:T}|x_{t+1})\mathbb{P}(d\omega), \quad (3.18)$$

where $C_{t+1:T}$ is conditioned on $x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)) := A_t\mathbf{x} + B_t\mathbf{u} + E_td_t(\omega)$ inside the integral above, and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined. Then, the following minimax equality holds:

$$\inf_{\pi \in \Pi_{t+1}} \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} H(Z, \pi) = \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} \inf_{\pi \in \Pi_{t+1}} H(Z, \pi). \quad (3.19)$$

Proof. We use Sion's Minimax Theorem [83, Thm. 1.1]. $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is a non-empty, convex, and compact set in $L^2 := L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$ with the weak topology by Lemma 7. $H(\cdot, \pi)$ is concave on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ for any $\pi \in \Pi_{t+1}$ by [79, Thm. 12] applied to a random cost and by linearity of expectation. For any $\pi \in \Pi_{t+1}$, $H(\cdot, \pi)$ is upper semi-continuous in the relative weak topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$ by [7, Prop. 2.10, p. 72] since $H(\cdot, \pi)$ is real-valued, concave, and upper semi-continuous in the relative norm topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$

(Lemma 6, Lipschitz $\implies$ upper semi-continuous). Moreover, since $U \subseteq \mathbb{R}^m$ is non-empty, bounded, and convex, Π_{t+1} is a non-empty convex set in $\{(\mu_{t+1}, \dots, \mu_{T-1}) \mid \mu_k : H_{t+1:k} \rightarrow \mathbb{R}^m, \|\mu_k\|_u < \infty, \text{ Borel measurable}\}$ with the norm topology. $H(Z, \cdot)$ is convex and continuous in the norm topology on Π_{t+1} for any $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ by Lemma 6. So, the conditions of Sion's Minimax Theorem hold, which completes the proof. $\square$

Equipped with the conclusions of Lemmas 3 and 4, we are ready to prove Theorem 4. In particular, we show that the value function J_t , as defined recursively by Algorithm 1, is equivalent to the CVAR-optimal cost-to-go function J_t^* , as defined by (3.15), for each t . This equivalence means that we can solve the CVAR-MDP problem (3.15) numerically using Algorithm 1 and hence estimate the risk-sensitive safe set under-approximation $\mathcal{U}_\alpha^r$ (3.8). Next, we prove Theorem 4 by induction.

Proof. $J_T^* = J_T$ by definition. Assume $J_{t+1}^* = J_{t+1}$ for some t . Take $(\mathbf{x}, \alpha) \in \mathbb{X}$. Then, the following equations hold:

$$\begin{aligned} J_t^*(\mathbf{x}, \alpha) &\stackrel{(i)}{=} \inf_{\mathbf{u} \in U} \left\{ c_t(\mathbf{x}, \mathbf{u}) + \inf_{\pi \in \Pi_{t+1}} \text{CVAR}_\alpha^\pi(C_{t+1:T} \mid x_t = \mathbf{x}, u_t = \mathbf{u}) \right\} \\ &\stackrel{(ii)}{=} \inf_{\mathbf{u} \in U} \left\{ c_t(\mathbf{x}, \mathbf{u}) + \inf_{\pi \in \Pi_{t+1}} \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} H_t^\alpha(Z, \pi, \mathbf{x}, \mathbf{u}) \right\} \\ &\stackrel{(iii)}{=} \inf_{\mathbf{u} \in U} \left\{ c_t(\mathbf{x}, \mathbf{u}) + \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} \inf_{\pi \in \Pi_{t+1}} H_t^\alpha(Z, \pi, \mathbf{x}, \mathbf{u}) \right\}, \end{aligned}$$

where

$$H_t^\alpha(Z, \pi, \mathbf{x}, \mathbf{u}) := \int_{\Omega} Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^\pi(C_{t+1:T} \mid x_{t+1}) \mathbb{P}(d\omega),$$

such that $C_{t+1:T}$ is conditioned on $x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)) := A_t \mathbf{x} + B_t \mathbf{u} + E_t d_t(\omega)$ in the integral above, and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined. We justify each equality as follows: (i) the equivalence between minimizing over $\mu_t : H_{t:t} \rightarrow U$ and $\mathbf{u} \in U$ when $x_t = \mathbf{x}$ is given, and CVAR is translation equivariant; (ii) Lemma 3; and (iii) Lemma 4. Moreover, the following equations hold:

$$\begin{aligned} \inf_{\pi \in \Pi_{t+1}} H_t^\alpha(Z, \pi, \mathbf{x}, \mathbf{u}) &\stackrel{(iv)}{=} \int_{\Omega} Z(\omega) \cdot \inf_{\pi \in \Pi_{t+1}} \text{CVAR}_{\alpha Z(\omega)}^\pi(C_{t+1:T} \mid x_{t+1}) \mathbb{P}(d\omega) \\ &\stackrel{(v)}{=} \int_{\Omega} Z(\omega) \cdot J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega) \\ &\stackrel{(vi)}{=} \int_{\Omega} Z(\omega) \cdot J_{t+1}(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega), \end{aligned}$$

where we justify each step above: (iv) the interchangeability assertion from the proof of [79, Thm. 18, p. 162];¹⁰ (v) the definition of J_{t+1}^* (3.15), and $C_{t+1:T}$ is conditioned on $x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega))$ in the previous line; and (vi) the induction hypothesis $J_{t+1}^* = J_{t+1}$. The above steps together with (3.14) show that $J_t^* = J_t$, which completes the induction. $\square$

We have shown that J_t^* , as defined by (3.15), is equal to J_t , as defined by Algorithm 1, for $t = 0, 1, \dots, T-1$ under the conditions specified by Assumption 1. In the next subsection, we prove the existence of tractable policies that attain J_t^* , if we pre-commit (restrict ourselves) to the dynamics of the risk-sensitivity level induced by Algorithm 1.

Synthesis of Pre-Commitment Control Policies for the CVAR-MDP Problem

Our approach to the synthesis of control policies for the CVAR-MDP problem (3.15) involves restricting the optimization space of history-dependent policies (3.5) to policies that commit to specific risk-sensitivity level dynamics *a priori*. Such policies are called *pre-commitment* policies. This restriction to pre-commitment policies is well-motivated for two reasons. First, due to time-inconsistency, a Conditional Value-at-Risk cost function is not amenable to dynamic programming unless the entire history is recorded [79]. Second, computing history-dependent policies is intractable generally since substantial memory and computation time are required. However, we can overcome these two challenges by restricting the optimization space to policies that pre-commit to risk-sensitivity level dynamics that summarize the history.

The purpose of this section is to synthesize tractable policies that attain the CVAR-optimal cost-to-go function J_t^* (3.15) under mild restrictions on the optimization space of history-dependent policies Π_t (3.5). We define a control law at time $k \geq t$ on the augmented state space $\mathbb{X} := X \times [0, 1]$ that is a function of the current state x_k and the current risk-sensitivity level α_k . Accordingly, this law is called a $\mathbb{X}$ -Markov control law. The current state $x_k \in X$ satisfies the linear time-varying dynamics specified by Assumption 1. The current risk-sensitivity level $\alpha_k \in [0, 1]$ satisfies particular dynamics induced by Algorithm 1. Under these dynamics, α_k is a one-dimensional parameter that summarizes the higher-dimensional prior history $(x_t, u_t, x_{t+1}, u_{t+1}, \dots, x_{k-1}, u_{k-1}) \in (X \times U)^{k-t}$.

This section is structured as follows. First, we specify a regularity condition on the risk envelope to facilitate policy synthesis (Assumption 2). Second, we define $\mathbb{X}$ -Markov control laws and $\mathbb{X}$ -Markov policies (Definition 7). We prove the existence of $\mathbb{X}$ -Markov control laws by invoking Assumptions 1 and 2. Third, we define the dynamics of the risk-sensitivity level induced by Algorithm 1 (Definition 8). Finally, in Theorem 5, we show that $\mathbb{X}$ -Markov policies attain the CVAR-optimal cost-to-go function (3.15), if we pre-commit to the risk-sensitivity level dynamics induced by Algorithm 1.

¹⁰Showing the interchangeability assertion from the proof of [79, Thm. 18, p. 162] explicitly (which has not been done to our knowledge and is out of the scope of this thesis) requires non-trivial arguments from functional analysis, measure theory, and topology, which could be part of a future doctoral dissertation.

Regularity Condition on Risk Envelope

Here we specify and justify a regularity condition on the risk envelope defined by (3.13).

Assumption 2 (Regularity Condition on Risk Envelope). *The set-valued mapping $(\mathbf{x}, \alpha, \mathbf{u}) \mapsto \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is lower semi-continuous on $\mathbb{X} \times U$.*¹¹

Assumption 2 is useful in general since it guarantees the existence of policies that optimize certain performance criteria. For example, in the classic setting of *games against nature*, where stochastic control systems are affected by adversarial players with unknown distributions, the set of admissible actions of the adversary (player 2) is assumed to be lower semi-continuous in the state and action of the controller (player 1) [37, Assumption 3.1 (g), p. 1632]. Also, Assumption 2 resembles conditions in the risk-averse dynamic programming literature. Ruszczyński [88] assumes that the stochastic transition kernel is continuous in the control and the risk envelope is lower semi-continuous in the probability measure, implying that the risk envelope is lower semi-continuous in the control by composition [88, Thm. 2 (i) (ii)]. The conditions in [88] are required for a time-consistent problem, whereas the condition in Assumption 2 is required for a time-inconsistent problem.

Assumption 2 is a *measurable selection condition* for time-inconsistent CVAR-MDP problems that guarantees the existence of a $\mathbb{X}$ -Markov control law.¹² Next, we define a $\mathbb{X}$ -Markov control law in terms of an argument that minimizes an objective function over a compact set. To ensure that a minimum argument exists, the objective function must be lower semi-continuous. Assumption 2 specifies a sufficient condition to guarantee that the objective function is indeed lower semi-continuous.

Existence of $\mathbb{X}$ -Markov Policies

Recall that $\mathbb{X} := X \times [0, 1]$ is the state space augmented by the space of risk-sensitivity levels. Here we define $\mathbb{X}$ -Markov control laws as Borel measurable mappings from $\mathbb{X}$ to U and $\mathbb{X}$ -Markov policies as time-based sequences of such mappings. Then, we prove the existence of $\mathbb{X}$ -Markov control laws using the conditions specified by Assumptions 1 and 2.

Definition 7 ($\mathbb{X}$ -Markov Control Law, Policy). *Fix t . A $\mathbb{X}$ -Markov control law at time t is a Borel measurable function $\mu_t^* : \mathbb{X} \rightarrow U$ that satisfies the following: for all $(\mathbf{x}, \alpha) \in \mathbb{X}$,*

$$\mu_t^*(\mathbf{x}, \alpha) \in \arg \min_{\mathbf{u} \in U} \left\{ c_t(\mathbf{x}, \mathbf{u}) + \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}) \right\}, \quad (3.20)$$

¹¹The condition specified in Assumption 2 is equivalent to the following statement. If $(\mathbf{x}_n, \alpha_n, \mathbf{u}_n) \rightarrow (\mathbf{x}, \alpha, \mathbf{u})$ in $\mathbb{X} \times U$ and $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$, then there exist $Z_n \in \mathcal{R}_t^{\alpha_n}(\mathbf{x}_n, \mathbf{u}_n)$ s.t. $Z_n \rightarrow Z$ [46, Prop. D.2, p. 182].

¹²Measurable selection conditions for standard MDPs, where expected costs are assessed, can be found in [46, Sec. 3.3]. An important direction for future work is to identify practical examples that satisfy Assumption 2.

such that $(x_t, \alpha_t) = (\mathbf{x}, \alpha)$ is the state and risk-sensitivity level at time t , and $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is defined as follows:

$$G_t^\alpha(Z, \mathbf{x}, \mathbf{u}) := \int_{\Omega} Z(\omega) \cdot J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega), \quad (3.21)$$

where $f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)) := A_t \mathbf{x} + B_t \mathbf{u} + E_t d_t(\omega)$, $x_t = \mathbf{x}$ is the state at time t , $u_t = \mathbf{u}$ is the control at time t , and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined. A sequence of $\mathbb{X}$ -Markov control laws $\pi_t^* := (\mu_t^*, \mu_{t+1}^*, \dots, \mu_{T-1}^*)$ is a $\mathbb{X}$ -Markov policy for time t .

It will become clear that the arguments in Lemmas 5, 8, and 9 together show the existence of $\mathbb{X}$ -Markov control laws by invoking Assumptions 1 and 2. Lemma 5 specifies that the objective function in (3.20) is lower semi-continuous in $\mathbf{u}$ to ensure the existence of a $\mathbb{X}$ -Markov control law, if $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous in $(Z, \mathbf{u})$. Lemma 8 specifies that $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous in $(Z, \mathbf{u})$, if J_{t+1}^* is lower semi-continuous (Appendix A). Finally, Lemma 9 specifies that J_t^* is lower semi-continuous for each t (Appendix A). We provide the statements and proofs of Lemmas 8 and 9 at the end of this thesis in Appendix A, while we provide the statement and proof of Lemma 5 next.

Lemma 5 (Existence of $\mathbb{X}$ -Markov Control Laws). *Fix t , and define $v_t : \mathbb{X} \times U \rightarrow \mathbb{R}$ as follows:*

$$v_t(\mathbf{x}, \alpha, \mathbf{u}) := c_t(\mathbf{x}, \mathbf{u}) + \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}), \quad (3.22)$$

where $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is given by (3.13), and $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is given by (3.21). Then, there exists a Borel measurable function $\mu_t^* : \mathbb{X} \rightarrow U$ that satisfies the following statement: for all $(\mathbf{x}, \alpha) \in \mathbb{X}$,

$$v_t(\mathbf{x}, \alpha, \mu_t^*(\mathbf{x}, \alpha)) = \inf_{\mathbf{u} \in U} v_t(\mathbf{x}, \alpha, \mathbf{u}) = \min_{\mathbf{u} \in U} v_t(\mathbf{x}, \alpha, \mathbf{u}). \quad (3.23)$$

Proof. Since U is compact, it suffices to show that $v_t(\mathbf{x}, \alpha, \cdot)$ is lower semi-continuous on U for every $(\mathbf{x}, \alpha) \in \mathbb{X}$ by [46, Prop. D.5 (a)]. Since the sum of lower semi-continuous functions is lower semi-continuous, it suffices to show that $\mathbf{u} \mapsto \psi_{\mathbf{x}}^\alpha(\mathbf{u}) := \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower

semi-continuous on U for every $(\mathbf{x}, \alpha) \in \mathbb{X}$. Since $\mathbf{u} \mapsto \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is a lower semi-continuous set-valued mapping by Assumption 2, for any $\{\mathbf{u}^{(j)}\} \subseteq U$ converging to a point $\mathbf{u} \in U$ and for any $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$, there exists $Z^{(j)} \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}^{(j)})$ such that $Z^{(j)} \rightarrow Z$. If $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous in $(Z, \mathbf{u})$, then one can show

$$\liminf_{j \rightarrow \infty} \psi_{\mathbf{x}}^\alpha(\mathbf{u}^{(j)}) \geq \liminf_{j \rightarrow \infty} G_t^\alpha(Z^{(j)}, \mathbf{x}, \mathbf{u}^{(j)}) \geq G_t^\alpha(Z, \mathbf{x}, \mathbf{u}),$$

by the definition of supremum and lower semi-continuity, respectively. Since the above inequalities hold for any $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$, and since the supremum is the least upper bound,

$$\liminf_{j \rightarrow \infty} \psi_{\mathbf{x}}^\alpha(\mathbf{u}^{(j)}) \geq \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}) =: \psi_{\mathbf{x}}^\alpha(\mathbf{u}),$$

showing that $\psi_{\mathbf{x}}^\alpha$ is lower semi-continuous on U , which is the desired result. (See [37, Lemma 3.2 (a)] for the original proof.) So, it suffices to show that $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous in $(Z, \mathbf{u})$, which is shown in Lemma 8 (Appendix A). $\square$

Next, we specify the dynamics of the risk-sensitivity level so that $\mathbb{X}$ -Markov policies under these dynamics are history-dependent.

Risk-Sensitivity Level Dynamics

Here we define the risk-sensitivity level dynamics in terms of a member of the risk envelope that maximizes a weighted CVAR-optimal cost-to-go in expectation.

Definition 8 (Risk-Sensitivity Level Dynamics). *Fix t , $x_t = \mathbf{x} \in X$, $u_t = \mathbf{u} \in U$, and $\alpha_t = \alpha \in [0, 1]$. Let $Z_t^* \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ satisfy the following statement:*

$$Z_t^* \in \arg \max_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}),$$

where $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is defined by (3.21). Then, the risk-sensitivity level at time $t + 1$ is given by $\alpha_{t+1} = \alpha \cdot Z_t^*$.

Please note the following remarks concerning the risk-sensitivity level dynamics defined above.

Remark 6. Lemma 7 and Lemma 10 ensure that $Z_t^* \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is well-defined since the supremum of an upper semi-continuous map on a compact topological space is attained [5, Thm. A6.3]. Lemma 7 specifies that the risk envelope $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is compact in $L^2 := L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$ endowed with the weak topology (Appendix A). Lemma 10 specifies that $G_t^\alpha(\cdot, \mathbf{x}, \mathbf{u})$ is upper semi-continuous in the relative weak topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$ (Appendix A).

Remark 7. The risk-sensitivity level at time $t + 1$, $\alpha_{t+1} = \alpha \cdot Z_t^*$, is a $\sigma(f_t(\mathbf{x}, \mathbf{u}, d_t))$ -measurable random variable since Z_t^* is. If $\omega \in \Omega$ is fixed (which implies that $d_t(\omega)$ occurs at time t), then the risk-sensitivity level that occurs at time $t + 1$ is given by $\alpha_{t+1}(\omega) = \alpha \cdot Z_t^*(\omega)$.

Remark 8. $\mathbb{X}$ -Markov policies (Definition 7) under the dynamics of the risk-sensitivity level (Definition 8) are history-dependent policies. Specifically, a $\mathbb{X}$ -Markov policy for time t is a function of the augmented state trajectory $(x_t, \alpha_t, x_{t+1}, \alpha_{t+1}, \dots, x_T, \alpha_T)$ by Definition 7, where α_k is a function of $(\alpha_t, x_t, u_t, \dots, x_{k-1}, u_{k-1})$ for $k = t + 1, t + 2, \dots, T$ by Definition 8 and [5, Thm. 6.4.2 (c), p. 251].

Optimality of Pre-Commitment $\mathbb{X}$ -Markov Policies

The CVAR-optimal cost-to-go J_t^* (3.15) is an infimum over a set of history-dependent policies Π_t (3.5). The final theorem of this chapter indicates that $\mathbb{X}$ -Markov policies (Definition 7) attain J_t^* , if we pre-commit (restrict ourselves) to the risk-sensitivity level dynamics given by Definition 8. This is a powerful practical result because history-dependent policies are generally intractable due to extensive memory requirements, but the following theorem enables tractable computations.

Theorem 5 (Optimal Pre-Commitment Policies). *Fix t . Let $\pi_t^* = (\mu_t^*, \mu_{t+1}^*, \dots, \mu_{T-1}^*)$ be a $\mathbb{X}$ -Markov policy for time t , where the risk-sensitivity level dynamics are given by Definition 8. Then, $\pi_t^* \in \Pi_t$ is an optimal pre-commitment policy, i.e., the following equality holds: for all $(\mathbf{x}, \alpha) \in \mathbb{X}$,*

$$J_t^*(\mathbf{x}, \alpha) = \text{CVAR}_{\alpha}^{\pi_t^*} (C_{t:T} | x_t = \mathbf{x}), \quad (3.24)$$

where

$$C_{t:T} = c_T(x_T) + \sum_{k=t}^{T-1} c_k(x_k, \mu_k^*(x_k, \alpha_k)),$$

and $(x_t, \alpha_t, \dots, x_T, \alpha_T)$ satisfies the linear time-varying dynamics (Assumption 1) and the risk-sensitivity level dynamics (Definition 8) under π_t^* initialized at $(x_t, \alpha_t) = (\mathbf{x}, \alpha)$.

Proof. We proceed by induction. Since $J_T^*(x_T, \alpha_T) = c_T(x_T)$ does not depend on the control when $(x_T, \alpha_T) = (\mathbf{x}, \alpha)$ is given, the base case is $t = T - 1$. In this case, $\pi_{T-1}^* = \mu_{T-1}^*$, where $u_{T-1} = \mu_{T-1}^*(\mathbf{x}, \alpha)$. Then, the following equations hold:

$$\begin{aligned} \text{CVAR}_{\alpha}^{\mu_{T-1}^*} (C_{T-1:T} | x_{T-1} = \mathbf{x}) - c_{T-1}(\mathbf{x}, \mu_{T-1}^*(\mathbf{x}, \alpha)) \\ \stackrel{(i)}{=} \text{CVAR}_{\alpha} (c_T(x_T) | x_{T-1} = \mathbf{x}, u_{T-1} = \mu_{T-1}^*(\mathbf{x}, \alpha)) \\ \stackrel{(ii)}{=} \sup_{Z \in \mathcal{R}_{T-1}^{\alpha}(\mathbf{x}, \mu_{T-1}^*(\mathbf{x}, \alpha))} \int_{\Omega} Z(\omega) \cdot c_T(x_T(\omega)) \mathbb{P}(d\omega), \end{aligned}$$

where $x_T(\omega) = f_{T-1}(\mathbf{x}, \mu_{T-1}^*(\mathbf{x}, \alpha), d_{T-1}(\omega))$, and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_{T-1} is defined. We justify each equality as follows: (i) CVAR is translation equivariant; and (ii) Lemma 3, and $\text{CVAR}_{\beta}(c_T(x_T) | x_T) = c_T(x_T)$ for any $\beta \in [0, 1]$ since c_T is bounded. On the right side of equality (ii) above, $c_T(x_T) = J_T^*(x_T, \alpha Z)$. Next, use Definition 7 to obtain

$$\text{CVAR}_{\alpha}^{\mu_{T-1}^*} (C_{T-1:T} | x_{T-1} = \mathbf{x}) = \inf_{\mathbf{u} \in U} \left\{ c_{T-1}(\mathbf{x}, \mathbf{u}) + \sup_{Z \in \mathcal{R}_{T-1}^{\alpha}(\mathbf{x}, \mathbf{u})} G_{T-1}^{\alpha}(Z, \mathbf{x}, \mathbf{u}) \right\} = J_{T-1}^*(\mathbf{x}, \alpha),$$

which shows the base case. Now, for some t , consider a $\mathbb{X}$ -Markov policy π_{t+1}^* , and assume the following: for all $(\bar{\mathbf{x}}, \bar{\alpha}) \in \mathbb{X}$,

$$J_{t+1}^*(\bar{\mathbf{x}}, \bar{\alpha}) = \text{CVAR}_{\bar{\alpha}}^{\pi_{t+1}^*} (C_{t+1:T} | x_{t+1} = \bar{\mathbf{x}}),$$

where $(x_{t+1}, \alpha_{t+1}, u_{t+1}, \dots, x_{T-1}, \alpha_{T-1}, u_{T-1}, x_T, \alpha_T)$ satisfies the linear time-varying dynamics (Assumption 1) and the risk-sensitivity level dynamics (Definition 8) under π_{t+1}^* with the initialization $(x_{t+1}, \alpha_{t+1}) = (\bar{\mathbf{x}}, \bar{\alpha})$. Consider $\pi_t^* = (\mu_t^*, \pi_{t+1}^*)$, where μ_t^* is a $\mathbb{X}$ -Markov control law. Then, for any $(x_t, \alpha_t) = (\mathbf{x}, \alpha) \in \mathbb{X}$, the following equations hold:

$$\begin{aligned} \text{CVAR}_\alpha^{(\mu_t^*, \pi_{t+1}^*)} (C_{t:T} | x_t = \mathbf{x}) &\stackrel{(i)}{=} \text{CVAR}_\alpha^{\pi_{t+1}^*} (c_t(x_t, u_t) + C_{t+1:T} | x_t = \mathbf{x}, u_t = \mu_t^*(\mathbf{x}, \alpha)) \\ &\stackrel{(ii)}{=} c_t(\mathbf{x}, \mu_t^*(\mathbf{x}, \alpha)) + \text{CVAR}_\alpha^{\pi_{t+1}^*} (C_{t+1:T} | x_t = \mathbf{x}, u_t = \mu_t^*(\mathbf{x}, \alpha)). \end{aligned}$$

We justify each equality as follows: (i) $\mu_t^*(\mathbf{x}, \alpha)$ is the control at time t since μ_t^* is a $\mathbb{X}$ -Markov control law at time t and $(x_t, \alpha_t) = (\mathbf{x}, \alpha)$; and (ii) CVAR is translation equivariant. Denote $\mathbf{u}_t^* := \mu_t^*(\mathbf{x}, \alpha)$ and use Lemma 3 to obtain

$$\text{CVAR}_\alpha^{\pi_{t+1}^*} (C_{t+1:T} | x_t = \mathbf{x}, u_t = \mathbf{u}_t^*) = \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}_t^*)} \int_{\Omega} Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^{\pi_{t+1}^*} (C_{t+1:T} | x_{t+1}) \mathbb{P}(d\omega),$$

where $C_{t+1:T}$ is conditioned on $x_{t+1} = f_t(\mathbf{x}, \mathbf{u}_t^*, d_t(\omega))$ in the integral above. Next, using the induction hypothesis and the definition of G_t^α given by (3.21), we have

$$\begin{aligned} \text{CVAR}_\alpha^{\pi_{t+1}^*} (C_{t+1:T} | x_t = \mathbf{x}, u_t = \mathbf{u}_t^*) &= \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}_t^*)} \int_{\Omega} Z(\omega) \cdot J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}_t^*, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega) \\ &= \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}_t^*)} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}_t^*). \end{aligned}$$

The risk-sensitivity level at time $t+1$ satisfies Definition 8 when using the policy $\pi_t^* = (\mu_t^*, \pi_{t+1}^*)$. Indeed, this risk-sensitivity level is given by $\alpha_{t+1} = \alpha \cdot Z_t^*$, where $Z_t^* \in \arg \max_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}_t^*)} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}_t^*)$ and $\mathbf{u}_t^* = \mu_t^*(\mathbf{x}, \alpha)$. Finally, combine the above equations to obtain the following statements:

$$\begin{aligned} \text{CVAR}_\alpha^{(\mu_t^*, \pi_{t+1}^*)} (C_{t:T} | x_t = \mathbf{x}) &= c_t(\mathbf{x}, \mu_t^*(\mathbf{x}, \alpha)) + \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mu_t^*(\mathbf{x}, \alpha))} G_t^\alpha(Z, \mathbf{x}, \mu_t^*(\mathbf{x}, \alpha)) \\ &\stackrel{(iii)}{=} \inf_{\mathbf{u} \in U} \left\{ c_t(\mathbf{x}, \mathbf{u}) + \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u}) \right\} \\ &\stackrel{(iv)}{=} J_t^*(\mathbf{x}, \alpha), \end{aligned} \tag{3.25}$$

where equality (iii) holds since μ_t^* is a $\mathbb{X}$ -Markov control law (Definition 7), and equality (iv) holds since $J_t = J_t^*$ (Theorem 4). Since the induction step has been shown, the proof is complete. $\square$

We have presented the theoretical contributions of this chapter. The next section provides a numerical example of risk-sensitive safety analysis applied to a stormwater system.

3.4 Numerical Example

We provide a numerical example of a two-tank gravity-driven stormwater system with an automated valve to illustrate risk-sensitive safety analysis (Fig. 3.10, Table 3.2).¹³ Consider the following non-linear discrete-time dynamics model: $x_{t+1} = x_t + F(x_t, u_t, d_t) \cdot \Delta\tau$ for $t = 0, 1, \dots, T-1$, where

$$\begin{aligned} F(x, u, d) &:= \left[\frac{d - q_{\text{valve}}(x, u)}{a_1}, \frac{d + q_{\text{valve}}(x, u) - q_{\text{drain}}(x)}{a_2} \right]^T \\ q_{\text{valve}}(x, u) &:= u \cdot \pi r_v^2 \cdot \text{sgn } h(x) \cdot \sqrt{2g|h(x)|} \\ h(x) &:= \max(x_1 - Z_1, 0) - \max(x_2 - Z_{1,\text{in}}, 0) \\ q_{\text{drain}}(x) &:= \begin{cases} C_d \pi r_d^2 \sqrt{2g(x_2 - Z_2)} & \text{if } x_2 \geq Z_2 \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and $x_t = [x_{1t}, x_{2t}]^T \in \mathbb{R}^2$ is the state vector, $x_{it} \in \mathbb{R}$ is the water elevation of tank i , $u_t \in [0, 1]$ is the valve setting (closed to open), $d_t \in \mathbb{R}$ is the random surface runoff (disturbance), and $\Delta\tau$ is the duration of $[t, t+1)$. The constraint set $K := [0, K_1] \times [0, K_2]$ specifies the maximum water elevation of the two tanks. The stage cost $g_K(x_t) := \max(x_{1t} - K_1, x_{2t} - K_2, 0)$ is the maximum overflow elevation when the system occupies state x_t .

We have identified a finite probability distribution for d_t using the first three empirical moments of time-averaged surface runoff samples (Table 3.3). We have obtained the surface runoff samples by simulating a design storm in PCSWMM (Computational Hydraulics International), which extends USEPA's Stormwater Management Model [85]. We have used the two-year Type II 24-hour design storm for Lenexa, Kansas to generate 100,000 surface runoff samples, and these samples prior to time-averaging are shown in Fig. 3.3.

We have estimated $\{\mathcal{U}_\alpha^r\}$ by approximating J_0^* on a grid of states and risk-sensitivity levels via Algorithm 1, and by utilizing the risk-sensitivity level interpolation approach proposed by [25], multi-linear interpolation of the state space, and uniform discretization of the action space. We have chosen the values of β and γ (Theorem 3) empirically, where the magnitude of γ is constrained by the limitations of numerical solvers to manage differently scaled constraints. The computation time of the value functions $J_0, J_1, \dots, J_{T-1}$ and an optimal pre-commitment policy $\pi_0^* = (\mu_0^*, \mu_1^*, \dots, \mu_{T-1}^*)$ was about 230 hours on a three-dimensional grid of 50,490 nodes over $T = 48$ time points when using a 4-core machine.¹⁴ We have executed Algorithm 1 serially on the state space grid, however at a given time point t , the computations at each state are independent and can be run in parallel to reduce computation time. We have approximated $\{\mathcal{S}_\alpha^r\}$ by performing 100,000 Monte Carlo simulations

¹³Our code is available at <https://github.com/risk-sensitive-reachability/IEEE-TAC-2019>.

¹⁴The three-dimensional grid consists of 15 risk-sensitivity levels, 51 values of x_1 , and 66 values of x_2 for a total of 50,490 nodes. To compute the value functions and an optimal pre-commitment policy, we have executed Algorithm 1 in a cluster computing session that was allocated 4 CPU cores running at 2.8 GHz. We have used the Tufts Linux Research Cluster (Medford, MA) running MATLAB (The Mathworks, Inc.) with MOSEK [100] and CVX [40] [41].

Table 3.2: Two-Tank Stormwater System Parameters

Symbol	Description	Value
a_1	surface area of tank 1	28292 ft ²
a_2	surface area of tank 2	25965 ft ²
C_d	discharge coefficient	0.61 (no units)
g	acceleration due to gravity	32.2 $\frac{\text{ft}}{\text{s}^2}$
K_1	maximum water elevation of tank 1	3.5 ft
K_2	maximum water elevation of tank 2	5 ft
π	circle circumference-to-diameter ratio	≈ 3.14
r_d	radius of drain	2/3 ft
r_v	radius of valve	1/3 ft
$\Delta\tau$	duration of $[t, t + 1)$	5 min
T	length of discrete time horizon	48 (= 4 hr)
U	action space	$[0, 1]$ (no units)
X	state space	$[0, 5]$ ft $\times$ $[0, 6.5]$ ft
Z_1	invert elevation of pipe from base of tank 1	1 ft
$Z_{1,\text{in}}$	invert elevation of pipe from base of tank 2	2.5 ft
Z_2	elevation from base of tank 2 to orifice	1 ft

ft = feet, s = seconds, min = minutes, hr = hours.

Table 3.3: Empirical Surface Runoff Probability Distribution

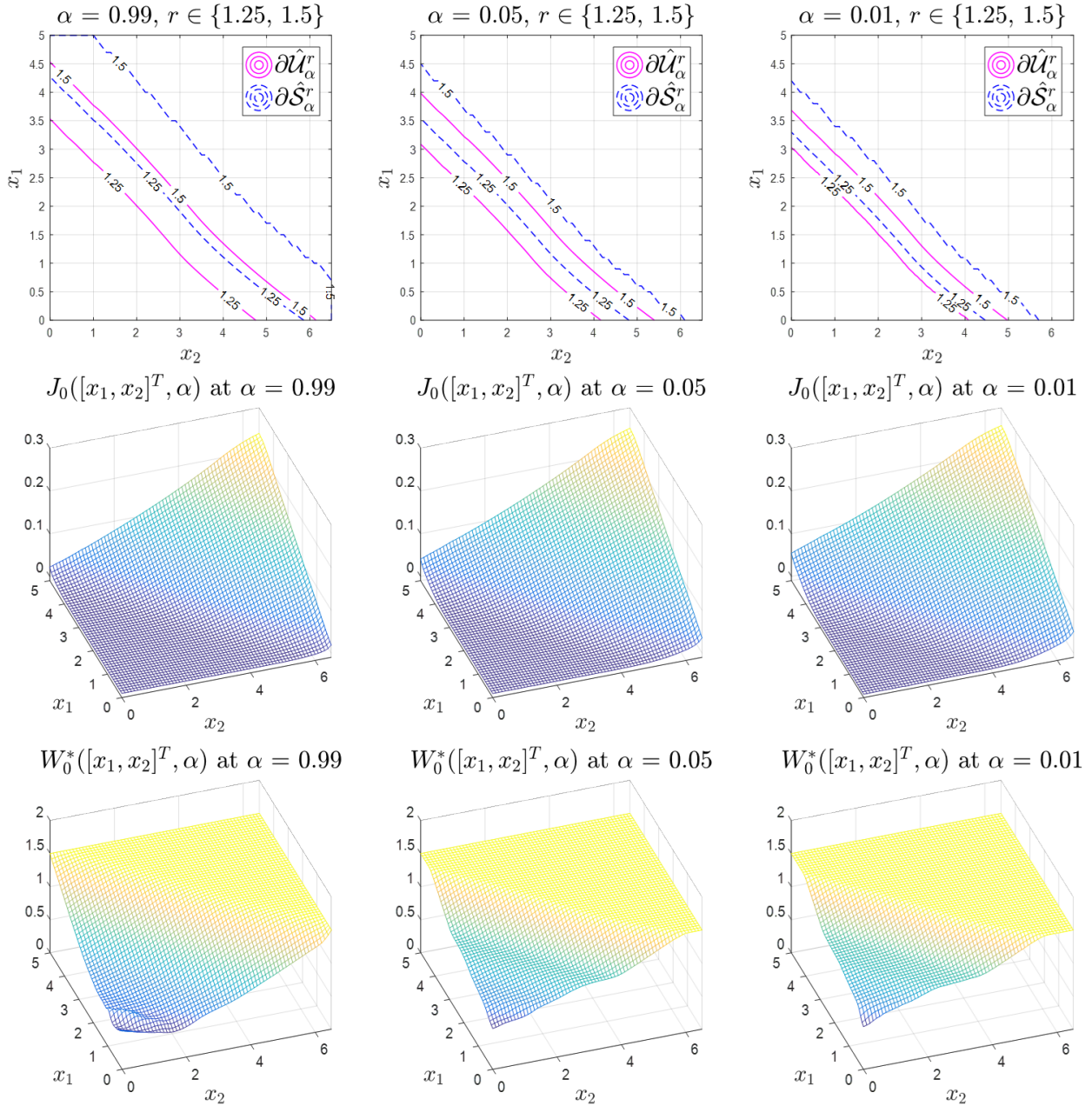
Possible value of d_t (ft ³ /s)	8.6	9.5	10.4	11.3	12.2	13.1	14.0	14.9	15.8	16.7
Probability $P(d_t)$	0.02	10^{-4}	10^{-4}	0.52	0.33	10^{-4}	10^{-4}	10^{-4}	10^{-4}	0.12

The first three empirical moments are: mean = 12.16 ft³/s, variance = 3.22 ft⁶/s², and skewness = 1.68 ft⁹/s³. Last digits are approximate.

of $\max\{g_K(x_t) : k = 0, 1, \dots, T\}$ initialized at each $x_0 = \mathbf{x}$ in the state space grid under a policy computed by Algorithm 1, and by utilizing a consistent CVAR estimator [92, p. 300].

Numerical approximations of $\mathcal{U}_\alpha^r$ and $\mathcal{S}_\alpha^r$ (denoted by $\hat{\mathcal{U}}_\alpha^r$ and $\hat{\mathcal{S}}_\alpha^r$, respectively) for the non-linear stormwater system model are presented in Fig. 3.11. Although we have shown the correctness of Algorithm 1 and the existence of optimal pre-commitment policies for a class of linear systems, we have found that $\hat{\mathcal{U}}_\alpha^r$ provides an under-approximation of $\hat{\mathcal{S}}_\alpha^r$ when $\beta e^{\gamma r}$ is sufficiently large.¹⁵ These results demonstrate that Algorithm 1 is tractable when applied to a realistic (yet low-dimensional) numerical example.

¹⁵ $\hat{\mathcal{U}}_\alpha^r$ provides an under-approximation of $\hat{\mathcal{S}}_\alpha^r$ when $\beta e^{\gamma r}$ is on the order of 10^{-4} or greater for $\alpha \in \{0.99, 0.05, 0.01\}$. When $\beta e^{\gamma r}$ is too small, numerical inaccuracies in the approximation of $\mathcal{U}_\alpha^r$ may be amplified by the transformation from $J_0 \leq \beta e^{\gamma r}$ to $\frac{1}{\gamma} \log(J_0/\beta) \leq r$. The numerical stability of the computations requires a careful selection of (β, γ) that balances the desire to make γ as large as possible with the understanding that β cannot be made arbitrarily small without introducing additional numerical issues.



Chapter 4

Future Directions

In the final chapter of this thesis, we present exciting future research directions. Some directions are more applied, and others are more theoretical. Many of the directions have been inspired by feedback and questions that were asked during the presentation of this thesis during the spring of 2020.

4.1 Risk-Sensitive Control of Cancer Systems

While the estimation of stochastic dynamical models for biological systems is challenging, developing treatment strategies that leverage ideas from risk-sensitive control has the potential to improve the management of various types of diseases. In the context of cancer, we would like to design treatment strategies that balance the need to kill cancer cells in a timely manner while also not causing undue harm to healthy cells. Designing treatments in a risk-sensitive context is especially important since a worst-case approach (e.g., applying very high doses of drugs over a short period of time) can have lasting long-term consequences on quality of life, even if the cancer is eradicated; and if the cancer is not eradicated, it can become resistant to future treatments. In a pre-clinical setting, important uncertainties to consider include drug-drug interactions when different drugs are applied sequentially and how the microenvironment can influence the response of the disease to treatment. In the near term, the extent of harm to healthy cells can be estimated by treating healthy cell populations with cancer therapies and measuring how the numbers of live and dying cells change over time.

In addition, one can focus on developing improved treatment strategies for blood cancers, such as *Chronic Lymphocytic Leukemia* (CLL), since the progression of this disease can be observed frequently in a minimally invasive way. In CLL, how to sequence therapy and when to administer therapy to individual patients are open questions (e.g., which patients require earlier intervention?). The number of white blood cells can be measured, and if this population grows too quickly, some oncologists will apply treatment and others will not. Ibrutinib is the front-line treatment for CLL, which is a pill that is taken daily. Some

patients develop resistance to this drug and relapse, and if this occurs, other drugs are tried. CLL progresses relatively slowly, so there is time to collect sufficient time series data for system identification and control.

One could design a clinical trial for CLL patients that involves: 1) collecting patient blood samples over time to measure how the numbers of cancer cells and healthy cells are evolving in response to treatments, 2) developing patient-specific dynamical models using these data, and 3) using these data and models for improved treatment design. Developing stochastic dynamical models and treatment strategies using blood samples is promising since these samples capture patient-specific variations over time that cannot be observed in cell lines. Moreover, using a risk-sensitive cost function, such as Conditional Value-at-Risk or Exponential Utility, is important for giving flexibility to the notion of worst case in the setting of disease management, where outcomes are highly uncertain.

Overall, there is a vital need to study certain diseases as *systems that evolve over time* in response to therapeutic interventions, where these responses are uncertain due to patient-specific genetic, epigenetic, and environmental factors. Developing new technologies that are designed to take measurements of biological systems over time is critical for designing improved treatment strategies that balance the risks of disease progression, the onset of therapeutic resistance, and the occurrence of adverse secondary reactions. Quantifying how the state of the disease is evolving under the current therapeutic strategy and re-designing the strategy based on this knowledge in a scientific way is necessary for improving quality of life in the long term. We are hopeful that transferring ideas from risk-sensitive control, classic feedback control, and system identification to the problem of disease management will benefit patient outcomes.

In this section, we have discussed an exciting applied research direction related to cancer systems and risk-sensitive control. In the next section, we present an important theoretical research direction for risk-sensitive safety analysis.

4.2 Approximate Risk-Sensitive Safety Analysis

Developing a new method for approximate risk-sensitive safety analysis is required for quantifying the safety of high-dimensional stochastic systems, such as societal-scale infrastructure systems. Specifically, we would like to estimate risk-sensitive safe sets and control policies with quality-of-approximation guarantees for high-dimensional systems which may not have analytical models. We plan to begin this effort by studying the foundations of approximate dynamic programming [11] and multi-armed bandits [94]. Approximate dynamic programming methods are scalable to high dimensions, and the multi-armed bandits literature contains well-established quality-of-approximation guarantees. Developing a new grid-free method that leverages ideas from approximate dynamic programming and multi-armed bandits may provide a more scalable risk-sensitive safety analysis approach.

In the next section, we discuss potential alternate approaches to risk-sensitive safety analysis (Chapter 3), which have been inspired by recent conversations during the presentation

of this thesis.

4.3 Potential Alternate Approaches to Risk-Sensitive Safety Analysis

State Augmentation via Maximum

It may be possible to estimate risk-sensitive safe sets without using the log-sum-exp function to approximate the maximum. For example, to solve the following stochastic optimal control problem,

$$\inf_{\pi \in \bar{\Pi}} \mathbb{E}^{\pi} \left(\max_{t=0,1,\dots,T} g_K(x_t) \middle| x_0 = \mathbf{x} \right) \quad (4.1a)$$

subject to a stochastic dynamical system model, $x_{t+1} = f_t(x_t, u_t, d_t)$ for $t = 0, 1, \dots, T-1$, where $\bar{\Pi}$ is an appropriate set of control policies, one may define

$$z_t := \max_{k=0,1,\dots,t} g_K(x_k) = \max(z_{t-1}, g_K(x_t)) \quad (4.1b)$$

and perform value iteration on the augmented state (x_t, z_t) . The value function at time t takes the following form:

$$V_t(x_t, z_t) := \inf_{u_t \in U} \mathbb{E} (V_{t+1}(f_t(x_t, u_t, d_t), z_{t+1})), \quad (4.1c)$$

where the expectation is taken with respect to the probability distribution of the disturbance d_t . It may be possible to adapt (4.1) in addition to using the machinery from Chapter 3 to estimate risk-sensitive safe sets (Definition 4).

CVAR-MDP via Dynamic Games

In this thesis, we have estimated risk-sensitive safe sets by first reducing the non-standard stochastic optimal control problem which defines these sets to the solution to a CVAR-MDP problem (Definition 4, Theorem 3). Then, we have estimated the solution to the CVAR-MDP problem for a class of linear systems by using a non-standard value iteration algorithm on the state space augmented by the space of risk-sensitivity levels (Theorem 4). This non-standard value iteration algorithm is based on a tower-like property for the Conditional Value-at-Risk measure, which allows us to express CVAR at a current risk-sensitivity level in terms of CVAR at a future risk-sensitivity level [79, Thm. 6 (iii)]. Proving the correctness of the non-standard value iteration algorithm for a class of linear systems requires us to show a minimax equality (Lemma 4).

Thus, a natural question is whether we can improve our method for solving the CVAR-MDP problem by leveraging the following well-established *dual representation* for the Conditional Value-at-Risk measure:

$$\text{CVAR}_\alpha(Y) = \sup_{\xi \in \mathcal{A}} \mathbb{E}_\xi(Y), \quad (4.2)$$

where $\alpha \in [0, 1]$ is the risk-sensitivity level, Y is a bounded random cost variable, and $\mathcal{A}$ is a specific set of probability density functions that depends on α [92, Thm. 6.4, Eqn. 6.40, Eqn. 6.70]. In the CVAR-MDP problem (3.15), we minimize an expression similar to “ $\text{CVAR}_\alpha(Y)$ ” over an appropriate set of control policies, meaning that we minimize an expression similar to “ $\sup_{\xi \in \mathcal{A}} \mathbb{E}_\xi(Y)$ ” by the dual representation (4.2). Hence, the CVAR-MDP problem is a minimax optimization problem that could possibly be solved by using existing results from the dynamic games literature [37].

CVAR-MDP via Hamilton-Jacobi-Bellman

A very interesting paper that has recently been published [6] provides the solution to a stochastic optimal control problem in continuous time, where the cost of the state trajectory can be assessed via Conditional Value-at-Risk, using a Hamilton-Jacobi-Bellman partial differential equation. A challenging, yet promising and important, future research direction is to transfer the methodology provided by [6] to a stochastic control system of the following form:

$$\dot{x} = f(x(t), u(t), d(t)), \quad (4.3)$$

where f satisfies the usual assumptions so that the ordinary differential equation (4.3) has a unique solution, and an appropriate set of control policies must be specified. Additional related research contributions include: 1) the development of a software toolbox to estimate the solution to the Hamilton-Jacobi-Bellman partial differential equation provided by [6] and a sub-optimal control policy *efficiently with quality-of-approximation guarantees*, and 2) leveraging this toolbox for risk-sensitive safety analysis of control systems. Key references for this research direction in addition to [6] include [69], [51], and [57], where the last is a classic reference on stochastic optimal control theory.

Application of Large Deviations Theory

The Conditional Value-at-Risk of a continuous random cost variable Y at risk-sensitivity level α is the expectation of the $\alpha \cdot 100\%$ worst realizations of Y . Hence, is it possible to solve the CVAR-MDP problem (3.15) by shifting the distribution of the random cost of the state trajectory and then utilize standard value iteration, which evaluates an expectation at each time step (e.g., see [10])? Studying Large Deviations Theory, which focuses on low probability events that are not characterized by the Central Limit Theorem, may help answer this question. References on Large Deviations Theory include [95] and [103].

Exponential Utility to Quantify Asymmetric Distributions

A key advantage of using Conditional Value-at-Risk for optimal control is the ability of this measure to quantify asymmetric cost distributions, i.e., distributions that cannot be well-approximated by bell-shaped curves. An interesting question is whether one can use Exponential Utility to quantify the skew of a distribution (and other higher-order moments) sufficiently well. Recall from (3.1) that the Exponential Utility of a random cost variable Y is defined as follows:

$$\rho_{e,\theta}(Y) := \frac{-2}{\theta} \log \mathbb{E}(e^{\frac{-\theta}{2}Y}) = \mathbb{E}(Y) - \frac{\theta}{4}\mathbb{V}(Y) + O(\theta^2),$$

where $\theta \in (-1, 0) \cup (0, 1)$ is the risk-sensitivity level, $\mathbb{E}(Y)$ is the expectation of Y , and $\mathbb{V}(Y)$ is the variance of Y [107, Eqn. 1.10, Eqn. 1.11]. The second equality holds since $0 < |\theta| < 1$, which can be shown using two Taylor expansions by grouping together the higher-order terms in $O(\theta^2)$. Specific research questions to study include the following:

1. What range of θ can be used to quantify the first, second, *and* third moments of the distribution of Y ?
2. Can the third moment of the distribution of Y be quantified sufficiently well for the range of θ specified above?
3. Or, is the weight on the third moment too small for practical applications? (This weight may be proportional to θ^p for some $p \geq 2$.)

One can use Taylor expansions of the logarithm and the exponential functions to investigate the above questions. If the third moment of the distribution of Y can be quantified sufficiently well using Exponential Utility, then one may not require the use of CVAR to represent asymmetric cost distributions, which may be desirable due to more straightforward theory and implementation when using Exponential Utility.

In this section, we have presented potential alternate and improved approaches to risk-sensitive safety analysis, which may be interesting to investigate in the future. In the next section, we present important extensions to risk-sensitive safety analysis to generalize the methodology to a larger class of dynamical systems and risk measures.

4.4 Extensions to Risk-Sensitive Safety Analysis

This section is organized into short-term and long-term extensions to risk-sensitive safety analysis based on perceived difficulty.

Short-Term Extensions

Interpretability of Regularity Condition on Risk Envelope. An important direction for future work is to identify standard examples that satisfy Assumption 2 to improve the

interpretability of this assumption. For example, if the probability space upon which the random disturbance d_t is defined is finite, can we prove that Assumption 2 holds? If the probability space is Gaussian, can we prove that Assumption 2 holds?

Time-Varying Probability Distributions. Disturbances to control systems may have probability distributions that depend on time, for example, when the disturbances are related to weather patterns. One can extend risk-sensitive safety analysis to the setting of time-varying probability distributions, where the random disturbances $(d_0, d_1, \dots, d_{T-1})$ are independent but may not be identically distributed. One can start by studying how the probability distributions are formalized in [32, Sec. II].

CVAR Objective versus CVAR Constraint. In this thesis, we use CVAR to evaluate the cost of the state trajectory, i.e., “ $\text{CVAR}_\alpha(\text{Cost}(x_0, x_1, \dots, x_T))$,” whereas other papers have used CVAR to constrain the cost at each state in the trajectory, i.e., “ $\text{CVAR}_\alpha(\text{Cost}(x_t))$ for all t ” [89] [102]. Using practical examples, it is important to evaluate the advantages and disadvantages of each approach. One can start this work by observing that our approach provides a risk-sensitive safety metric for the entire state trajectory and is more conservative.

Long-Term Extensions

Generalization to a Larger Class of Risk Measures. We have developed risk-sensitive safety analysis using the Conditional Value-at-Risk measure, and it would be interesting to extend this analysis to other types of risk measures. One could start by studying the *Optimized Certainty Equivalent* risk measures, which have been utilized recently in an optimal control setting by Backhoff-Veraguas and Tangpi [6]. This research direction could include the following steps:

1. Define a more general notion of a risk-sensitive safe set using an Optimized Certainty Equivalent risk measure;
2. Specify a dynamical system equipped with an appropriate space of control policies and transfer the methodology in [6] to this system;
3. Develop a *scalable* numerical method to approximately solve the Hamilton-Jacobi-Bellman partial differential equation proposed by [6] with quality-of-approximation guarantees; and
4. Create an open-source, well-documented software toolbox to compute approximate risk-sensitive safe sets using the numerical method above.

An especially exciting component of the above research direction is to specify the mathematical assumptions on the dynamical system and on the set of control policies so that the resulting methodology is practical for many applications (i.e., efficient computations are possible for a broad class of systems).

Generalization to a Larger Class of Dynamical Systems. The computational approach to risk-sensitive safety analysis in this thesis is applicable in theory to a class of linear systems,

and we would like to extend its theoretical applicability to a broader class of dynamical systems. One could explore extensions to non-linear systems by utilizing operators that transform non-linear dynamics to linear dynamics on an infinite-dimensional space, such as the Koopman operator [56].¹ Another approach is to impose structural assumptions, such as the system being feedback linearizable or having cone-bounded non-linearities [90].

Very challenging but important extensions include the development of efficient computational approaches to risk-sensitive safety analysis for *partially observable* systems, *high-dimensional* systems, and systems that do not necessarily have dynamics that can be fully expressed in terms of analytical equations. A societal-scale infrastructure system (e.g., food, energy, water, or transportation systems, or a subset of these systems in combination), for example, is partially observable, high-dimensional, and may not have a complete analytical dynamics representation. Developing a safety analysis method for a part of one of these cases would be a fundamental research contribution. Considering an infinite time horizon, which has been studied by [25], may be useful so that the computation time does not scale with the number of time points.

It is also critical to consider systems where the probability distribution of the disturbance is not known exactly but lives within a known family of distributions, or *ambiguity set* (e.g., see [110]). In this setting, one could develop a *distributionally robust* risk-sensitive safety analysis method. An important extension involves developing a risk-sensitive safety analysis method with *non-asymptotic guarantees*, where only a finite number of samples can be used to estimate the probability distribution of the disturbance. One may even consider developing an efficient *risk-sensitive uncertainty estimation* method using a finite number of samples to approximate the $\alpha \cdot 100\%$ worst mean, variance, and skew of the distribution with non-asymptotic guarantees. In the final section, we present additional exciting future research directions and concluding remarks.

4.5 Additional Exciting Future Directions

This thesis has inspired a wide range of interesting future research directions, many of which we have discussed above. Additional future projects include: the extension of finite-time stability analysis for stochastic switched dynamical systems with application to cancer treatment [112] [109] [24] [111] [35]; the application of multi-armed bandits to optimize therapeutic schedules for cancer cell populations that stochastically evolve into subpopulations of varying degrees of sensitivity;² the application of risk measures, such as Conditional Value-at-Risk, to alleviate conservativeness in the analysis and design of optimization algorithms and

¹A challenge with using the Koopman operator is that the lifted dynamical system is control-affine.

²Consider a setting with d distinct therapies and d distinct types of cancer cells. Each therapy can kill a particular type of cancer cell over time but may cause the other types of cancer cells to proliferate. The cancer cell population is assumed to stochastically evolve after a random waiting time into subpopulations containing the original cell type(s) and a randomly chosen cell type. The cancer cell population is homogeneous initially. We can only administer one therapy at a time, and we would like to design a treatment strategy that is guaranteed in probability to reduce the size of the cancer cell population to a given threshold by a given

to broaden the assumptions on the objective functions that are typically used [61] [71] [68]; leveraging risk measures [92] and non-asymptotic statistical analysis [105] to develop uncertainty quantification methods that balance the risk-neutral and worst-case perspectives with non-asymptotic guarantees that use the number of available data samples; the application of risk-sensitive control to develop food processing or production strategies with reduced water and carbon footprints using sensor data; and writing a tutorial-style paper on practical examples of linear control systems. We conclude this thesis by presenting, what we believe to be, the fundamental open research areas in the field of systems and control.

Open theoretical research areas in the field of systems and control concern the development of: 1) numerical methods that scale to high-dimensional systems but also provide safety or quality-of-approximation guarantees for these systems by exploiting domain-specific structure; 2) numerical methods with safety or quality-of-approximation guarantees that are devised with respect to more realistic assumptions about uncertain disturbances using available data;³ 3) controller synthesis methods that more rigorously integrate physics-based models, data-driven models, partially observable states, and states that have no analytical dynamics representation; and 4) new approaches to decentralized control for high-dimensional systems (in adversarial or non-adversarial settings) that leverage the ability of modern computers to easily exchange information and facilitate coordination between low-dimensional subsystems.

Open applied research areas in the field of systems and control concern the translation and transfer of mathematical and numerical methods to address safety-critical challenges in society, which involve the protection of the *health of the planet*, including the health of humankind. Specific applications include transferring systems and control methodologies to inform decision-making under uncertainty in disease and healthcare systems and in societal-scale infrastructure systems, such as food, water, energy, transportation, and their interactions. We are hopeful that these ideas will inspire scholarship that is focused on developing, not technologies for convenience, but technologies that promote the safety and well-being of the planet and its people.

finite time. How do we efficiently design a treatment strategy for this setting or further specify the setting so that a treatment strategy can be efficiently designed?

³Both non-asymptotic and asymptotic theoretical guarantees are important as well as the integration of real-time data samples and previously existing data sets.

Bibliography

- [1] Alessandro Abate et al. “Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems”. In: *Automatica* 44.11 (2008), pp. 2724–2734.
- [2] Anayo Akametalu. “A Learning-Based Approach to Safety for Uncertain Robotic Systems”. PhD thesis. Electrical Engineering and Computer Sciences Department, University of California, Berkeley, May 2018. URL: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-41.html>.
- [3] Vanessa Almendro et al. “Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity”. In: *Cell Reports* 6.3 (2014), pp. 514–527.
- [4] Philippe Artzner et al. “Coherent multiperiod risk adjusted values and Bellman’s principle”. In: *Annals of Operations Research* 152.1 (2007), pp. 5–22.
- [5] Robert B. Ash. *Real Analysis and Probability*. New York: Academic Press, 1972.
- [6] Julio Backhoff-Veraguas and Ludovic Tangpi. “On the dynamic representation of some time-inconsistent risk measures in a Brownian filtration”. In: *Mathematics and Financial Economics* (2020), pp. 1–28.
- [7] Viorel Barbu and Teodor Precupanu. *Convexity and Optimization in Banach Spaces*. Dordrecht: Springer Science & Business Media, 2012.
- [8] Tamer Başar and Pierre Bernhard. *H-infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. 2nd ed. Boston: Birkhauser, 1995.
- [9] Dimitri P. Bertsekas. “Control of Uncertain Systems with a Set-Membership Description of the Uncertainty.” PhD thesis. Massachusetts Institute of Technology, 1971.
- [10] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. 4th ed. Vol. 1. Belmont: Athena Scientific, 2017.
- [11] Dimitri P. Bertsekas. *Reinforcement Learning and Optimal Control*. Belmont: Athena Scientific, 2019.
- [12] Dimitri P. Bertsekas and Ian B. Rhodes. “On the minimax reachability of target sets and target tubes”. In: *Automatica* 7.2 (1971), pp. 233–247.
- [13] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

- [14] Stephen Boyd et al. *Linear Matrix Inequalities in System and Control Theory*. Vol. 15. Philadelphia: Society for Industrial and Applied Mathematics, 1994.
- [15] Lon R. Cardon and Gary D. Stormo. “Expectation maximization algorithm for identifying protein-binding sites with variable lengths from unaligned DNA fragments”. In: *Journal of Molecular Biology* 223.1 (1992), pp. 159–170.
- [16] Tony F. Chan and Chiu-Kwong Wong. “Convergence of the alternating minimization algorithm for blind deconvolution”. In: *Linear Algebra and its Applications* 316.1 (2000), pp. 259–285.
- [17] Margaret P. Chapman et al. “A model of phenotypic state dynamics initiates a promising approach to control heterogeneous malignant cell populations”. In: *55th IEEE Conference on Decision and Control*. © 2016 IEEE. 2016, pp. 2481–2487.
- [18] Margaret P. Chapman et al. “A Risk-Sensitive Finite-Time Reachability Approach for Safety of Stochastic Dynamic Systems”. In: *2019 American Control Conference*. © 2019 IEEE. 2019, pp. 2958–2963.
- [19] Margaret P. Chapman et al. “Modeling differentiation-state transitions linked to therapeutic escape in triple-negative breast cancer”. In: *PLoS Computational Biology* 15.3 (2019), e1006840.
- [20] Margaret P. Chapman et al. “On the Analysis of Cyclic Drug Schedules for Cancer Treatment using Switched Dynamical Systems”. In: *57th IEEE Conference on Decision and Control*. © 2018 IEEE. Reprinted with permission. 2018, pp. 3503–3509.
- [21] Margaret P. Chapman et al. “Reachability Analysis as a Design Tool for Stormwater Systems”. In: *6th IEEE Conference on Technologies for Sustainability*. © 2018 IEEE. 2018, pp. 1–8.
- [22] Margaret P. Chapman et al. “Risk-sensitive safety specifications for stochastic systems using Conditional Value-at-Risk”. In: *arXiv Preprint arXiv:1909.09703* (2019).
- [23] Mo Chen and Claire J. Tomlin. “Hamilton-Jacobi Reachability: Some recent theoretical advances and applications in unmanned airspace management”. In: *Annual Review of Control, Robotics, and Autonomous Systems* 1 (2018), pp. 333–358.
- [24] Weisheng Chen and L. C. Jiao. “Finite-Time Stability Theorem of Stochastic Nonlinear Systems”. In: *Automatica* 46.12 (2010), pp. 2105–2108.
- [25] Yinlam Chow et al. “Risk-Sensitive and Robust Decision-Making: a CVaR Optimization Approach”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 1522–1530.
- [26] I. Csiszár and Gábor Tusnády. “Information geometry and alternating minimization procedures”. In: *Statistics and Decisions* (1984), pp. 205–237.
- [27] Jamal Daafouz, Pierre Riedinger, and Claude Iung. “Stability Analysis and Control Synthesis for Switched Systems: A Switched Lyapunov Function Approach”. In: *IEEE Transactions on Automatic Control* 47.11 (2002), pp. 1883–1887.

- [28] Russell Davidson and Emmanuel Flachaire. “The wild bootstrap, tamed at last”. In: *Journal of Econometrics* 146.1 (2008), pp. 162–169.
- [29] Jerry Ding et al. “A stochastic games framework for verification and control of discrete time stochastic hybrid systems”. In: *Automatica* 49.9 (2013), pp. 2665–2674.
- [30] Chuong B. Do and Serafim Batzoglou. “What is the expectation maximization algorithm?” In: *Nature Biotechnology* 26.8 (2008), pp. 897–899.
- [31] Lorenzo Farina and Sergio Rinaldi. *Positive Linear Systems: Theory and Applications*. New York: John Wiley & Sons, 2000.
- [32] Michael Fauß and Abdelhak M. Zoubir. “Old Bands, New Tracks—Revisiting the Band Model for Robust Hypothesis Testing”. In: *IEEE Transactions on Signal Processing* 64.22 (2016), pp. 5875–5886.
- [33] Jaime F. Fisac et al. “Reach-avoid problems with time-varying dynamics, targets and constraints”. In: *18th International Conference on Hybrid Systems: Computation and Control*. 2015, pp. 11–20.
- [34] Gerald B. Folland. *Real Analysis: Modern Techniques and their Applications*. New York: John Wiley & Sons, 2013.
- [35] Kunal Garg and Dimitra Panagou. “Finite-Time Stability of Hybrid Systems: A Multiple Generalized Lyapunov Functions Approach”. In: *arXiv preprint arXiv:1906.09109* (2019).
- [36] Aaron Goldman et al. “Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition”. In: *Nature Communications* 6.6139 (2015).
- [37] J. I. González-Trejo, Onésimo Hernández-Lerma, and Luis F. Hoyos-Reyes. “Minimax control of discrete-time stochastic systems”. In: *SIAM Journal on Control and Optimization* 41.5 (2002), pp. 1626–1659.
- [38] Roy Z. Granit et al. “Regulation of cellular heterogeneity and rates of symmetric and asymmetric divisions in triple-negative breast cancer”. In: *Cell Reports* 24.12 (2018), pp. 3237–3250.
- [39] Roy Z. Granit, Michal Slyper, and Ittai Ben-Porath. “Axes of differentiation in breast cancer: untangling stemness, lineage identity, and the epithelial to mesenchymal transition”. In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 6.1 (2014), pp. 93–106.
- [40] Michael Grant and Stephen Boyd. *CVX: Matlab Software for Disciplined Convex Programming, Version 2.1*. 2014. URL: <http://cvxr.com/cvx>.
- [41] Michael Grant and Stephen Boyd. “Graph Implementations for Nonsmooth Convex Programs”. In: *Recent Advances in Learning and Control*. Ed. by V. Blondel, S. Boyd, and H. Kimura. Vol. 371. Lecture Notes in Control and Information Sciences. London: Springer, 2008, pp. 95–110.

- [42] J. E. Grilley-Olson, P. L. Bedard, A. Fasolo, et al. “A Phase Ib Dose-Escalation Study of the MEK Inhibitor Trametinib in Combination with the PI3K/mTOR Inhibitor GSK2126458 in Patients with Advanced Solid Tumors”. In: *Investigational New Drugs* 34.6 (2016), pp. 740–749. ISSN: 1573-0646. DOI: 10.1007/s10637-016-0377-0. URL: <https://doi.org/10.1007/s10637-016-0377-0>.
- [43] Piyush B. Gupta et al. “Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells”. In: *Cell* 146.4 (2011), pp. 633–644.
- [44] Astghik Hakobyan, Gyeong Chan Kim, and Insoon Yang. “Risk-Aware Motion Planning and Control Using CVaR-Constrained Optimization”. In: *IEEE Robotics and Automation Letters* 4.4 (2019), pp. 3924–3931.
- [45] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Second Edition. Springer Science & Business Media, 2009.
- [46] Onésimo Hernández-Lerma and Jean B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Vol. 30. New York: Springer-Verlag, 1996.
- [47] Esteban A. Hernandez-Vargas, Patrizio Colaneri, and Richard H. Middleton. “Optimal Therapy Scheduling for a Simplified HIV Infection Model”. In: *Automatica* 49.9 (2013), pp. 2874–2880.
- [48] Esteban A. Hernandez-Vargas, Patrizio Colaneri, and Richard H. Middleton. “Switching Strategies to Mitigate HIV Mutation”. In: *IEEE Transactions on Control Systems Technology* 22.4 (2014), pp. 1623–1628.
- [49] Esteban A. Hernandez-Vargas et al. “Discrete-time Control for Switched Positive Systems with Application to Mitigating Viral Escape”. In: *International Journal of Robust and Nonlinear Control* 21.10 (2011), pp. 1093–1111.
- [50] Arthur E. Hoerl and Robert W. Kennard. “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. In: *Technometrics* 12.1 (1970), pp. 55–67.
- [51] Ying Hu, Hanqing Jin, and Xun Yu Zhou. “Time-Inconsistent Stochastic Linear-Quadratic Control: Characterization and Uniqueness of Equilibrium”. In: *SIAM Journal on Control and Optimization* 55.2 (2017), pp. 1261–1279.
- [52] Gareth James et al. *An Introduction to Statistical Learning with Applications in R*. New York: Springer Science & Business Media, 2013.
- [53] Mikael Johansson and Anders Rantzer. “Computation of Piecewise Quadratic Lyapunov Functions for Hybrid Systems”. In: *IEEE Transactions on Automatic Control* 43.4 (1998), pp. 555–559.
- [54] Maryam Kamgarpour et al. “Discrete Time Stochastic Hybrid Dynamical Games: Verification & Controller Synthesis”. In: *50th IEEE Conference on Decision and Control and European Control Conference*. IEEE. 2011, pp. 6122–6127.

- [55] Jakob Kisiala. “Conditional Value-at-Risk: Theory and Applications”. MA thesis. The School of Mathematics, The University of Edinburgh, Aug. 2015. URL: https://www.maths.ed.ac.uk/~prichtar/docs/Kisiala_Dissertation.pdf.
- [56] Milan Korda and Igor Mezić. “On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator”. In: *Journal of Nonlinear Science* 28.2 (2018), pp. 687–710.
- [57] Harold J. Kushner. *Stochastic Stability and Control*. Tech. rep. Brown University, Providence, Rhode Island, 1967.
- [58] R. H. Kwong. “On the LQG problem with correlated noise and its relation to minimum variance control”. In: *26th IEEE Conference on Decision and Control*. Vol. 26. IEEE. 1987, pp. 763–767.
- [59] Giovanni M. Lasio, Bruce R. Whiting, and Jeffrey F. Williamson. “Statistical reconstruction for x-ray computed tomography using energy-integrating detectors”. In: *Physics in Medicine and Biology* 52.8 (2007), p. 2247.
- [60] Brian D. Lehmann et al. “Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies”. In: *The Journal of Clinical Investigation* 121.7 (2011), pp. 2750–2767.
- [61] Laurent Lessard, Benjamin Recht, and Andrew Packard. “Analysis and Design of Optimization Algorithms via Integral Quadratic Constraints”. In: *SIAM Journal on Optimization* 26.1 (2016), pp. 57–95.
- [62] Daniel Liberzon and Roger W. Brockett. “Nonlinear feedback systems perturbed by noise: Steady-state probability distributions and optimal control”. In: *IEEE Transactions on Automatic Control* 45.6 (2000), pp. 1116–1130.
- [63] Hai Lin and Panos J. Antsaklis. “Stability and Stabilizability of Switched Linear Systems: A Survey of Recent Results”. In: *IEEE Transactions on Automatic Control* 54.2 (2009), pp. 308–322.
- [64] Zhiyun Lin and Mireille E. Broucke. “On a reachability problem for affine hypersurface systems on polytopes”. In: *Automatica* 47.4 (2011), pp. 769–775.
- [65] Xingwen Liu. “Stability Analysis of Switched Positive Systems: A Switched Linear Copositive Lyapunov Function Method”. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 56.5 (2009), pp. 414–418.
- [66] Anirudha Majumdar, Amir Ali Ahmadi, and Russ Tedrake. “Control and verification of high-dimensional systems with DSOS and SDSOS programming”. In: *53rd IEEE Conference on Decision and Control*. IEEE. 2014, pp. 394–401.
- [67] Anirudha Majumdar, Amir Ali Ahmadi, and Russ Tedrake. “Control design along trajectories with sums of squares programming”. In: *2013 IEEE International Conference on Robotics and Automation*. IEEE. 2013, pp. 4054–4061.

- [68] Simon Michalowsky, Carsten Scherer, and Christian Ebenbauer. “Robust and Structure Exploiting Optimization Algorithms: An Integral Quadratic Constraint Approach”. In: *International Journal of Control* (2020), pp. 1–51.
- [69] Christopher W. Miller and Insoon Yang. “Optimal Control of Conditional Value-at-Risk in Continuous Time”. In: *SIAM Journal on Control and Optimization* 55.2 (2017), pp. 856–884.
- [70] Ian M. Mitchell, Alexandre M. Bayen, and Claire J. Tomlin. “A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games”. In: *IEEE Transactions on Automatic Control* 50.7 (2005), pp. 947–957.
- [71] Hesameddin Mohammadi, Meisam Razaviyayn, and Mihailo R. Jovanović. “Robustness of Accelerated First-Order Algorithms for Strongly Convex Optimization Problems”. In: *arXiv preprint arXiv:1905.11011* (2019).
- [72] Richard E. Mortensen and Kevin P. Haggerty. “A stochastic computer model for heating and cooling loads”. In: *IEEE Transactions on Power Systems* 3.3 (1988), pp. 1213–1219.
- [73] *NCI Dictionary of Cancer Terms*. URL: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/signaling-pathway>.
- [74] Arnab Nilim and Laurent El Ghaoui. “Robust Control of Markov Decision Processes with Uncertain Transition Matrices”. In: *Operations Research* 53.5 (2005), pp. 780–798.
- [75] Petter Nilsson and Necmiye Ozay. “Synthesis of separable controlled invariant sets for modular local control design”. In: *2016 American Control Conference*. IEEE. 2016, pp. 5656–5663.
- [76] Yuanling Niu, Yue Wang, and Da Zhou. “The phenotypic equilibrium of cancer cells: From average-level stability to path-wise convergence”. In: *Journal of Theoretical Biology* 386 (2015), pp. 7–17.
- [77] Andy Packard et al. “A Collection of Robust Control Problems Leading to LMIs”. In: *30th IEEE Conference on Decision and Control*. IEEE. 1991, pp. 1245–1250.
- [78] Georg Ch. Pflug and Alois Pichler. “Time-Consistent Decisions and Temporal Decomposition of Coherent Risk Functionals”. In: *Mathematics of Operations Research* 41.2 (2016), pp. 682–699.
- [79] Georg Ch. Pflug and Alois Pichler. “Time-inconsistent multistage stochastic programs: Martingale bounds”. In: *European Journal of Operational Research* 249.1 (2016), pp. 155–163.
- [80] Warren B. Powell. “A unified framework for stochastic optimization”. In: *European Journal of Operational Research* 275.3 (2019), pp. 795–821.

- [81] Maria Prandini and Jianghai Hu. “Application of reachability analysis for stochastic hybrid systems to aircraft conflict prediction”. In: *47th IEEE Conference on Decision and Control*. IEEE. 2008, pp. 4036–4041.
- [82] Aleix Prat et al. “Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer”. In: *Breast Cancer Research* 12.5 (2010), R68.
- [83] Biagio Ricceri and Stephen Simons. *Minimax Theory and Applications*. Vol. 26. Dordrecht: Springer Science & Business Media, 2013.
- [84] Tyler Risom et al. “Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer”. In: *Nature Communications* 9.3815 (2018).
- [85] Lewis A. Rossman. *Storm Water Management Model User’s Manual, Version 5.0*. Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency, 2010.
- [86] Bartek Roszak and Mireille E. Broucke. “Necessary and sufficient conditions for reachability on a simplex”. In: *44th IEEE Conference on Decision and Control*. IEEE. 2005, pp. 4706–4711.
- [87] Bartek Roszak and Mireille E. Broucke. “Reachability of a set of facets for linear affine systems with $n - 1$ inputs”. In: *IEEE Transactions on Automatic Control* 52.2 (2007), pp. 359–364.
- [88] Andrzej Ruszczyński. “Risk-averse dynamic programming for Markov Decision Processes”. In: *Mathematical Programming* 125.2 (2010), pp. 235–261.
- [89] Samantha Samuelson and Insoon Yang. “Safety-Aware Optimal Control of Stochastic Systems Using Conditional Value-at-Risk”. In: *2018 American Control Conference*. IEEE. 2018, pp. 6285–6290.
- [90] B. M. Scherzinger and R. H. Kwong. “Estimation and Control of Discrete Time Stochastic Systems having Cone-Bounded Non-Linearities”. In: *International Journal of Control* 36.1 (1982), pp. 33–52.
- [91] Pete Seiler and Raja Sengupta. “An H-infinity Approach to Networked Control”. In: *IEEE Transactions on Automatic Control* 50.3 (2005), pp. 356–364.
- [92] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics, Mathematical Programming Society, 2009.
- [93] Sumeet Singh et al. “A Framework for Time-Consistent, Risk-Sensitive Model Predictive Control: Theory and Algorithms”. In: *IEEE Transactions on Automatic Control* 64.7 (2018), pp. 2905–2912.
- [94] Aleksandrs Slivkins. “Introduction to Multi-Armed Bandits”. In: *Foundations and Trends® in Machine Learning* 12.1-2 (2019), pp. 1–286.
- [95] Daniel W. Stroock. *An Introduction to the Theory of Large Deviations*. Springer Science & Business Media, 2012.

- [96] Sean Summers and John Lygeros. “Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem”. In: *Automatica* 46.12 (2010), pp. 1951–1961.
- [97] Zhendong Sun. *Switched Linear Systems: Control and Design*. Springer Science & Business Media, 2006.
- [98] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.
- [99] Richard S. Sutton, Andrew G. Barto, and Ronald J. Williams. “Reinforcement learning is direct adaptive optimal control”. In: *IEEE Control Systems Magazine* 12.2 (1992), pp. 19–22.
- [100] *The MOSEK Optimization Toolbox for MATLAB Manual, Version 9.0*. 2019. URL: <http://docs.mosek.com/9.0/toolbox/index.html>.
- [101] Claire J. Tomlin. “Hybrid Control of Air Traffic Management Systems”. PhD thesis. Electrical Engineering and Computer Sciences Department, University of California, Berkeley, Sept. 1998. URL: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/1998/3511.html>.
- [102] Bart P. G. Van Parys et al. “Distributionally Robust Control of Constrained Stochastic Systems”. In: *IEEE Transactions on Automatic Control* 61.2 (2015), pp. 430–442.
- [103] S. R. S. Varadhan. *Large Deviations and Applications*. Vol. 46. Philadelphia: Society for Industrial and Applied Mathematics, 1984.
- [104] Zev A. Wainberg et al. “A Multi-Arm Phase I Study of the PI3K/mTOR Inhibitors PF-04691502 and Gedatolisib (PF-05212384) plus Irinotecan or the MEK Inhibitor PD-0325901 in Advanced Cancer”. In: *Targeted Oncology* (2017), pp. 1–11.
- [105] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Vol. 48. Cambridge University Press, 2019.
- [106] Yilun Wang et al. “A New Alternating Minimization Algorithm for Total Variation Image Reconstruction”. In: *SIAM Journal on Imaging Sciences* 1.3 (2008), pp. 248–272.
- [107] Peter Whittle. *Risk-sensitive Optimal Control*. Vol. 20. Chichester: John Wiley & Sons, 1990.
- [108] Chien-Fu Jeff Wu. “Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis”. In: *The Annals of Statistics* 14.4 (1986), pp. 1261–1295.
- [109] Zhiguo Yan, Weihai Zhang, and Guoshan Zhang. “Finite-Time Stability and Stabilization of Itô Stochastic Systems with Markovian Switching: Mode-Dependent Parameter Approach”. In: *IEEE Transactions on Automatic Control* 60.9 (2014), pp. 2428–2433.
- [110] Insoon Yang. “A dynamic game approach to distributionally robust safety specifications for stochastic systems”. In: *Automatica* 94 (2018), pp. 94–101.

- [111] Ying Yang, Junmin Li, and Guopei Chen. “Finite-Time Stability and Stabilization of Nonlinear Stochastic Hybrid Systems”. In: *Journal of Mathematical Analysis and Applications* 356.1 (2009), pp. 338–345.
- [112] Juliang Yin et al. “Finite-Time Stability and Instability of Stochastic Nonlinear Systems”. In: *Automatica* 47.12 (2011), pp. 2671–2677.
- [113] Boyang Zhao et al. “Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution”. In: *Cell* 165.1 (2016), pp. 234–246.
- [114] Da Zhou, Yue Wang, and Bin Wu. “A multi-phenotypic cancer model with cell plasticity”. In: *Journal of Theoretical Biology* 357 (2014), pp. 35–45.
- [115] Da Zhou et al. “Population dynamics of cancer cells with cell state conversions”. In: *Quantitative Biology* 1.3 (2013), pp. 201–208.
- [116] Kemin Zhou and John Comstock Doyle. *Essentials of Robust Control*. Upper Saddle River: Prentice Hall, 1998.

Appendix A

Supplementary Results for Chapter 3

A.1 Lemmas 6 and 7

Here we show two results that are required to prove the minimax equality (Lemma 4). Lemma 6 states that an expected CVAR of the cost of the state trajectory is convex and continuous on the policy space and is continuous on the risk envelope. Lemma 7 specifies important topological properties of the risk envelope.

Remark 9. Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined. The risk envelope $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is a subset of $L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$ since it is a collection of $\mathbb{P}$ -integrable, $\sigma(f_t(\mathbf{x}, \mathbf{u}, d_t))$ -measurable bounded functions on Ω by definition (3.13). “Bounded” means bounded in $\|\cdot\|_\infty$, which implies bounded in $\|\cdot\|_2$. We study properties of the risk envelope in $L^2 := L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$ since L^2 is reflexive and equals its dual, which simplifies proofs.

Lemma 6 (Convexity and Continuity). *Fix t , $x_t = \mathbf{x} \in X$, $u_t = \mathbf{u} \in U$, and $\alpha_t = \alpha \in [0, 1]$. Define $H : \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \times \Pi_{t+1} \rightarrow \mathbb{R}$ as follows:*

$$H(Z, \pi) := \int_{\Omega} Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^\pi(C_{t+1:T} | x_{t+1}) \mathbb{P}(d\omega),$$

where $C_{t+1:T}$ is conditioned on $x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)) := A_t \mathbf{x} + B_t \mathbf{u} + E_t d_t(\omega)$ in the integral above. Then, the following properties hold:

1. $H(Z, \cdot)$ is convex and continuous in the norm topology on Π_{t+1} for any $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$.
2. $H(\cdot, \pi)$ is Lipschitz continuous in the relative norm topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2 := L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$ for any $\pi \in \Pi_{t+1}$.

Proof. $C_{t+1:T}$ is convex on Π_{t+1} since $C_{t+1:T}$ is a convex function, the state trajectory is affine on Π_{t+1} , and the composition of a convex function with an affine function is convex. Further, $H(Z, \cdot)$ is convex on Π_{t+1} for any $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ because $C_{t+1:T}$ is convex on Π_{t+1} ,

Conditional Value-at-Risk is a convex functional, $Z(\omega) \geq 0$ for almost every $\omega \in \Omega$, and expectation is a linear functional.

Define the product norm on Π_{t+1} as the maximum uniform norm, i.e., $\|\pi\| := \max\{\|\mu_k\|_u : k = t+1, \dots, T-1\}$. Take $\{\pi^{(n)}\}$ in Π_{t+1} converging to $\pi \in \Pi_{t+1}$. Fix $r > 0$, define $\gamma_n := \frac{\|\pi - \pi^{(n)}\|}{r + \|\pi - \pi^{(n)}\|}$, and define $\bar{\pi}^{(n)} \in \Pi_{t+1}$ so that $\pi = \gamma_n \bar{\pi}^{(n)} + (1 - \gamma_n) \pi^{(n)}$. Fix $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$. Since $H(Z, \cdot)$ is convex on Π_{t+1} , we have

$$H(Z, \pi) \leq \gamma_n H(Z, \bar{\pi}^{(n)}) + (1 - \gamma_n) H(Z, \pi^{(n)}).$$

As $n \rightarrow \infty$, $\gamma_n H(Z, \bar{\pi}^{(n)}) \rightarrow 0$ since $H(Z, \cdot)$ is bounded and $\gamma_n \rightarrow 0$. So, $H(Z, \pi) \leq \liminf_{n \rightarrow \infty} H(Z, \pi^{(n)})$, which shows lower semi-continuity. To show upper semi-continuity, i.e., $\limsup_{n \rightarrow \infty} H(Z, \pi^{(n)}) \leq H(Z, \pi)$, define $\bar{\pi}^{(n)} \in \Pi_{t+1}$ so that $\pi^{(n)} = \gamma_n \bar{\pi}^{(n)} + (1 - \gamma_n) \pi$, use convexity to obtain $H(Z, \pi^{(n)}) \leq \gamma_n H(Z, \bar{\pi}^{(n)}) + (1 - \gamma_n) H(Z, \pi)$, and take the limit superior as $n \rightarrow \infty$.

Now fix $\pi \in \Pi_{t+1}$ and $\omega \in \Omega$. Use the definition of CVAR (3.2) as an integral over VAR if $\alpha Z(\omega) > 0$ and as an essential supremum if $\alpha Z(\omega) = 0$ to show that $Z(\omega) \mapsto Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^\pi(C_{t+1:T} | x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)))$ is Lipschitz continuous on $\mathbb{R}$ for almost every $\omega \in \Omega$; set the Lipschitz constant equal to the essential supremum of $C_{t+1:T}$. Since $\int_\Omega |Z(\omega)| \mathbb{P}(d\omega) := \|Z\|_1 \leq \|Z\|_2$ for any $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$, $H(\cdot, \pi)$ is Lipschitz continuous in the relative norm topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$ for any $\pi \in \Pi_{t+1}$. $\square$

Lemma 7 (Properties of Risk Envelope). *Fix t , $x_t = \mathbf{x} \in X$, $u_t = \mathbf{u} \in U$, and $\alpha_t = \alpha \in [0, 1]$. $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$, as defined by (3.13), is a non-empty, convex, and compact set in $L^2 := L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$ endowed with the weak topology.*

Proof. $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is non-empty since it contains $Z(\omega) = 1$ for almost every $\omega \in \Omega$. $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$ since bounded random variables have bounded second moments. So, $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is bounded in $\|\cdot\|_2$. $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is a convex subset of L^2 by inspection. Since only bounded costs are evaluated in our setting, for any $\alpha \in [0, 1]$, CVAR_α is a real-valued coherent risk measure on L^2 . Moreover, for any $\alpha \in [0, 1]$ and $Y \in L^2$ such that Y is bounded, the following equality holds by [92, Thm. 6.4, Eqn. 6.40, Eqn. 6.70]: $\text{CVAR}_\alpha(Y) = \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} \int_\Omega Z(\omega) Y(\omega) \mathbb{P}(d\omega)$. So, $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is closed in the weak* topology on L^2 by [92, Thm. 6.6, p. 264]. Finally, a bounded and weakly* closed set is weakly* compact by the Banach-Alaoglu Theorem [92, Thm. 7.70, p. 401]. Thus, $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is compact in the weak* topology on L^2 . Since L^2 is reflexive, the weak and the weak* topologies on L^2 coincide, hence $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is compact in the weak topology on L^2 . $\square$

Remark 10. L^2 endowed with the weak topology is a *locally convex* topological vector space since the weak topology in a topological vector space is locally convex [5, p. 161].

A.2 Lemmas 8-10

Here we provide three results that facilitate the synthesis of pre-commitment policies for the CVAR-MDP problem. Lemmas 8 and 9 are required to prove the existence of $\mathbb{X}$ -Markov control laws. Lemma 10 is critical for ensuring that the risk-sensitivity level dynamics are well-defined. These results require the following definition repeated from (3.21) for convenience:

$$G_t^\alpha(Z, \mathbf{x}, \mathbf{u}) := \int_{\Omega} Z(\omega) \cdot J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega),$$

where $(\mathbf{x}, \alpha) \in \mathbb{X}$, $\mathbf{u} \in U$, $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$, $f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)) := A_t \mathbf{x} + B_t \mathbf{u} + E_t d_t(\omega)$, and $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space upon which the random disturbance d_t is defined.

Lemma 8 (G_t^α is LSC in $(Z, \mathbf{u})$). *For any $(\mathbf{x}, \alpha) \in \mathbb{X}$, $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous in $(Z, \mathbf{u})$.*

Proof. Fix t and $(x_t, \alpha_t) = (\mathbf{x}, \alpha) \in \mathbb{X}$. Since $\mathbf{u} \mapsto f_t(\mathbf{x}, \mathbf{u}, d_t(\omega))$ is continuous for almost every $\omega \in \Omega$ by Assumption 1, $(Z(\omega), \mathbf{u}) \mapsto (f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ is continuous for almost every $\omega \in \Omega$. If J_{t+1}^* is lower semi-continuous on $\mathbb{X}$, then $(Z(\omega), \mathbf{u}) \mapsto J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ would be lower semi-continuous for almost every $\omega \in \Omega$, since a lower semi-continuous function composed with a continuous function is lower semi-continuous. Then, $(Z(\omega), \mathbf{u}) \mapsto Z(\omega) \cdot J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ would be lower semi-continuous for almost every $\omega \in \Omega$, since $Z(\omega)$ is nonnegative and bounded for almost every $\omega \in \Omega$. (The product of a nonnegative, continuous bounded function and a lower semi-continuous bounded function is lower semi-continuous.) Since expectation preserves lower semi-continuity, $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ would be lower semi-continuous in $(Z, \mathbf{u})$ to complete the proof. So, it suffices to show that J_{t+1}^* is lower semi-continuous on $\mathbb{X}$, which is shown in Lemma 9 below. $\square$

Lemma 9 (Lower Semi-Continuity of J_t^*). *Fix t . J_t^* , as defined in (3.15), is lower semi-continuous on $\mathbb{X} := X \times [0, 1]$.*

Proof. Recall that each J_t^* is bounded since the stage costs and the terminal cost are bounded. It suffices to show that J_t , as defined in (3.14), is lower semi-continuous on $\mathbb{X}$ by Theorem 4. We proceed by induction. Let $(\mathbf{x}, \alpha) \in \mathbb{X}$. $J_T(\mathbf{x}, \alpha) = c_T(\mathbf{x})$, which is continuous. Now, assume J_{t+1} is lower semi-continuous on $\mathbb{X}$ for some t . Since U is compact, c_t is continuous, and lower semi-continuity is preserved through summation, it suffices to show that the following (bounded) map $(\mathbf{x}, \alpha, \mathbf{u}) \mapsto \sup_{Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})} G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous on $\mathbb{X} \times U$

by [46, Prop. D.5 (b)], where $G_t^\alpha(Z, \mathbf{x}, \mathbf{u}) = \int_{\Omega} Z(\omega) \cdot J_{t+1}(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega)) \mathbb{P}(d\omega)$ by Theorem 4. Since $(\mathbf{x}, \alpha, \mathbf{u}) \mapsto \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ is a lower semi-continuous set-valued mapping by Assumption 2, if $G_t^\alpha(Z, \mathbf{x}, \mathbf{u})$ is lower semi-continuous in $(Z, \mathbf{x}, \alpha, \mathbf{u})$, then the desired result holds by [37, Lemma 3.2 (a)]. Since $(\mathbf{x}, \mathbf{u}) \mapsto f_t(\mathbf{x}, \mathbf{u}, d_t(\omega))$ is continuous for almost every $\omega \in \Omega$ by Assumption 1, $(Z(\omega), \mathbf{x}, \alpha, \mathbf{u}) \mapsto (f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ is continuous for almost every $\omega \in \Omega$. Since J_{t+1} is lower semi-continuous on $\mathbb{X}$ by the induction

hypothesis, $(Z(\omega), \mathbf{x}, \alpha, \mathbf{u}) \mapsto J_{t+1}(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ is lower semi-continuous for almost every $\omega \in \Omega$. Moreover, since $Z(\omega)$ is nonnegative and bounded for a.e. $\omega \in \Omega$, $(Z(\omega), \mathbf{x}, \alpha, \mathbf{u}) \mapsto Z(\omega) \cdot J_{t+1}(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ is lower semi-continuous for almost every $\omega \in \Omega$. Since expectation preserves lower semi-continuity, the proof is complete. $\square$

Lemma 10 (G_t^α is USC on Risk Envelope). *Fix t , $x_t = \mathbf{x}$, $u_t = \mathbf{u}$, and $\alpha_t = \alpha$. $G_t^\alpha(\cdot, \mathbf{x}, \mathbf{u})$, as defined by (3.21), is upper semi-continuous in the relative weak topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2 := L^2(\Omega, \sigma(f_t(\mathbf{x}, \mathbf{u}, d_t)), \mathbb{P})$.*

Proof. It suffices to show that $G_t^\alpha(\cdot, \mathbf{x}, \mathbf{u})$ is real-valued concave and upper semi-continuous in the relative norm topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$ by [7, Prop. 2.10]. $G_t^\alpha(\cdot, \mathbf{x}, \mathbf{u})$ is real-valued since $Z \in \mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ and J_{t+1}^* are bounded. $G_t^\alpha(\cdot, \mathbf{x}, \mathbf{u})$ is concave on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u})$ by [79, Thm. 12] applied to a random cost, since the pointwise infimum of concave functions is concave, by the definition of J_{t+1}^* , and by linearity of expectation. It suffices to show that $G_t^\alpha(\cdot, \mathbf{x}, \mathbf{u})$ is Lipschitz continuous in the relative norm topology on $\mathcal{R}_t^\alpha(\mathbf{x}, \mathbf{u}) \subseteq L^2$, as this implies upper semi-continuous in this topology. One can show that for any $\pi \in \Pi_{t+1}$, $Z(\omega) \mapsto Z(\omega) \cdot \text{CVAR}_{\alpha Z(\omega)}^\pi(C_{t+1:T} | x_{t+1} = f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)))$ is Lipschitz continuous on $\mathbb{R}$ for almost every $\omega \in \Omega$. It follows (after a few steps) that $Z(\omega) \mapsto Z(\omega) \cdot J_{t+1}^*(f_t(\mathbf{x}, \mathbf{u}, d_t(\omega)), \alpha Z(\omega))$ is Lipschitz continuous on $\mathbb{R}$ for almost every $\omega \in \Omega$, by using the definition of J_{t+1}^* as an infimum over the policy space Π_{t+1} . Then, take the expectation with respect to $\mathbb{P}$, and use $|\mathbb{E}(\cdot)| \leq \mathbb{E}(|\cdot|) := \|\cdot\|_1 \leq \|\cdot\|_2$ to complete the proof. $\square$