STOCHASTIC OPTIMAL CONTROL: THRESHOLD-AWARE POLICIES AND IMPACT OF RANDOM DISRUPTIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

MingYi Wang

August 2024

STOCHASTIC OPTIMAL CONTROL: THRESHOLD-AWARE POLICIES AND IMPACT OF RANDOM DISRUPTIONS

MingYi Wang, Ph.D.

Cornell University 2024

Stochastic optimal control theory encompasses various types of stochasticity and notions of optimality. The standard risk-neutral approach minimizes or maximizes an expected total cost, but this approach often yields non-robust results. In this thesis, we introduce a particular type of robust control framework of indefinite-horizon processes, maximizing the probability of desired outcomes while keeping the cumulative cost within a threshold.

For diffusive processes, our framework results in second-order parabolic Hamilton-Jacobi-Bellman (HJB) Partial Differential Equations (PDEs). We develop an efficient algorithm to solve these equations by leveraging the inherent causality of the framework. This allows us to recover the optimal "threshold (risk)-aware" feedback policies for all initial configurations and a range of threshold values simultaneously in a single sweep. We first apply this methodology to adaptive cancer therapy under stochastic cancer dynamics. In particular, we aim to maximize the probability of achieving treatment goals while keeping the total treatment cost within a specific cost threshold/budget.

We then extend this threshold-aware approach to hybrid control problems, specifically through sailboat routing under stochastically evolving wind conditions. This application involves solving a pair of quasi-variational inequalities in a Hamilton-Jacobi framework. Monte Carlo simulations are used to generate cumulative distribution functions (CDFs), demonstrating the advantages of threshold-aware policies over risk-neutral ones.

In the final section, we investigate bacterial competition influenced by environmental extreme events (dilutions). We propose an explanation for why toxin-sensitive bacteria, usually

outcompeted by toxin-producers *in vitro*, can thrive under frequent dilutions. We consider both deterministic periodic dilutions and randomly timed dilutions modeled by a Poisson process. Through a series of optimized toxin-regulation behaviors for toxin-producers, we demonstrate that toxin-sensitive strains still have a reasonable chance of winning. The numerical approach involves solving Hamilton-Jacobi-type equations (including a specific type of non-local coupling emerging from the jump-discontinuities induced by the Poisson process) using semi-Lagrangian schemes.

BIOGRAPHICAL SKETCH

MingYi Wang was born to parents Zhiyuan Wang and Hui Xu on November 12, 1993, in Shanghai, China. He developed a passion for mathematics during his time at the High School for Dual Languages and Asian Studies in New York City, from which he graduated in 2014. He then pursued an undergraduate degree at Rensselaer Polytechnic Institute in Troy, New York, earning a Bachelor of Science in Mathematics in December 2017. Subsequently, he joined the Center for Applied Mathematics at Cornell University, where he began his doctoral studies in applied mathematics in 2018 and earned a Master of Science in Applied Mathematics in 2021.

To my parents, for their sacrifices and love; to my friends and CAMsters, for their companionship and support; and to my advisor, Alex, for guiding me through this journey with wisdom and patience.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Alexander Vladimirsky, for his unwavering support, invaluable guidance, and continuous encouragement throughout this journey and the writing of this thesis. His mentorship has been a cornerstone of my academic development, and without his insights and patience, this thesis would not have been possible.

I am deeply thankful to my committee members, Anil Damle and Philippe Sosoe, for their constructive feedback and insightful guidance. I would also like to thank my unofficial committee member, Andrea Giometto, for his creative ideas and valuable suggestions, which have significantly shaped my last project.

To my family, and especially my mother, Hui Xu, I owe the deepest gratitude. Your love, patience, and encouragement have been my pillars of strength. Your unwavering belief in me has been a source of inspiration and motivation throughout this journey. Your sacrifices and constant support have given me the courage to pursue my dreams, and for that, I cannot thank you enough.

A special thanks to my academic family, particularly Marissa and Mallory, for their camaraderie, collaboration, and the countless discussions that enriched my research experience. The shared moments of struggle and success with you both and all the CAMsters have made this journey much more memorable. I sincerely cherish the support and friendship I have received in CAM.

This journey would not have been possible without my collaborators. Besides the aforementioned Alex and Andrea, I extend my gratitude to Jacob Scott for his numerous contributions to my first paper. Chapter 4 began as a summer REU project with Natasha Patnaik, Anne Somalwar, and Jingyi Wu. Your decision to continue working with me after the summer program made this chapter possible. The weekly discussions we had are some of my fondest memories. I also want to thank Alexander Anderson and Mark Robertson-Tessi

from the Integrated Mathematical Oncology Department at Moffitt Cancer Center for exploring various approaches to applying Optimal Control Theory to adaptive cancer therapies together with me during my internship.

I am grateful to all my friends at Moffitt Cancer Center; your encouragement made my internship far more worthwhile. I am also thankful to Roberto Ferretti and Lars Grüne for their advice on numerical methods used in Chapter 2. Additionally, I appreciate Mallory Gaspard, Cole Miles, and Elliot Cartee for their website templates that helped me create public web pages to show my projects to a broader audience.

Lastly, I am profoundly grateful for the funding provided by the National Science Foundation (NSF) Division of Mathematical Sciences (DMS) under Awards 1645643, 1738010, and 2111522, as well as the Air Force Office of Scientific Research under Award FA9550-22-1-0528. Their support provided the resources necessary to pursue my research goals.

This journey has been a collective effort. Thank you all for believing in me and for making this achievement possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

We start by considering traditional formulations of *deterministic* optimal control to achieve desired outcomes. A classical *finite-horizon* problem aims to attain specific goals, such as minimizing cumulative costs, by the predetermined end time, T. In a continuous-time setting, we often assume the process dynamics are modeled by an Ordinary Differential Equation (ODE) system (1.1)

$$\begin{cases} \dot{\boldsymbol{x}}(\tau) = \boldsymbol{f}(\boldsymbol{x}(\tau),\ \boldsymbol{a}(\tau)), \quad \tau \in [t, T]; \\ \boldsymbol{x}(t) = \boldsymbol{\xi} \in \mathbb{R}^n, \end{cases} \tag{1.1}$$

where $\boldsymbol{a} : \mathbb{R}_+ \to \mathcal{A}$ defines the measurable control function over a compact set $\mathcal{A}$ of all possible control values. We here assume the rate function $\boldsymbol{f}$ takes as inputs only the current state $\boldsymbol{x}(\tau)$ and the current control $\boldsymbol{a}(\tau)$. (It can certainly depend on the time τ as well but we omit here for a later comparison.) Suppose the controller seeks to minimize the cost functional

$$\mathcal{J}\Big(\boldsymbol{\xi}, t, \boldsymbol{a}(\cdot)\Big) = \int_t^T K(\boldsymbol{x}(\tau),\ \boldsymbol{a}(\tau))\, d\tau + \ g\Big(\boldsymbol{x}(T)\Big), \tag{1.2}$$

where K is some *running cost* and g is some *terminal cost* depending on the final state of the system. We can define a *value function*

$$u(\boldsymbol{\xi}, t) = \inf_{\boldsymbol{a}(\cdot)}\ \mathcal{J}\Big(\boldsymbol{\xi}, t, \boldsymbol{a}(\cdot)\Big) \tag{1.3}$$

as the minimal cost till termination. Via tools of dynamic programming [8, 61], one can show that u is the unique *viscosity solution* [39] to the following *time-dependent* Hamilton-Jacobi-Bellman (HJB) type Partial Differential Equation (PDE)

$$-\frac{\partial u}{\partial t}(\boldsymbol{\xi}, t) = \min_{\boldsymbol{a} \in \mathcal{A}} \Big\{ K(\boldsymbol{x}, \boldsymbol{a}) + \nabla u(\boldsymbol{\xi}, t) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{a}) \Big\} \tag{1.4}$$

with the terminal condition $u(\boldsymbol{\xi}, T) = g(\boldsymbol{\xi})$. It is important to note that by solving (1.4) for all $t \in [0, T]$, we effectively obtain solutions to a range of horizons simultaneously. That is, each $u(\boldsymbol{\xi}, t)$ serves as the initial time slice for a reduced horizon $T - t$.

However, the finite-horizon framework does not suit all problems. There exists a broad class of *indefinite-horizon* problems, named *exit-time* problems, where the process concludes once its state reaches a specific terminal or target set Δ. Under this framework, the terminal time (or exit time) now depends on the chosen policy $\boldsymbol{a}(\cdot)$

$$T\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big) = \inf\Big\{t > 0 \mid \boldsymbol{x}(t) \in \Delta,\ \boldsymbol{x}(0) = \boldsymbol{\xi},\ \text{following } \boldsymbol{a}(\cdot)\Big\}. \tag{1.5}$$

Different from (1.2), the controller now assesses the overall cost $\mathcal{J}$ by integrating the running cost K along the path from $\boldsymbol{\xi}$ to Δ and adding the terminal cost g depending on its final state. I.e,

$$\mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big) = \int_0^{T(\boldsymbol{\xi}, \boldsymbol{a}(\cdot))} K(\boldsymbol{x}(\tau),\ \boldsymbol{a}(\tau))\, d\tau + \ g\Big(\boldsymbol{x}(T)\Big). \tag{1.6}$$

The value function

$$u(\boldsymbol{\xi}) = \inf_{\boldsymbol{a}(\cdot)}\ \mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big) \tag{1.7}$$

is now a minimal cost-to-target and standard methods [8, 61] show that u is the unique viscosity solution [39] to the *time-independent* HJB PDE

$$0 = \min_{\boldsymbol{a} \in \mathcal{A}} \Big\{K(\boldsymbol{x}, \boldsymbol{a}) + \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{a})\Big\} \tag{1.8}$$

with the boundary condition $u = g$ on Δ.

Now consider a stochastic extension of the system (1.1) to a *drift-diffusion* process:

$$\begin{cases} \mathrm{d}\boldsymbol{X} = \boldsymbol{b}(\boldsymbol{X},\ \boldsymbol{a})\ \mathrm{d}\tau + \boldsymbol{\Sigma}(\boldsymbol{X}, \boldsymbol{a})\ \mathrm{d}\boldsymbol{W}, \\ \boldsymbol{X}(0) = \boldsymbol{\xi} \in \mathbb{R}^n, \end{cases} \tag{1.9}$$

where $\boldsymbol{\Sigma}$ is a diffusion coefficient function and $\boldsymbol{W}$ is a standard m-dimensional Brownian motion. Note that the system state $\boldsymbol{X}(\tau)$ now follows a Stochastic Differential Equation (SDE). As a result, the terminal time $T\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big)$ and the total cost $\mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big)$ become *random variables* even if their definitions remain the same. The standard (*risk-neutral*)

stochastic optimal control framework aims to minimize the expected total cost $\mathbb{E}[\mathcal{J}]$. The value function

$$w(\boldsymbol{\xi}) = \inf_{\boldsymbol{a}(\cdot)} \mathbb{E}\Big[\mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big)\Big] \tag{1.10}$$

defines the unique viscosity solution [62] to a second-order elliptic HJB PDE

$$\min_{\boldsymbol{a}\in\mathcal{A}} \left\{ K(\boldsymbol{\xi}, \boldsymbol{a}) + \nabla w(\boldsymbol{\xi}) \cdot \boldsymbol{b}(\boldsymbol{\xi}, \boldsymbol{a}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial \xi_i \partial \xi_j} w(\boldsymbol{\xi}) \boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{a})_{i,j} \right\} = 0, \tag{1.11}$$

where $\boldsymbol{B} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ and $w = g$ on Δ.

However, this approach becomes problematic when we require $g = +\infty$ on some subset of Δ (e.g., when entering a certain region of the domain is strictly prohibited or should be avoided as much as possible). Namely, it would result in $\mathbb{E}[\mathcal{J}] = +\infty$ due to a positive probability of entering that subset. Furthermore, even if the terminal cost always stay finite, the risk-neutral approach is indifferent to the level of variability in the distribution of $\mathcal{J}$. Consequently, this non-robustness may result in an impractical feedback policy if $\mathcal{J}$ has a right-heavy-tailed distribution. To address the above drawbacks, in this thesis we develop new tools for robust stochastic control of indefinite-horizon processes. In particular, we maximize the probability of desired outcomes while keeping $\mathcal{J}$ under any specific cost threshold (or "budget") $\bar{s}$

$$v(\boldsymbol{\xi}, \bar{s}) = \sup_{\boldsymbol{a}(\cdot)} \mathbb{P}\Big(\mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big) \leq \bar{s}\Big). \tag{1.12}$$

We show this novel stochastic optimal control formulation leads to a second-order parabolic PDE and use tools of dynamic programming to find "threshold-aware" (or *risk-aware*) optimal feedback policies. Our approach further enables an efficient algorithm to compute these policies for a range of threshold values simultaneously. The mathematical derivation of the HJB PDE that v has to satisfy if sufficiently smooth and the numerics are provided in Chapter 3.

In Chapter 2, we demonstrate the application of our threshold-aware approach to adaptive cancer therapies. In cancer applications, where stochastic cancer models are used, the standard approach is often ineffective because the cost of therapy failure is typically considered infinite. Thus, our goal is to select a strategy that maximize the probability of achieving therapy goals without exceeding a pre-established treatment cost threshold. We illustrate the broad applicability of our method using two distinct stochastic cancer models. The first model is based on Evolutionary Game Theory (EGT) [96], while the second model involves a drug-Sensitive/Resistant cells competition [30]. They are further tested against two different therapeutic goals: stabilizing the tumor and eradicating it. We show threshold-aware policies adjust the treatment plan based on the responsiveness of tumor to the drug dosage accumulated and the treatment duration already elapsed. Furthermore, we demonstrate the advantage of threshold-aware policies over deterministic-optimal policies by comparing their *empirical cumulative distribution functions* (ECDFs) through Monte Carlo simulations. The numerical implementation details specific to these two examples and additional numerical results are also provided in Chapter 3.

In Chapter 4, we extend the threshold-aware approach to *hybrid* control problems, illustrated by sailboat path-planning. This is a natural hybrid control scenario due to the combination of continuous steering and discrete tack-switching maneuvers, significantly influenced by stochastically evolving wind conditions. Previous studies primarily developed risk-neutral policies that aim to minimize the expected time of arrival [59, 29, 117]. However, such path-planing strategies can become impractical when there is a substantial risk of exceeding the expected time beyond an affordable level. Our method, therefore, focuses on maximizing the probability of reaching the target before a specified time deadline or threshold. While both Chapter 2 and Chapter 4 aim to maximize the probability of achieving desired outcomes, they differ significantly in their methodologies. In Chapter 4, we address the hybrid control problem by solving two quasi-variational inequalities based on second-

order HJB PDEs with degenerate parabolicity. We show that risk-awareness in sailing is particularly useful when a carefully calculated bet on the evolving wind direction might yield a reduction in the number of tack-switches.

We would also like to emphasize that the scope of stochastic optimal control problems extends beyond drift-diffusion processes. For instance, consider a framework for randomly-terminated events as another stochastic extension to the system (1.1). Suppose the process begins at $\boldsymbol{x}(0) = \boldsymbol{\xi}$ and follows the dynamics governed by (1.1). This process is subject to a random termination at time $\mathcal{T}$, where $\mathcal{T}$ follows an exponential distribution with rate $\lambda > 0$. Under these conditions, the expected cost-to-termination is defined as:

$$\mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big) = \int_0^{+\infty} \lambda \mathrm{e}^{-\lambda t} \left[\int_0^t K(\boldsymbol{x}(\tau),\ \boldsymbol{a}(\tau))\, d\tau +\ g\Big(\boldsymbol{x}(t)\Big)\right] \mathrm{d}t. \tag{1.13}$$

Thus, the standard expectation minimizing value function can be defined as

$$v(\boldsymbol{\xi}) = \inf_{a(\cdot)} \mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{a}(\cdot)\Big), \tag{1.14}$$

which can be shown to satisfy another first-order HJB PDE

$$0 = \lambda\Big(g(\boldsymbol{\xi}) - v(\boldsymbol{\xi})\Big) + \min_{\boldsymbol{a}\in\mathcal{A}} \Big\{K(\boldsymbol{x}, \boldsymbol{a}) + \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{a})\Big\}. \tag{1.15}$$

This idea is first introduced in [4] and further classified under "initial uncertainty" problems in [136]. However, one limitation of this approach is that it only seeks to optimize the problem up to the occurrence of the first random extreme event. In this thesis, we explore extensions of this idea by considering objectives involving potentially *infinitely many* random extreme events.

In Chapter 5, we explore a puzzle of bacterial competition observed in natural ecosystems such as animal guts, soil, and lakes. While toxin-producing bacteria typically outcompete toxin-sensitive bacteria in *in vitro* environments, the latter often coexist with or even outnumber the former in nature. We hypothesize that this discrepancy may be explained by the

frequent (re)occurrence of extreme events, such as flushing or dilution in the gut, which disrupt the anticipated outcomes of bacterial competition. To test this hypothesis, we consider a specific competition model and investigate the impact of both regular and randomly-timed dilutions, modeled by a Poisson process with rate λ. Using several control-theoretic models, we show that the toxin-producers cannot always win even if they develop *population sensing* and have evolved to take advantage of the known frequency of such extreme events. Mathematically, this involves solving a range of Hamilton-Jacobi-type equations with one of them leading to a *non-local coupling* due to Poisson distributed dilutions. To address this, we propose using an efficient Value-Policy Iterations (VPI) method [78] to compute the numerical solutions.

We acknowledge that there are various methods for incorporating random perturbations into deterministic systems. By employing stochastic optimal control theory, we can leverage our understanding of the specific types of stochasticity involved to develop diverse frameworks for a wide spectrum of applications. Throughout this thesis, we will explore the aforementioned intriguing applications with innovative stochastic optimal control formulations. We aim to propose efficient numerical schemes to solve a variety of Hamilton-Jacobi-type equations, thereby addressing both the mathematical and computational challenges inherent in these systems.

Chapter 2 is based on a paper co-authored with Jacob Scott and Alexander Vladimirsky, published in PLoS Computational Biology [169]. Additionally, Chapter 3 has been submitted as its Supplementary Materials accompanying this publication. Chapter 4 is based on a paper collaborated with Natasha Patnaik, Anne Somalwar, Jingyi Wu, and Alexander Vladimirsky [168] that has been accepted by American Control Conference (ACC) 2024. Finally, Chapter 5 is based on an ongoing project collaborated with Andrea Giometto and Alexander Vladimirsky. A preprint will soon to be submitted to PNAS.

CHAPTER 2

THRESHOLD-AWARENESS IN ADAPTIVE CANCER THERAPY

2.1 Introduction

Optimizing the schedule and composition of drug therapies for cancer patients is an important and active research area, with mathematical tools often employed to improve the outcomes and reduce the negative side effects. Tumor heterogeneity is increasingly viewed as a key aspect that can be leveraged to improve therapies through the use of optimal control theory [145]. Most researchers using this perspective focus on deterministic models of tumor evolution, with typical optimization objectives of maximizing the survival time [113], minimizing the tumor size [30], or minimizing the time until the tumor size is stabilized [31]. In models that address the stochasticity in tumor evolution, a typical optimization goal is to find treatment policies that maximize the likelihood of patient's eventual cure (e.g., [45, 36, 94]) or minimize the likelihood of negative events (e.g., metastasis) after specified time [60]. However, this ignores the need for more nuanced treatment policies that maximize the likelihood of different levels of success – e.g., the probability of reaching remission or tumor stabilization without exceeding the specified amount of drugs and/or the specified treatment duration. The primary goal of this chapter is to introduce a rigorous and computationally efficient approach that addresses such challenging objectives in stochastic cancer models.

The advent of personalized medicine in cancer has changed the way we think about therapy for patients whose tumors have actionable mutations. This has been a game changer for some patients, drastically increasing life spans, reducing toxicity and improving quality of life. Frustratingly, however, this population of patients is still small; it was estimated in 2020

that only $\approx 5\%$ of patients benefit from these targeted therapies [82]. Further, despite the many advantages of personalized therapies, they rarely, if ever, lead to a complete cure since tumors develop resistance through the process of Darwinian evolution [71]. In response to this realization, a new approach called "evolutionary therapy" seeks to use the evolutionary dynamics of diseases to alter therapeutic schedules and drug choices. Through a combination of mathematical and experimental modeling, investigators have worked to understand a range of theoretical questions of practical importance. E.g., how does the emergence of resistance to one drug affect the sensitivity to another? Do heterogeneous (phenotypically or genotypically mixed) populations within tumors respond to drugs differently depending on their current state? The insights gained in these investigations have already led to progress in rational drug ordering/cycling for bacterial infections [89, 122, 123, 111] as well as for a number of cancers [179, 46]. In the study of therapeutic scheduling, adaptive therapy, which uses mathematical tools from Evolutionary Game Theory (EGT), has shown promise not only in theory [68], but also in a phase 2 trial for men with metastatic prostate cancer [177]. Experimentally, there have been confirmations of EGT principles *in vivo*[51] as well as more quantitatively focused assay development *in vitro* [95], and observations of game interactions using these methods [55]. There are also many other models capturing the competition within heterogeneous tumors without using game-theoretic derivations; e.g., [30, 35, 103, 76]. The majority of theoretical work in this space has focused on optimization of different drug regimens for *deterministic* models of cancer evolution [41, 174, 173, 40, 73]. In contrast, our goal here is to provide efficient computational tools for nuanced therapeutic scheduling in cancer models that directly account for stochastic perturbations.

Cancers (and other populations of living things) are comprised of individual cells (or organisms) with their own behaviours and evolutionary histories. Stochastic phenomena are ubiquitous in their interactions and life histories. These include individual genetic differences, fate transitions [79], varying reactions to drugs [102], differences in signalling, and

small-scale variations in the tumor environment. Many of these are instances of *demographic stochasticity* [106], which often can be "averaged-out" when dealing with a sufficiently large population. Indeed, this notion is crucial for any description of tumor heterogeneity through splitting the cells into sub-populations. Such splitting is natural if the mutation-selection balance is tuned so that only closely related genotypes, encoding the same phenotype, will stably exist. These groups are also referred to as quasispecies [175, 107] and exist as distributions around a central genotype, with all cells in the group behaving in a similar manner despite random birth/death events [50, 106] and small within-the-group genetic heterogeneities [90]. In contrast, our focus here is on *environmental stochasticity*, which cannot be ignored even in large populations since it describes random events that simultaneously affect the entire groups. Such perturbations are typically external [50, 106]; e.g., for cancer they might result from therapy-unrelated drugs or from frequent small changes in the host's physiology. Of course, any such event will also cause varying responses of individual cells within each subpopulation; so, our use of the term "environmental stochasticity" should be interpreted as direct modeling of subpopulation-averaged responses to such system-wide perturbations.

Modeling such perturbations in continuous-time usually results in *Stochastic Differential Equations* (SDEs) [25, 2], whose behavior can be optimized using Stochastic Optimal Control Theory [61]. The latter provides a mathematical framework for handling sequential decision making (e.g., how much drug to administer at each point in time) under random perturbations (e.g., stochastic changes in respective fitness of competing subpopulations of cancer cells). Any fixed treatment strategy will result in a random tumor-evolutionary trajectory and a random cumulative "cost" (e.g., cumulative amount of drugs used, or time to recovery, or a combination of these two metrics). The key idea of *Dynamic Programming* (DP) is to pose equations for the cumulative cost of the optimal strategy and to recover that strategy *in feedback form*: i.e., decisions about the dose and duration of therapy are frequently re-evaluated based on the current state of the tumor instead of selecting a fixed

time-dependent treatment schedule in advance. This idea is applicable across a wide range of cancer models and therapy types, including those intended to stabilize the tumor and those aiming to eradicate it. We follow this approach here, but with an important caveat: instead of selecting an *on-average optimal* strategy (e.g., the one which minimizes the expected cost of treatment) as would be usual in stochastic DP, we select a strategy maximizing the probability of some desirable outcome (e.g., reaching the goals of the therapy *without exceeding a specific cost threshold*). The resulting risk-aware (or, more precisely, "threshold-aware") policies are designed to be adaptive, adjusting the treatment plan along the way based on the responsiveness of tumor to drugs already used (and the cost already incurred) so far. In contrast to standard methods of constrained stochastic optimal control, our approach makes it easy to compute such threshold-aware policies for a range of thresholds simultaneously.

As is often the case, there remains a significant gap between simplified mathematical models and clinical applications. Much work remains in refining and calibrating EGT models, and also in measuring different aspects of biological stochasticity. But once high-fidelity personalized models become available, our general approach could potentially be used by clinicians to choose the most suitable threshold after a detailed discussion of a specific patient's goals (to include the trade-offs between toxicity and quality of life, for example).

2.2 Methods and Models

To emphasize the broad applicability of our "risk-aware" adaptive therapy optimization approach, we first describe it for a fairly generic cancer model. Two specific examples are then studied in detail in §2.2.2 and §2.2.3.

2.2.1 Traditional and risk-aware control in drug therapy optimization

We note that most of the literature on dynamic programming in cancer models starts with positing a specific known/fixed treatment horizon T, with the success or failure of therapy assessed after that time (or earlier, in case of the modeled patient's death). This makes it easier to use the standard equations and algorithms of "finite-horizon" optimal control theory. But such a pre-determined T is not well-aligned with the notion of adaptive therapies. Instead, we adopt the *indefinite-horizon* framework, in which the process terminates as soon as the tumor's state satisfies some predefined conditions, with the terminal time T thus dependent on the chosen treatment policy. We use this framework in all of the control approaches described below, even though many of them have direct finite-horizon analogs as well.

We begin by describing several "traditional" optimal control formulations, followed by the threshold-aware version, which addresses some of their shortcomings in cancer applications. Starting with the deterministic setting summarized in Box 1, we use $\boldsymbol{x}(t) \in \mathbb{R}^n$ to encode the time-dependent state of a tumor (e.g., this could be the size or the relative abundance of n different sub-types of cancer cells). Tumor dynamics are modeled by an ODE system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{d})$, where the rate function $\boldsymbol{f}$ takes as inputs both the current state $\boldsymbol{x}(t)$ and the current control, the "therapy intensity" $\boldsymbol{d}(t)$. In models with a single drug, this is just a scalar $d(t) \in [0, d_{\max}]$ indicating the current rate of that drug's delivery, where $d_{\max}$ encodes the Maximum Tolerated Dose (MTD), which can in principle be patient-specific. But the same framework can also be used for multiple drugs, with a separate upper bound specified for each element of $\boldsymbol{d}(t)$. Given an initial tumor configuration $\boldsymbol{x}(0) = \boldsymbol{\xi}$, a successful therapy aims to drive the tumor state to a set Δ_{succ} while ensuring that the set Δ_{fail} is avoided. E.g.,

Box 1: Problem setup of a typical deterministic optimal cancer-control problem

Evolutionary dynamics with control on therapy intensity:

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{d}), \\ \boldsymbol{x}(0) = \boldsymbol{\xi}. \end{cases} \tag{2.1}$$

Process *terminates* as soon as either $\begin{cases} \boldsymbol{x}(t) \in \Delta_{\text{succ}}, & \text{if therapy succeeds;} \\ \boldsymbol{x}(t) \in \Delta_{\text{fail}}, & \text{if therapy fails.} \end{cases}$

Definitions and Parameters:

- $\boldsymbol{x} \in \mathbb{R}^n$, n-dimensional cancer state;
- $\boldsymbol{d} : \mathbb{R}_+ \to \mathcal{D}$ ($\mathcal{D}$ compact), time-dependent intensity of the therapy (control);
- $\Delta_{\text{succ}} \subset \mathbb{R}^n$, success region;
- $\Delta_{\text{fail}} \subset \mathbb{R}^n$, failure region;
- $\Delta = \Delta_{\text{succ}} \cup \Delta_{\text{fail}}$, terminal set.

Total treatment time: $T(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)) = \inf\Big\{ t \in \mathbb{R}_+ \;\big|\; \boldsymbol{x}(t) \in \Delta,\ \boldsymbol{x}(0) = \boldsymbol{\xi} \Big\}.$ (2.2)

Treatment cost function: $\mathcal{J}(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)) = \int_0^T K(\boldsymbol{x}(\tau), \boldsymbol{d}(\tau))\, \mathrm{d}\tau + g\Big(\boldsymbol{x}\,(T)\Big),$ (2.3)

where $T := T(\boldsymbol{\xi}, \boldsymbol{d}(\cdot))$ is the terminal time, $K(\boldsymbol{x}, \boldsymbol{d})$ is the running cost, and the terminal cost is

$$g(\boldsymbol{x}) = \begin{cases} +\infty, & \text{if } \boldsymbol{x}(T) \in \Delta_{\text{fail}}, \\ 0, & \text{if } \boldsymbol{x}(T) \in \Delta_{\text{succ}}. \end{cases}$$

Value function: $u(\boldsymbol{\xi}) = \inf_{\boldsymbol{d}(\cdot)} \mathcal{J}(\boldsymbol{\xi}, \boldsymbol{d}(\cdot))$

is found by numerically solving a first-order HJB PDE $\min_{\boldsymbol{d} \in \mathcal{D}} \Big\{ K(\boldsymbol{\xi}, \boldsymbol{d}) + \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{d}) \Big\} = 0,$ (2.4)

with the boundary condition $u = g$ on Δ. See §3.3.1 for the derivation and §3.4.2 for the numerics.

in eradication therapy models, Δ_{succ} might correspond to tumors below the detection level, while Δ_{fail} might specify a much larger size that effectively kills a patient; see §2.2.3. On the other hand, for models that only track the relative abundance of cancer subpopulations, Δ_{succ} might be defined in terms of the desired low abundance of specific subpopulations affected by $\boldsymbol{d}(t)$, with the idea that the tumor size stabilizes or an entirely different therapy strategy is adopted after $\boldsymbol{x}(t)$ enters Δ_{succ}; see §2.2.2.

If the therapy manages to reach Δ_{succ} while avoiding Δ_{fail}, its overall "cost" $\mathcal{J}$ is assessed by integrating some running cost $K = K(\boldsymbol{x}, \boldsymbol{d})$ along the "trajectory" from $\boldsymbol{\xi}$ to $\Delta = \Delta_{\text{succ}} \cup \Delta_{\text{fail}}$ and adding the "terminal cost" g depending on its final state. E.g., g might be defined as $+\infty$ on Δ_{fail} to make such outcomes unacceptable. The running cost K depends on the current state of the tumor and the current drug usage levels and can be used to model the direct impact on the patient of the tumor size and composition as well as the side effects of the therapy. The key idea of dynamic programming [8] is to define a *value function* $u(\boldsymbol{\xi})$ encoding the minimal overall cost for each specific initial tumor state and to show that this u must satisfy a stationary Hamilton-Jacobi-Bellman (HJB) equation (2.4). Once that partial differential equation (PDE) is solved numerically, the globally optimal rate of treatment can be obtained in *feedback form* for all cancer states (i.e., $\boldsymbol{d} = \boldsymbol{d}_\star(\boldsymbol{\xi})$), which makes it suitable for the adaptive therapy framework. Throughout this chapter, we will use $\boldsymbol{d}_\star(\boldsymbol{\xi})$ to denote an optimal feedback policy for the deterministic version of each control problem. If K is chosen so that the overall cost of a successful therapy $\mathcal{J}$ reflects a weighted sum of the total therapy duration and the cumulative use of each drug, the weights can be adjusted to reflect the relative importance of these optimization criteria. In this case, if $\boldsymbol{f}$ is also a linear function of $\boldsymbol{d}$, it is easy to show that the optimal treatment policy $\boldsymbol{d}_\star(\boldsymbol{\xi})$ is generally *bang-bang*; i.e., for each drug, it prescribes either no usage or the maximal (MTD) usage in every cancer state $\boldsymbol{\xi}$.

In a generic continuous-time stochastic cancer model (summarized in Box 2), the tumor state $\boldsymbol{X}(t)$ becomes a random variable, with the dynamics specified by a Stochastic Differential Equation (SDE) (2.5), which replaces the deterministic Ordinary Differential Equation (ODE) (2.1). The definitions of the *total treatment time* $T(\boldsymbol{\xi}, \boldsymbol{d}(\cdot))$ and the *overall treatment cost* $\mathcal{J}(\boldsymbol{\xi}, \boldsymbol{d}(\cdot))$ remain the same, but both of them become random variables. The standard (*risk-neutral*) approach of stochastic optimal control is to find a feedback-form treatment policy that minimizes the expected treatment cost $\mathbb{E}\Big[\mathcal{J}\Big]$. As explained in Box 3, the resulting

Box 2: Problem setup of the stochastic optimal control problem

Stochastic evolution dynamics with control on therapy intensity (a *drift-diffusion* process):

$$\begin{cases} \mathrm{d}\boldsymbol{X} = \boldsymbol{a}(\boldsymbol{X}, \boldsymbol{d})\,\mathrm{d}t + \boldsymbol{\Sigma}(\boldsymbol{X}, \boldsymbol{d})\,\mathrm{d}\boldsymbol{W}, \\ \boldsymbol{X}(0) = \boldsymbol{\xi}. \end{cases} \tag{2.5}$$

Process *terminates* as soon as either

$$\begin{cases} \boldsymbol{X}(t) \in \Delta_{\text{succ}}, & \text{if therapy succeeds;} \\ \boldsymbol{X}(t) \in \Delta_{\text{fail}}, & \text{if therapy fails.} \end{cases}$$

Definitions and Parameters:

- $\boldsymbol{X} \in \mathbb{R}^n$, n-dimensional cancer state;
- $\boldsymbol{W}$, standard m-dimensional Brownian motion;
- $\boldsymbol{a}(\boldsymbol{X}, \boldsymbol{d}) \in \mathbb{R}^n$, the drift function;
- $\boldsymbol{\Sigma}(\boldsymbol{X}, \boldsymbol{d}) \in \mathbb{R}^{n\times m}$, the diffusion function.

Note: Definitions of the *total treatment time* $T := T(\boldsymbol{\xi}, \boldsymbol{d}(\cdot))$ and the *treatment cost function* $\mathcal{J}(\boldsymbol{\xi}, \boldsymbol{d}(\cdot))$ stay the same as in Box 1. But they are now *random variables* as we will replace $\boldsymbol{x}(t)$ by $\boldsymbol{X}(t)$.

Box 3: Standard stochastic dynamic programming approaches

A risk-neutral (expectation-minimizing) approach [62] :

Value function: $w(\boldsymbol{\xi}) = \inf_{\boldsymbol{d}(\cdot)} \mathbb{E}\Big[\mathcal{J}\big(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)\big)\Big]$ (2.6)

can be found by solving a second-order Hamilton-Jacobi-Bellman (HJB) equation:

$$\min_{\boldsymbol{d}\in\mathcal{D}} \left\{ K(\boldsymbol{\xi}, \boldsymbol{d}) + \nabla w(\boldsymbol{\xi}) \cdot \boldsymbol{a}(\boldsymbol{\xi}, \boldsymbol{d}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial \xi_i \partial \xi_j} w(\boldsymbol{\xi}) \boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{d})_{i,j} \right\} = 0, \tag{2.7}$$

where $\boldsymbol{B} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ and $w = g$ on $\Delta = \Delta_{\text{succ}} \cup \Delta_{\text{fail}}$.

Note: If one uses $g(\boldsymbol{X}(T)) = +\infty$ when therapy fails (i.e., when $\boldsymbol{X}(T) \in \Delta_{\text{fail}}$), the diffusion will generally result in $w = +\infty$ for most if not all initial tumor configurations outside of Δ_{succ}.

An alternative is to maximize the probability of eventual goal attainment:

Value function: $\tilde{w}(\boldsymbol{\xi}) = \sup_{\boldsymbol{d}(\cdot)} \mathbb{P}\Big(\boldsymbol{X}\big(T\big) \in \Delta_{\text{succ}}\Big)$ (2.8)

can be found by solving a second-order Hamilton-Jacobi-Bellman (HJB) equation:

$$\max_{\boldsymbol{d}\in\mathcal{D}} \left\{ \nabla \tilde{w}(\boldsymbol{\xi}) \cdot \boldsymbol{a}(\boldsymbol{\xi}, \boldsymbol{d}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial \xi_i \partial \xi_j} \tilde{w}(\boldsymbol{\xi}) \boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{d})_{i,j} \right\} = 0, \tag{2.9}$$

with the boundary condition $\tilde{w} = 1$ on Δ_{succ} and $\tilde{w} = 0$ on Δ_{fail}.

value function satisfies another stationary HJB PDE (2.7). The choice of suitable boundary conditions is more subtle here: setting $g = +\infty$ on Δ_{fail} is no longer an option since the probability of entering Δ_{fail} before Δ_{succ} is usually positive under every treatment policy, which would result in $\mathcal{J} = +\infty$ for every occasional failure and the overall $\mathbb{E}\big[\mathcal{J}\big] = +\infty$. This makes it necessary to either choose a specific finite "cost" of failure (which can be problematic both for practical and ethical reasons) or switch to an entirely different optimization objective. For example, one can try to simply maximize the probability of fulfilling the therapy goals (i.e., eventually reaching Δ_{succ} while avoiding Δ_{fail}) by solving the equation (2.9). But this latter formulation ignores many important practical considerations: e.g., it can easily result in an unreasonably long treatment time or in significant side effects from a prolonged MTD-level drug administration.

In contrast, the approach we are pursuing here allows for a more nuanced definition of success (e.g., taking into account the total drug usage, the treatment duration, and/or the cumulative burden from the tumor). Choosing a running cost K to reflect the above factors, we define the overall cost as $\mathcal{J}(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)) = \int_0^T K(\boldsymbol{X}(\tau), \boldsymbol{d}(\tau))\,\mathrm{d}\tau + g\left(\boldsymbol{X}(T)\right)$, which might be infinite if $\boldsymbol{X}(T) \in \Delta_{\text{fail}}$. We then maximize the probability of reaching the policy goals, but constraining the overall cost by some pre-specified threshold $\bar{s}$. I.e., we need to find an adaptive therapy that maximizes $\mathbb{P}\left(\mathcal{J} \leq \bar{s}\right)$. Our goal is to compute such *threshold-aware* policies efficiently for all starting tumor configurations $\boldsymbol{\xi}$ and a broad range of threshold levels simultaneously. It is easy to see that here good treatment policies will have to also take into account the cost accumulated so far. This makes it natural to treat our chosen threshold $\bar{s}$ as an *initial cost budget*, tracking the remaining budget $s(t)$ by solving equation (2.13) in Box 4. The value function can be found by solving the parabolic PDE (2.11) numerically, and the optimal feedback policy $\boldsymbol{d}_*^{\bar{s}}(\boldsymbol{\xi}, s)$ is recovered in the process for all $\bar{s} \in (0, \bar{\mathcal{S}}]$.

We note that, in classical stochastic optimal control, parabolic HJB equations are usually

encountered when dealing with finite horizon problems, where the terminal time T is specified in advance. See [60, 27, 125] for typical examples in cancer-related literature. In contrast, the parabolicity in PDE (2.11) arises because of the monotone decrease in the remaining budget $s(t)$.

The details of our numerical method based on Box 4 are included in Chapter 3 §3.4.1. In the interest of computational reproducibility, we provide the source code for approximating value functions and computing threshold-aware policies for all the examples from §2.3 at `https://github.com/eikonal-equation/Stochastic-Cancer`

Box 4: Our threshold-aware approach

Value function: $$v(\boldsymbol{\xi}, \bar{s}) = \sup_{\boldsymbol{d}(\cdot)} \mathbb{P}\Big(\mathcal{J}\big(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)\big) \le \bar{s}\Big) \tag{2.10}$$

can be found by solving a different second-order HJB equation:

$$\max_{\boldsymbol{d}\in\mathcal{D}} \left\{ -\frac{\partial}{\partial s} v(\boldsymbol{\xi}, s) K(\boldsymbol{\xi}, \boldsymbol{d}) + \nabla v(\boldsymbol{\xi}, s) \cdot \boldsymbol{a}(\boldsymbol{\xi}, \boldsymbol{d}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial \xi_i \partial \xi_j} v(\boldsymbol{\xi}, s) \boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{d})_{i,j} \right\} = 0, \tag{2.11}$$

where $s \in [0, \bar{\mathcal{S}}]$. See the detailed derivation in §3.3.2 of Chapter 3.

The boundary conditions of HJB equation:

$$\begin{cases} v(\boldsymbol{\xi}, s) = 1, & \text{if } \boldsymbol{\xi} \in \Delta_{\text{succ}},\ s \in [0, \bar{\mathcal{S}}]; \\ v(\boldsymbol{\xi}, s) = 0, & \text{if } \boldsymbol{\xi} \in \Delta_{\text{fail}},\ s \in [0, \bar{\mathcal{S}}]; \\ v(\boldsymbol{\xi}, 0) = 0, & \text{if } \boldsymbol{\xi} \notin \Delta_{\text{succ}}. \end{cases} \tag{2.12}$$

The (random) ODE describing the reduction of budget:

$$\dot{s} = -K(\boldsymbol{X}(t), \boldsymbol{d}(t)), \qquad s(0) = \bar{s} \in (0, \bar{\mathcal{S}}]. \tag{2.13}$$

While this threshold-aware framework has important advantages illustrated below, it also brings to the forefront several subtleties avoided in the more traditional stochastic optimal control approaches. First, an adaptive treatment policy optimal for one specific threshold is usually not optimal for another. (The starting budget in (2.13) is important for deciding when to administer drugs.) This would make it necessary for a practitioner to have a detailed discussion with their patient to choose a suitable threshold value before the treatment is started. Second, stochastic perturbations make the outcome random, and the budget might

run out under any treatment policy; i.e., we might see $s(t_\sharp) = 0$ at some random time $t_\sharp$. But this scenario is only a failure in the sense that the overall cost $\mathcal{J}$ will now definitely exceed the threshold value $\bar{s}$. If the patient is still alive ($\boldsymbol{X}(\tau) \notin \Delta_{\text{fail}},\ \forall \tau \in [0, t_\sharp]$) and interested in continuing treatment, one has to make a decision on the new strategy. This can be done either by posing a new threshold for future treatment costs or by switching to an entirely different policy – e.g., either by employing some traditional stochastic optimal control approach (based on equations (2.7) or (2.9)) or by using a deterministic-optimal policy based on equation (2.4). The latter version is used in all stochastic simulations in the following sections.

Informally, threshold-aware policies reflect a tension between two objectives which are often (but not always) in conflict. Maximizing the probability of treatment attaining its primary goals (e.g., tumor stabilization or eradication) is balanced against reducing the cost (a combination of tumor and treatment burdens) suffered along the way. The former is optimized but only over the scenarios where the latter stays below the prescribed threshold. We close this subsection by highlighting connections of our approach to general multiobjective optimal control and optimal control with integral constraints.

In deterministic optimal control theory, the idea of treating some version of cumulative cost as an additional state variable is well-known. But the resulting ODE systems are typically treated within the framework of *Pontryagin's Maximum Principle* (PMP) [135], which has an important advantage (its suitability for high-dimensional problems) but also a number of serious drawbacks: the fact that policies are not recovered in feedback form, the fact that these policies are generally not guaranteed to be *globally* optimal, occasional difficulties in ensuring the convergence of numerical methods needed to find such policies, and challenges in handling non-trivial state constraints. In cancer literature, this PMP-based approach has been used to impose "isoperimetric constraints" on the amount of administered

chemotherapy [182] or immunotherapy [81]. In addition to the issues listed above, we note that the suitability of equality (isoperimetric) constraints is not obvious in many cancer applications. Indeed, the fact that a less aggressive treatment may in some cases improve the outcomes is one of the main reasons for the interest in adaptive therapies. Thus, insisting that all available drugs must be used is hard to justify, and inequality constraints (e.g., imposing an upper bound on the cumulative drug use) seem much more reasonable.

The first dynamic programming (HJB-based) formulation for handling such constraints in general deterministic control problems was developed in [101]. It circumvents all these PMP-associated difficulties with an added benefit of finding globally optimal policies for a range of inequality constraint levels simultaneously. The threshold-aware method presented here extends many of the same ideas to a stochastic setting.

2.2.2 Example 1: an EGT-based competition model.

To develop our first example, we adopt the base model of cancer evolution proposed by Kaznatcheev et al. in [96], which describes a competition of 3 types of cancer cells. Glycolytic cells (GLY) are anaerobic and produce lactic acid, which damages the surrounding non-cancerous tissue. The other two types are aerobic and benefit from better vasculature, development of which is promoted by production of the VEGF signaling protein. Thus, the VEGF (over)-producing cells (VOP) devote some of their resources to vasculature development, while the remaining aerobic cells are essentially free-riders or *defectors* (DEF) in game-theoretic terminology. If $(z_{\mathrm{G}}(t), z_{\mathrm{D}}(t), z_{\mathrm{V}}(t))$ encode the time-dependent subpopulation sizes of these three cancer types, their dynamics are given by $\dot{z}_i = \psi_i z_i$, where $i \in \{G, D, V\}$ and $(\psi_{\mathrm{G}}, \psi_{\mathrm{D}}, \psi_{\mathrm{V}})$ are the respective type fitnesses. The actual expressions for these ψ_i are derived from the inter-population competition in the usual EGT framework; see Chapter 3

§3.1.1. This competition of cells in the tumor is modeled as a "public goods" / "club goods" game: VEGF is a "club good" since it benefits only VOP and DEF cells, while the acid generated by GLY is a "public good" since the damage to healthy tissue is assumed to benefit all cancer cells. The base model in [96] assumes that each cell interacts with n others nearby. How much it benefits from these interactions depends on its own type and the proportions of different cell types among those nearby cells. Assuming that all participants are drawn uniformly at random from a large well-mixed population, one can derive all fitnesses ψ_i as expected payoffs in this game of $(n+1)$ players. Those expected payoffs will naturally depend on the current subpopulation fractions (or relative abundances) $x_{\mathrm{G}} = \dfrac{z_{\mathrm{G}}}{z_{\mathrm{G}} + z_{\mathrm{D}} + z_{\mathrm{V}}}$, $x_{\mathrm{D}} = \dfrac{z_{\mathrm{D}}}{z_{\mathrm{G}} + z_{\mathrm{D}} + z_{\mathrm{V}}}$, and $x_{\mathrm{V}} = \dfrac{z_{\mathrm{V}}}{z_{\mathrm{G}} + z_{\mathrm{D}} + z_{\mathrm{V}}}$. A *Replicator* Ordinary Differential Equation (ODE) [156, 86] is a standard EGT model for predicting the changes in these subpopulation-fractions as a function of time.

In both the original deterministic case and its stochastic extension, it is easier to view the replicator equation as a 2-dimensional system (e.g., by noting that $x_{\mathrm{D}} = 1 - x_{\mathrm{G}} - x_{\mathrm{V}}$). Following [96], we use a slightly different reduction, rewriting everything in terms of the proportion of glycolytic cells in the tumor $p(t) = x_{\mathrm{G}}(t)$ and the proportion of VOP among aerobic cells $q(t) = \dfrac{x_{\mathrm{V}}(t)}{x_{\mathrm{V}}(t) + x_{\mathrm{D}}(t)}$. A drug therapy (in this example, affecting the fitness of GLY cells only) is similarly easy to encode by modifying the Replicator ODE; see equation (2.14) in Box 5 and the Supplementary Materials in [96] for the derivation. The goal of the drug therapy here is to drive the GLY fraction $p(t) = x_{\mathrm{G}}(t)$ down below a specified "stabilization barrier" γ_{r}. (In [96], this goal is justified by noting that, with GLY gone, the DEF cells will then quickly overcome VOP, leading to "an aerobic tumor with no - or significantly diminished - ability to recruit blood vessels," which stabilizes (or at least significantly slows down the growth of) the tumor.) For a range of parameter values, this model yields periodic behavior of cancer subpopulations: without drugs, $x_{\mathrm{G}}(t)$, $x_{\mathrm{D}}(t)$, and $x_{\mathrm{V}}(t)$ alternate in being dominant in the tumor, with the amplitude of oscillations determined

by the initial conditions [96]. This highlights the importance of proper timing in therapies: starting from the same initial tumor composition (q_0, p_0), the same MTD therapy of a fixed duration could lead to either a stabilization ($p(t)$ falling below γ_{r}) or a death ($p(t)$ rising above the specified "failure barrier" γ_{f}) depending on how long we wait until this therapy starts; see Fig 2 in Kaznatcheev et al. [96].

Box 5: Example 1 (an EGT-based competition model adopted from [96, 73])

The deterministic base model (components for the approach in Box 1):

$$\boldsymbol{x} = \begin{bmatrix} q \\ p \end{bmatrix} := \begin{bmatrix} \dfrac{x_{\mathrm{V}}}{x_{\mathrm{V}} + x_{\mathrm{D}}} \\ x_{\mathrm{G}} \end{bmatrix} \text{ and } \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, d) = \begin{bmatrix} q(t)\Big(1 - q(t)\Big)\Big(\dfrac{b_v}{n+1}\sum\limits_{k=0}^{n}(p(t))^k - c\Big) \\ p(t)\Big(1 - p(t)\Big)\Big(\dfrac{b_a}{n+1} - (b_v - c)q(t) - d(t)\Big) \end{bmatrix} \tag{2.14}$$

The above reflects the formulas for subpopulation fitnesses $(\psi_{\mathrm{G}}, \psi_{\mathrm{D}}, \psi_{\mathrm{V}})$; see details in Chapter 3 § 3.1.1.

The stochastic model (components for the approaches in Boxes 2-4):

$$\boldsymbol{X} = \begin{bmatrix} Q \\ P \end{bmatrix} := \begin{bmatrix} \dfrac{X_{\mathrm{V}}}{X_{\mathrm{V}} + X_{\mathrm{D}}} \\ X_{\mathrm{G}} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} W_{\mathrm{G}} \\ W_{\mathrm{D}} \\ W_{\mathrm{V}} \end{bmatrix},$$

$$\boldsymbol{a}(\boldsymbol{X}, d) = \begin{bmatrix} Q(1-Q)\left\{\left(\dfrac{b_v}{n+1}\left[\sum\limits_{k=0}^{n}P^k\right] - c\right) + \Big[(1-Q)\sigma_2^2 - Q\sigma_3^2\Big]\right\} \\ P(1-P)\left\{\left(\dfrac{b_a}{n+1} - (b_v - c)Q - d\right) - \Big[\sigma_1^2 P - \sigma_2^2(1-P)(1-Q)^2 - \sigma_3^2(1-P)Q^2\Big]\right\} \end{bmatrix},$$

$$\boldsymbol{\Sigma}(\boldsymbol{X}, d) = \begin{bmatrix} 0 & -\sigma_{\mathrm{D}}Q(1-Q) & \sigma_{\mathrm{V}}Q(1-Q) \\ \sigma_{\mathrm{G}}P(1-P) & \sigma_{\mathrm{D}}P(1-P)(1-Q) & \sigma_{\mathrm{V}}P(1-P)Q \end{bmatrix}. \tag{2.15}$$

Definitions and Parameters:

- $d : \mathbb{R}_+ \to [0, d_{\max}]$, time-dependent intensity of GLY-targeting therapy;
- $\Delta_{\mathrm{succ}} = \{(q,p) \in [0,1]^2 \mid p < \gamma_{\mathrm{r}}\}$, success region where γ_{r} is the *stabilization barrier*;
- $\Delta_{\mathrm{fail}} = \{(q,p) \in [0,1]^2 \mid p > \gamma_{\mathrm{f}}\}$, failure region where γ_{f} is the *failure barrier*;
- $K(\boldsymbol{X}, d) = d + \delta$, running cost function where δ is the treatment time penalty;
- $g = \begin{cases} +\infty, & \text{if } \boldsymbol{x}(T) \in \Delta_{\mathrm{fail}}, \\ 0, & \text{if } \boldsymbol{x}(T) \in \Delta_{\mathrm{succ}}, \end{cases}$ terminal cost;
- $\boldsymbol{W} = (W_{\mathrm{G}}, W_{\mathrm{D}}, W_{\mathrm{V}})$, standard 3D Brownian motion for (GLY, DEF, VOP) cells;
- $(\sigma_{\mathrm{G}}, \sigma_{\mathrm{D}}, \sigma_{\mathrm{V}})$, volatilities for (GLY, DEF, VOP) cells;
- b_a, the benefit per unit of acidification;
- b_v, the benefit from the oxygen per unit of vascularization;
- c, the cost of production of VEGF;
- $(n+1)$, the number of cells in the interaction group.

Conditions for the heterogeneous regime (coexistence of all cell types):

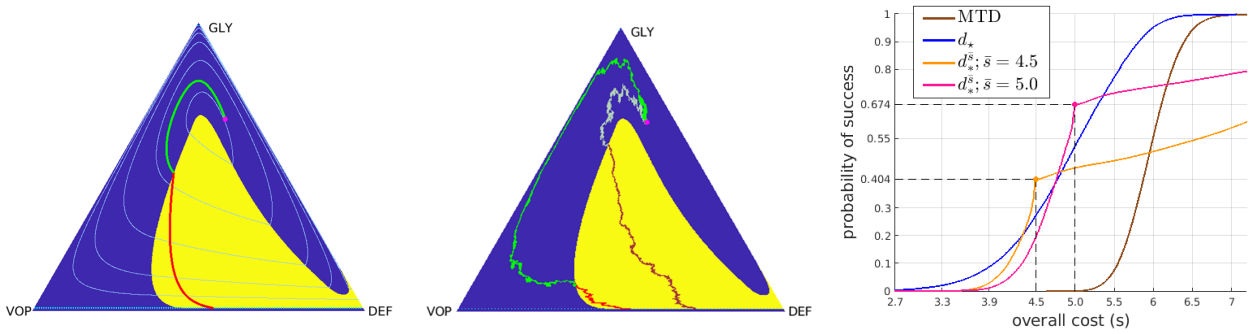
$$\frac{b_a}{n+1} < b_v - c < cn. \tag{2.16}$$

The optimal threshold-aware policy in feedback form:

$$d_*(q,p,s) = \begin{cases} d_{\max}, & \text{if } \left(\dfrac{\partial v}{\partial p}p(1-p) + \dfrac{\partial v}{\partial s}\right) < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

This strongly suggests the advantage of *adaptive therapies*, which prescribe the amount of drugs based on continuous or occasional monitoring of $(q(t), p(t))$ or some proxy (non-invasively measured) variables. A natural question is how to optimize such policies to reduce the total amount of drugs used and the total duration of treatment until $p(t) < \gamma_{\mathrm{r}}$. Gluzman et al. have addressed this in [73] using the framework of deterministic optimal control theory [61]. A time-dependent intensity of the therapy $d(t)$ (ranging from 0 to the MTD level $d_{\max}$) was chosen to minimize the overall cost of treatment $\mathcal{J}\left(q_0, p_0, d(\cdot)\right) = \int_0^T d(t)\, dt + \delta T + g\left(q(T), p(T)\right)$, where T is the time till stabilization (or failure, if $(q(T), p(T)) \in \Delta_{\text{fail}}$ and $g = +\infty$) while the value of $\delta > 0$ reflects the relative importance of two optimization goals (total drugs vs total time). In the framework of deterministic dynamic programming [8] summarized in Box 1, this corresponds to minimizing the integral of the running cost $K = d(t) + \delta$. In [73], the deterministic-optimal policy is obtained in *feedback form* (i.e., $d = d_{\bigstar}(q, p)$) by numerically solving the Hamilton-Jacobi-Bellman (HJB) PDE (2.4). As explained in §2.2.1, this policy is bang-bang. Fig 2.1(a) summarizes it (showing in yellow the MTD region where $d_{\bigstar}(q, p) = d_{\max}$) and illustrates the corresponding trajectory for one specific initial (q_0, p_0).

A natural way to introduce stochastic perturbations into this base model is to assume that the rates of subpopulation growth/decay are actually random and normally distributed at any instant, with the fitness functions $(\psi_{\mathrm{G}}, \psi_{\mathrm{D}}, \psi_{\mathrm{V}})$ encoding the expected values of those rates and the scale of random perturbations specified by $(\sigma_{\mathrm{G}}, \sigma_{\mathrm{D}}, \sigma_{\mathrm{V}})$. This approach, originating from Fudenberg and Harris paper [65], is suitable for modeling heterogeneous tumors, in which subpopulations not only interact [114] but can also vary in their growth rates over time [163]. Adopting the usual probabilistic notation of using capital letters for random variables, we can again start with the subpopulation sizes $(Z_{\mathrm{G}}, Z_{\mathrm{D}}, Z_{\mathrm{V}})$ evolving based on the *Stochastic Differential Equations* (SDEs) $dZ_i = (\psi_i dt + \sigma_i dW_i)Z_i$, where $i \in \{G, D, V\}$ and each W_i is a standard one-dimensional Brownian motion, modeling independent perturbations to



(a) **(b)** **(c)**

Figure 2.1. Deterministic-optimal policy in the EGT-model. The (GLY-VOP-DEF) triangle represents all possible relative abundances of respective subpopulations. Since the optimal policy is bang-bang, we show it by using the yellow background where drugs should be used at the MTD rate and the blue background where no drugs should be used at all. Starting from an initial state $(q_0, p_0) = (0.26, 0.665)$ (magenta dot), the subfigures show (a) the optimal trajectory found from the truly deterministically driven system (2.14) with cost 5.13; (b) two representative sample paths generated under the deterministic-optimal policy but subject to stochastic fitness perturbations (the brighter one incurs a total cost of 3.33, whereas the duller-colored path incurs a much higher 6.23); (c) CDFs of the cumulative cost $\mathcal{J}$ approximated using 10^5 random simulations. In both (a) & (b), *the green parts of trajectories* correspond to not prescribing drugs and *the red parts of trajectories* correspond to prescribing drugs at the MTD rate. In (a), the level sets of the value function in the deterministic case are shown in *light blue*. In (c), the *blue* curve is the CDF generated with the deterministic-optimal policy $d_\star$. Its observed median and mean conditioning on success are 4.95 and 4.91 respectively. The *brown* curve is the CDF generated with the MTD-based therapy, which in this example also maximizes the chances of "budget-unconstrained" tumor stabilization. Its observed median and mean conditioning on success are 5.95 and 5.96 respectively. *Orange* and *pink* curves show the CDFs for two different threshold-aware policies (with $\bar{s} = 4.5$ and $\bar{s} = 5$ respectively). The large dot on each of them represents the maximized probability of not exceeding the corresponding threshold. The term "threshold-specific advantage" refers to the fact that, at $\bar{s}$, the CDF of $d_*^{\bar{s}}$ is above the CDFs of all other policies.

the fitness of the respective subpopulation. This can be used to derive the SDEs for the corresponding fractions $(X_{\rm G}, X_{\rm D}, X_{\rm V})$. We note that similar Stochastic Replicator Equations arise naturally in ecology, where they have been studied in depth to address a possible coexistence of species in randomly perturbed environments [149, 84].

The summary of Replicator SDEs for the reduced $(Q,\ P)$ coordinates is provided in Box 5; the derivation can be found in Chapter 3 §3.2.2. The terminal set Δ is still the same: the process terminates as soon as $P(t)$ crosses a stabilization barrier (GLY's are low, leaving

mostly aerobic cells in the tumor) or the failure barrier (GLY's are high, the patient dies). But the terminal time T and the incurred cumulative cost $\mathcal{J}$ will also be random even if we fix the initial tumor configuration (q_0, p_0) and choose a specific treatment policy $d(\cdot)$. Fig 2.1(b) shows one example of using the deterministic-optimal policy $d(t) = d_\star(Q(t), P(t))$ in this stochastic setting. Gathering statistics from many random simulations that start from the same (q_0, p_0), we can approximate the *Cumulative Distribution Function* (CDF), measuring the probability of keeping $\mathcal{J}$ below any given threshold s if the deterministic-optimal policy is employed:

$$F_{d_\star}(s) \;=\; \mathbb{P}\left(\mathcal{J} \leq s\right),$$

whose graph is shown in blue in Fig 2.1(c). If one instead opts to solve the PDE (2.9) to maximize the probability of reaching Δ_{succ} while avoiding Δ_{fail}, this yields a simple MTD-policy $d = d_{\max}$, whose CDF (shown in brown in Fig 2.1(c)) is strictly worse than that of $d_\star$. This is not surprising since the more selective $d_\star$ is quite safe for this particular (q_0, p_0), with Δ_{fail} avoided in all of our 10^5 simulations. However, its resulting "cost" can be still high in many scenarios. E.g., in 47.4% of the $d_\star$-based simulations, $\mathcal{J}$ exceeded 5; in 72.6% of all cases it exceeded 4.5.

This motivates our optimization approach: deriving a *threshold-aware optimal policy* $d_*^{\bar{s}}$ to maximize the probability of stabilization without exceeding a specific cost threshold $\bar{s}$. As explained in §2.2.1 and summarized in Box 4, this is accomplished for a range of threshold values and all initial cancer configurations simultaneously. Fig 2.1(c) already shows that such policies can provide significant threshold-specific advantages over the deterministic-optimal therapy. Additional simulation results and the actual policies are illustrated in §2.3.1.

2.2.3 Example 2: a Sensitive-Resistant competition model.

We also illustrate our approach by extending a model proposed by Carrère [30], which focuses on the actual size of lung cancer cell populations studied *in vitro*. They consider a heterogeneous tumor that consists of two types of lung cancer cells: the sensitive (S) "A549" (sensitive to the drug "Epothilene") and the resistant (R) "A549 Epo40". This was based on the data from a series of experiments conducted by Manon Carrè at the Center for Research in Oncobiology and Oncopharmacology, Aix-Marseille Université. Mutation events were neglected due to their rarity at the considered dosages of Epothilene and due to relatively short treatment durations. The competition model presented below was derived based on phenotypical observations, with fluorescent marking used to trace and differentiate the cells.

Considered separately, both of these types obey a logistic growth model with respective intrinsic growth rates g_{S} and g_{R}. The carrying capacity of the Petri dish (C) is assumed to be shared, with the resistant cells assumed to be m times bigger than the sensitive; so, the fraction of space used at the time t is $\dfrac{z_{\mathrm{S}}(t) + m z_{\mathrm{R}}(t)}{C}$. When cultivated together, it was observed that the sensitive cells quickly outgrow the resistant ones despite the fact that their intrinsic growth rates are similar [30]. To model this competitive advantage, they have used an additional competition term $-\beta z_{\mathrm{S}} z_{\mathrm{R}}$ to describe the rate of change of $z_{\mathrm{R}}(t)$, with the coefficient β calibrated based on experimental data. It was further assumed that R cells are completely resistant to a specific drug, which reduces the population of S cells at the rate of $\alpha z_{\mathrm{S}}(t) d(t)$, with $d(t)$ reflecting the current rate of drug delivery and the constant coefficient α reflecting that drug's effectiveness. With a normalization $z_{\mathrm{S}}(t) \to z_{\mathrm{S}}(t)/C$, $z_{\mathrm{R}}(t) \to z_{\mathrm{R}}(t)/C$,

the resulting dynamics are summarized by

$$\begin{aligned}
\dot{z_{\mathrm{S}}}(t) &= g_{\mathrm{S}}\Big(1 - z_{\mathrm{S}}(t) - m z_{\mathrm{R}}(t)\Big) z_{\mathrm{S}}(t) - \alpha z_{\mathrm{S}}(t) d(t), \\
\dot{z_{\mathrm{R}}}(t) &= g_{\mathrm{R}}\Big(1 - z_{\mathrm{S}}(t) - m z_{\mathrm{R}}(t)\Big) z_{\mathrm{R}}(t) - \beta C z_{\mathrm{S}}(t) z_{\mathrm{R}}(t).
\end{aligned} \tag{2.18}$$

In both the original deterministic case and its stochastic extension, it is more convenient to restate the dynamics in terms of the *effective tumor size* $p(t) = z_{\mathrm{S}}(t) + m z_{\mathrm{R}}(t)$ and the fraction of effective tumor size comprised of the sensitive cells $q(t) = z_{\mathrm{S}}(t)/p(t)$. Note that the proportion of sensitive cells in the tumor, by number, is $\dfrac{mq(t)}{1 + (m-1)q(t)}$ instead of just $q(t)$ due to the size ratio m between S and R cells. This change of coordinates yields an ODE model (2.19) summarized in Box 6; see Chapter 3 §3.1.2 for the derivation. In this case, the goal of our adaptive therapy is eradication: i.e., driving the total tumor size $p(t)$ below some remission barrier γ_{r} (e.g., a physical detection level) while ensuring that throughout the treatment this $p(t)$ stays below a significantly higher failure barrier $\gamma_{\mathrm{f}} < 1$.

Box 6: Example 2 (a Sensitive-Resistant competition model adopted from [30])

The deterministic base model (components for the approach in Box 1):

$$\boldsymbol{x} = \begin{bmatrix} q \\ p \end{bmatrix} := \begin{bmatrix} \dfrac{z_{\mathrm{S}}}{z_{\mathrm{S}} + m z_{\mathrm{R}}} \\ z_{\mathrm{S}} + m z_{\mathrm{R}} \end{bmatrix}$$

$$\text{and} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, d) = \begin{bmatrix} (1-p)q(1-q)(g_{\mathrm{S}} - g_{\mathrm{R}}) + \beta C p^2 q^2 (1-q) - \alpha q (1-q) d(t) \\ p(1-p)[g_{\mathrm{S}} q + g_{\mathrm{R}}(1-q)] - \beta C p^2 q (1-q) - \alpha q p d(t) \end{bmatrix} \tag{2.19}$$

The stochastic model (components for the approaches in Boxes 2-4):

$$\boldsymbol{X} = \begin{bmatrix} Q \\ P \end{bmatrix} := \begin{bmatrix} \dfrac{Z_{\mathrm{S}}}{Z_{\mathrm{S}} + m Z_{\mathrm{R}}} \\ Z_{\mathrm{S}} + m Z_{\mathrm{R}} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} B_t \end{bmatrix},$$

$$\boldsymbol{a}(\boldsymbol{X}, d) = \begin{bmatrix} Q(1-Q)\Big\{(1-P)(g_{\mathrm{S}} - g_{\mathrm{R}}) - \alpha d + \beta C Q P + (1-P)^2[\sigma_{\mathrm{R}}^2(1-Q) - \sigma_{\mathrm{S}}^2 Q + \sigma_{\mathrm{S}}\sigma_{\mathrm{R}}]\Big\} \\ P(1-P)(g_{\mathrm{S}} Q + g_{\mathrm{R}}(1-Q)) - \alpha Q P d - \beta C P^2 Q (1-Q) \end{bmatrix},$$

$$\boldsymbol{\Sigma}(\boldsymbol{X}, d) = \begin{bmatrix} (1-P)Q(1-Q)(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}}) \\ P(1-P)[\sigma_{\mathrm{S}} Q + \sigma_{\mathrm{R}}(1-Q)] \end{bmatrix}. \tag{2.20}$$

Definitions and Parameters:

- $d : \mathbb{R}_+ \to [0, d_{\max}]$, time-dependent intensity of S-targeting therapy;
- $\Delta_{\mathrm{succ}} = \{(q,p) \in [0,1]^2 \mid p < \gamma_{\mathrm{r}}\}$, success region where γ_{r} is the *remission barrier*;
- $\Delta_{\mathrm{fail}} = \{(q,p) \in [0,1]^2 \mid p > \gamma_{\mathrm{f}}\}$, failure region where γ_{f} is the *failure barrier*;
- $K(\boldsymbol{X}, d) = d + \delta$, running cost function where δ is the treatment time penalty;
- $g = \begin{cases} +\infty, & \text{if } \boldsymbol{x}(T) \in \Delta_{\mathrm{fail}}, \\ 0, & \text{if } \boldsymbol{x}(T) \in \Delta_{\mathrm{succ}}, \end{cases}$ terminal cost;
- $(g_{\mathrm{S}}, g_{\mathrm{R}})$, growth rate for the sensitive and resistant cells, respectively;
- B_t, standard 1D Brownian motion;
- $(\sigma_{\mathrm{S}}, \sigma_{\mathrm{R}})$, volatilities for the sensitive and resistant cells, respectively;
- m, size ratio between S and R cells;
- C, Petri dish carrying capacity;
- α, drug efficiency;
- β, action of sensitive on resistant.

Parameter values are specified in Chapter 3 §3.5.2.

The optimal threshold-aware policy in feedback form:

$$d_*(q,p,s) = \begin{cases} d_{\max}, & \text{if } \left(\dfrac{\partial v}{\partial q}\alpha q(1-q) + \dfrac{\partial v}{\partial p}\alpha q p + \dfrac{\partial v}{\partial s}\right) < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.21}$$

Fig 2.2(a) illustrates the natural dynamics of this model with no drug use. In this case, the competitive pressure reduces the population R, which at first decreases the tumor size for many initial conditions. But a rapid growth in S eventually increases the overall tumor, leading to an inevitable failure ($p(t) > \gamma_{\rm f}$). The deterministic-optimal drug therapy is again sought to minimize a weighted sum of total drugs used and the time of treatment (with the running cost $K = d(t) + \delta$) until the eradication. It is obtained in feedback form $d = d_\star(q, p)$ after solving the PDE (2.4). Fig 2.2(b) shows that, for smaller tumor sizes, this $d_\star$ prescribes MTD-level treatment only after this initial tumor reduction is over, once S gets rid of most R cells which are not sensitive to Epothilene. However, for larger initial p, this deterministic-optimal policy starts using the drugs much earlier, planning to keep S cells in check as soon as they are numerous enough to control R.

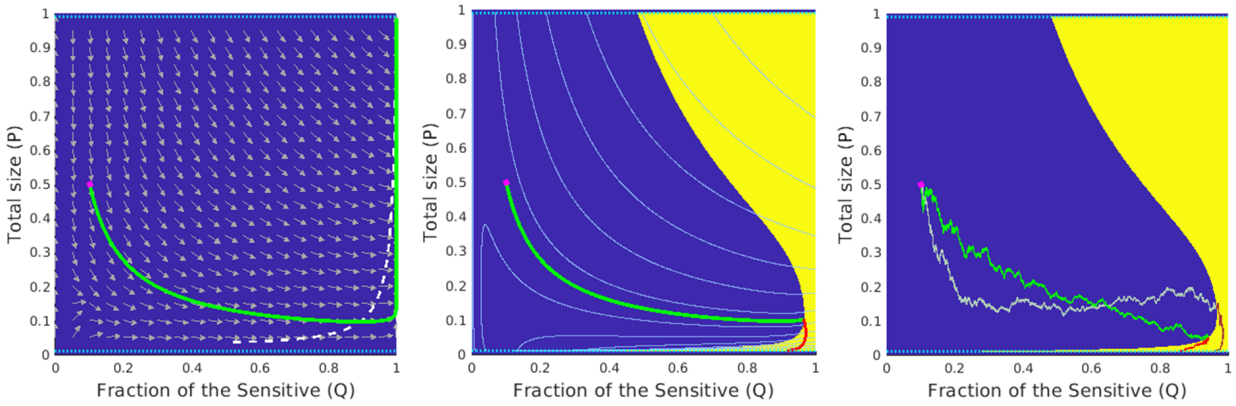


Figure 2.2. Deterministic-optimal policy in the Sensitive-Resistant model. Starting from an initial state $(q_0, p_0) = (0.1, 0.5)$ (magenta dot), the subfigures show (a) the deterministic trajectory without therapy that ends in the $\Delta_{\rm fail}$; (b) the optimal trajectory found from the deterministically driven system (2.1,2.19) with cost 49.30; (c) two representative sample paths generated under the deterministic-optimal policy but subject to stochastic perturbations in $(g_{\rm S}, g_{\rm R})$ (the brighter one incurs a total cost of 49.43, versus a much higher 70.45 for the duller-colored path); In (a), the *white dashed-line* is part of the nullcline where $\dot{p} = 0$; In both (b)&(c), *the green parts of trajectories* correspond to not prescribing drugs and *the red parts of trajectories* correspond to prescribing drugs at the MTD rate. The level sets of the value function u in the deterministic case are shown in *light blue*.

Stochastic perturbations can be similarly introduced here by assuming that the intrinsic

growth rates are actually random and normally distributed at any instant. (This approach was also used in modeling persistence strategies among bacteria in [27].) In Example 1, we assumed that the fitness function of each subpopulation was affected by its own random perturbations. The Brownian motion in Box 5 was three-dimensional, corresponding to subpopulations impacted by three separate and uncorrelated aspects of the fluctuating environment. Depending on the nature of perturbations, a similar assumption might be reasonable in the current example as well. But this is not a necessary feature for the threshold-aware optimization approach to be applicable. To demonstrate this, we will instead assume here that the same aspect of fluctuating environment impacts both subpopulations, and thus a single (1D) Brownian motion perturbs the intrinsic growth rates of both S and R. We will use $(g_{\mathrm{S}}, g_{\mathrm{R}})$ to represent their expected growth rates and $(\sigma_{\mathrm{S}}, \sigma_{\mathrm{R}})$ to denote their respective volatilities. This yields SDEs for the stochastic evolution of (Q, P), which are derived in §3.2.3 of Chapter 3 and summarized in Box 6. As shown in Fig 2.2(c), if the deterministic-optimal policy $d_\star$ is used in this stochastic setting, the initiation time of the MTD-based therapy (and the resulting overall cost $\mathcal{J}$) can vary significantly. This motivates us again to use the threshold-aware approach based on the PDE (2.11), with the policies illustrated and advantages quantified in §2.3.2.

2.3 Results

2.3.1 Policies, trajectories, and CDFs for the EGT-based model

We explore the structure and performance of threshold-aware policies computed for the system described in §2.2.2. The parameter values $d_{\max} = 3, b_a = 2.5, b_v = 2, c = 1, n = 4$ are the same ones provided in Kaznatcheev et al. [96] and Gluzman et al. [73]. However,

we use $\gamma_{\mathrm{r}} = 1 - \gamma_{\mathrm{f}} = 10^{-2}$ and $\delta = 0.05$ as opposed to $\gamma_{\mathrm{r}} = 1 - \gamma_{\mathrm{f}} = 10^{-1.5}$ and $\delta = 0.01$ in [73]. Additionally, we consider small uniform constant volatilities $\sigma_{\mathrm{G}} = \sigma_{\mathrm{D}} = \sigma_{\mathrm{V}} = 0.15$, characterizing the scale of random perturbations in fitness function for all 3 cancer subpopulations. The details of our Monte-Carlo simulations used to build all CDFs can be found in Chapter 3 §3.4.3. Additional examples, including those with higher volatilities, in which the threshold-performance advantages are even more significant, can be found in Chapter 3 §3.5.

In Fig 2.3, we present some representative s-slices of threshold-aware optimal policies and their corresponding optimal probability of success for respective threshold values. Since these policies are also bang-bang, the drugs-on region (at the MTD level) is shown in yellow and the drugs-off region is shown in blue in all of our figures, following the convention from [73]. We observe that this drugs-on region is strongly s-dependent and completely different from the one in the deterministic-optimal case shown in Fig 2.1(a). Since the cancer evolution considered here has stochastic dynamics given in (2.5, 2.15), different realizations of random perturbations will result in entirely different sample paths even if the starting configuration and the feedback policy remain the same. Three such representative sample paths are shown in Fig 2.4, starting from the same initial tumor configuration $(q_0, p_0) = (0.26,\, 0.665)$ already used in Fig 2.1 and focusing on a threshold $\bar{s} = 5$. We use the example from Fig 2.4(a), in which the stabilization is achieved while incurring the total cost of $\mathcal{J} = 4.70 < \bar{s}$, to illustrate the general use of threshold-aware policies. Starting from the initial budget $s = \bar{s}$, the optimal decision on whether to use drugs right away is based on the first diagram in Fig 2.3(a). For our initial tumor state, this indicates that $d_*(q_0, p_0, 5) = 0$ (not prescribing drugs initially) would maximize the probability of stabilizing the tumor without exceeding the threshold $\bar{s} = 5.0$. As time passes, we accumulate the cost, thus decreasing the budget, even if the drugs are not used. If we stay in the blue region for the time $\theta = 1/\delta$, the second diagram (the "$s = 4.0$" case) in Fig 2.3(a) becomes relevant, with subsequent budget decreases shifting

us to lower and lower s slices. Of course, in reality we constantly reevaluate the decision on d_* (as s changes continuously while Fig 2.3(a) presents just a few representative slices) taking into account the changing tumor configuration $(Q(t), P(t))$. (Movies with additional information for Figs 2.3 and 2.4 are available at `https://eikonal-equation.github.io/Stochastic-Cancer/examples.html`.)

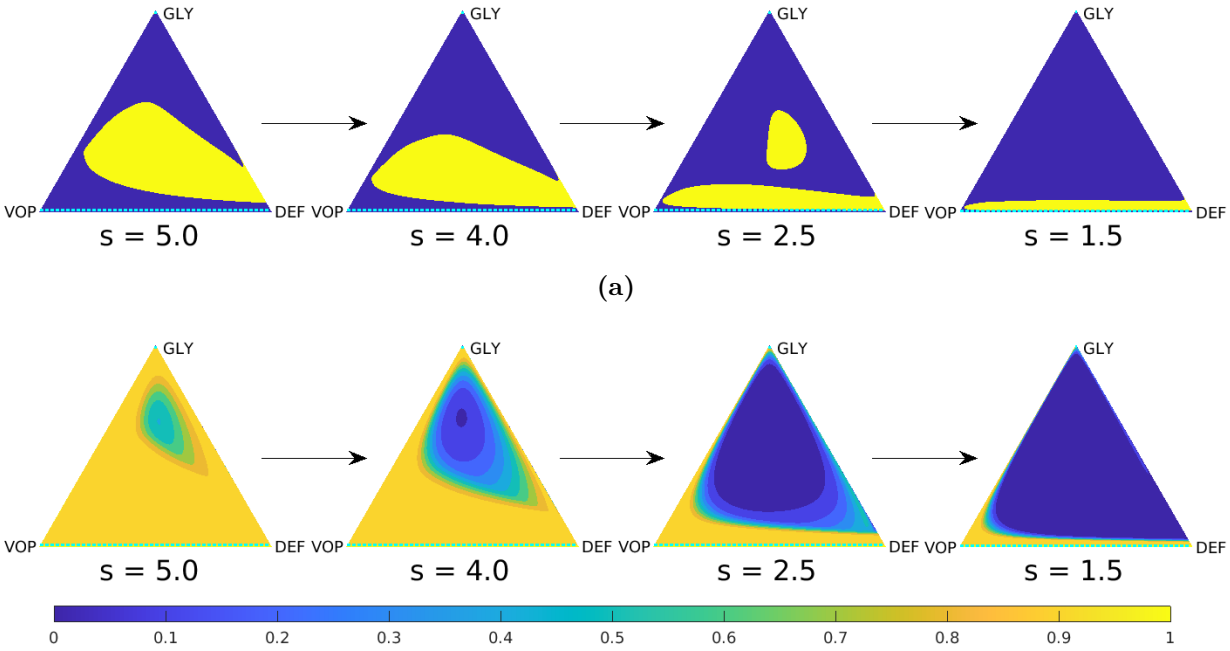


Figure 2.3. Representative slices of the threshold-aware optimal policy (top row) and the corresponding probability of success (bottom row) for the EGT-based model. Each triangle represents all possible tumor compositions (proportions of GLY/VOP/DEF cells in the population). Top row shows the policy, which prescribes the optimal instantaneous decisions on drug usage given the indicated remaining budget (s) and the current tumor state. Bottom row shows the probability of "stabilization within the budget" if the optimal policy is followed from this time point and onward. Each column corresponds to a specific budget level s, which is shown below each triangle. The arrows indicate the natural decrease of the remaining budget while implementing the policy.

In contrast to the success story in Fig 2.4(a), we note that there are two very different ways of "failing". First, the process can stop if the proportion of GLY cells becomes too high, as in Fig 2.4(b). When VOP is relatively low, the deterministic portion of the dynamics can bring us close to the failure barrier, with random perturbations resulting in a noticeable probability

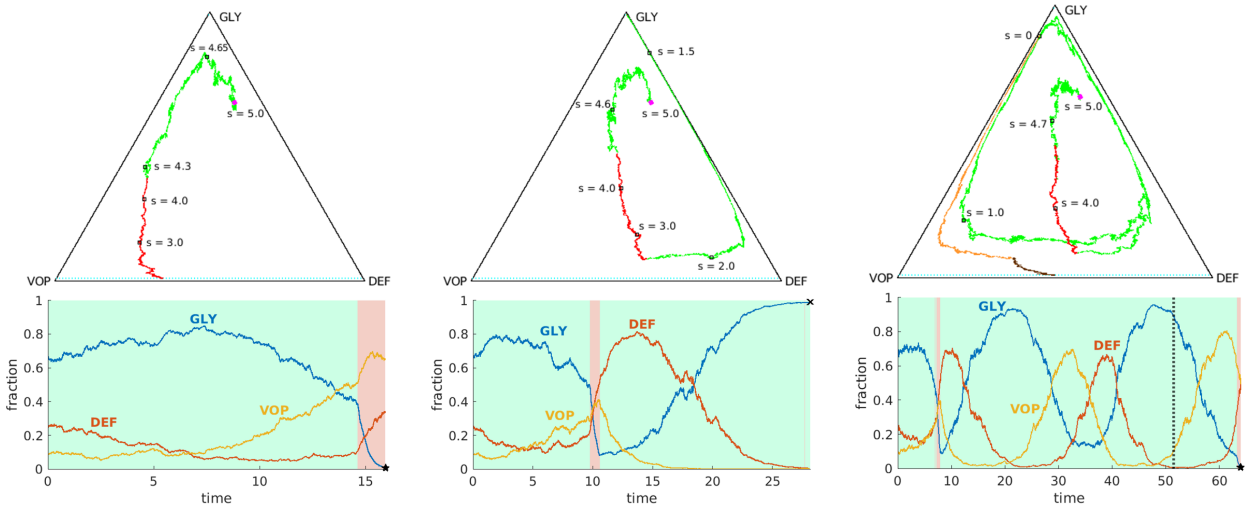


Figure 2.4. Representative sample paths starting from the same initial state $(q_0, p_0) = (0.26, 0.665)$ (magenta dot) and the same initial budget $\bar{s} = 5$. **Top row:** sample paths on a GLY-DEF-VOP triangle. (a) eventual stabilization with a cost of 4.70 (within the budget); (b) eventual death; (c) failure by running out of budget (eventual stabilization with a total cost of 7.80 by switching to the deterministic-optimal policy after $s = 0$). Some representative tumor states along these paths (with indications of how much budget is left) are marked by *black squares*. In (c), the part where $\mathcal{J} > 5$ is specified in *orange* (no drugs) and *brown* (at MTD level). **Bottom row:** evolution of sub-populations with respect to time based on the sample paths from the top row. Here we use *light green* and *light pink backgrounds* to indicate the time interval(s) of prescribing no drugs and of prescribing drugs at the MTD-rate, respectively. We use *black pentagrams* and *black crosses* to indicate eventual stabilization and death, respectively. In (c), we use a *dashed black* line to indicate the budget depletion time $t_\sharp$.

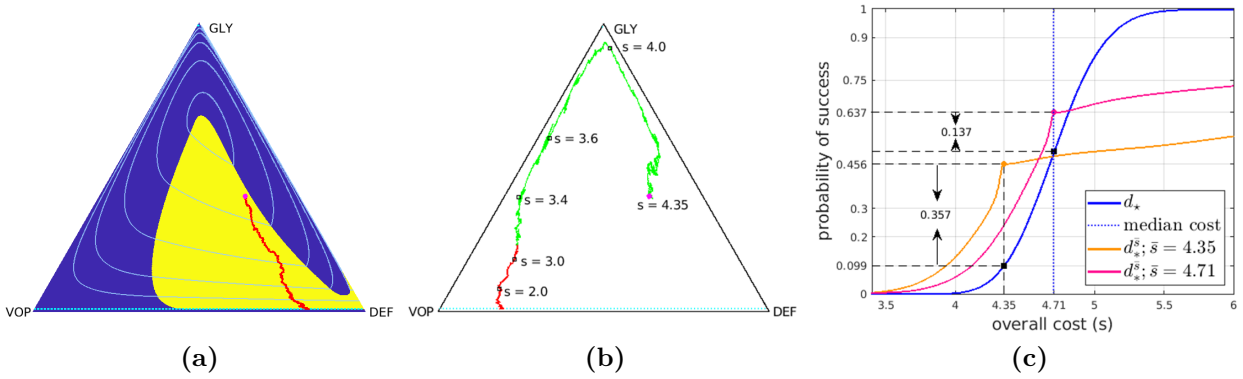
of crossing into Δ_{fail}. Second, even if we stay away from Δ_{fail}, the budget might be exhausted before reaching Δ_{succ}, as in Fig 2.4(c). Threshold-aware policies provide no guidance once $s = 0$, but it is reasonable to continue (using some different treatment policy) since the patient is still alive. In our numerical simulations, we switch in this case to a deterministic-optimal policy $d_\star$ illustrated in Fig 2.1. This decision is somewhat arbitrary; e.g., one could choose instead to switch to an MTD-based policy, which in this example maximizes the probability of reaching Δ_{succ} while avoiding Δ_{fail} without any regard to additional cost incurred thereafter. For this initial tumor configuration and parameter values, continuing with $d_\star$ typically yields smaller costs while only slightly increasing the chances of eventual

failure (e.g., $\approx 0.36\%$ of crossings into Δ_{fail} using $d_\star$ versus $\approx 0.07\%$ using the full MTD once the original budget of $\bar{s} = 5.0$ is exhausted). But whatever new policy is chosen for such "unlucky" cases, this choice will only affect the right tail of $\mathcal{J}$'s distribution; i.e., $\mathbb{P}(\mathcal{J} \geq \tilde{s})$ will be affected only for $\tilde{s} > \bar{s}$.

Returning to the optimal probability of success $v(q, p, s)$ shown in Fig 2.3(b), we observe that v has particularly large gradient near the level curves of the deterministic-optimal value function u shown in Fig 2.1(a). (The particular level curve of u near which v changes the most is again s-dependent as the budget decreases.) If the remaining budget is relatively low (e.g., $s = 1.5$), one can see from Fig 2.3(b) that there is no chance to stabilize the tumor within this budget unless the GLY is already low (and a short burst of drug therapy would likely be enough) or VOP is high (and the no-drugs dynamics will bring us to a low GLY concentration later on). Consequently, the optimal policy for $s = 1.5$ is to not use drugs for the majority of tumor states.

The contrast in threshold-specific performance is easy to explain when the deterministic-optimal and threshold-aware policies prescribe different actions from the very beginning. To illustrate this, we consider $(q_0, p_0) = (0.27, 0.4)$, for which $d_\star = d_{\max}$ while $d_*^{\bar{s}} = 0$ for a range of $\bar{s}$ values; see Fig 2.5(a) and 2.5(b) for representative paths and 2.5(c) for the respective CDFs. Under the deterministic-optimal policy (whose CDF is shown in blue), only 50% of simulations yield the cost not exceeding 4.71. A threshold-aware policy (implemented for $\bar{s} = 4.71$, with CDF shown in pink) maximizes this $\mathbb{P}(\mathcal{J} \leq \bar{s})$ and succeeds in 63.7% of all cases. The potential for improvement is even more significant with lower threshold values. For instance, we see that $\mathbb{P}(\mathcal{J}(d_\star) \leq 4.35) < 10\%$, while our threshold-aware policy (implemented for $\bar{s} = 4.35$, with CDF shown in orange) ensures that $\mathbb{P}\left(\mathcal{J}(d_*^{\bar{s}}) \leq 4.35\right) = v(q_0, p_0, \bar{s}) \approx 45.6\%$. This improvement can also be translated to simple medical terms: starting from this initial tumor configuration, the deterministic-optimal policy will likely

keep using the drugs at the maximum rate $d_{\max}$ all the way to stabilization; see Fig 2.5(a). In contrast, our threshold-aware policies tend not to prescribe drugs until GLY is relatively low and VOP is relatively high; see Fig 2.5(b). As a result, the patient would suffer less toxicity from drugs in most scenarios.



(a) **(b)** **(c)**

Figure 2.5. Comparison between threshold-aware policies and the deterministic-optimal policy. Starting from an initial state $(q_0, p_0) = (0.27, 0.4)$ (magenta dot): (a) a sample path with cost 4.75 under the deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 4.35$ with a realized total cost of 4.02 under the (orange) threshold-aware policy; (c) CDFs of the cumulative cost $\mathcal{J}$ approximated using 10^5 random simulations. In (c), the *solid blue* curve is the CDF generated with the deterministic-optimal policy. Its median (*dashed blue* line) is 4.71 while its mean conditioning on success is 4.72. The *solid orange* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.35$; and the *solid pink* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.71$.

It is worth noting that each threshold-aware policy maximizes the probability of success for a single/specific threshold value only. E.g., for all the pink/orange CDFs we have provided, the probability of success is only maximized at those pink/orange dots. Moreover, we clearly see from Fig 2.5(c) that the probability of $\mathcal{J}$ not exceeding *any* $\tilde{s} \leq 4.35$ is lower on the pink CDF than on the orange CDF (computed for $\bar{s} = 4.35$). Intuitively, this is not too surprising. In the early stages of treatment, a (pink) policy computed to maximize the chances of not exceeding $\bar{s} = 4.71$ is more aggressive in using the drugs and thus spends the "budget" quicker than the (orange) policy, which starts from a lower initial budget $\bar{s} = 4.35$. This is also consistent with the budget-dependent sizes of drugs-on regions in Fig 2.3(a).

2.3.2 Policies, trajectories, and CDFs for the SR-model

We now turn to the SR model system described in §2.2.3. Our numerical experiments use $d_{\max} = 3$, $\gamma_{\rm r} = 1 - \gamma_{\rm f} = 10^{-2}$, $\delta = 0.05$, and volatilities $\sigma_{\rm R} = \sigma_{\rm S} = 0.15$. For other parameter values, see §3.5.2 of Chapter 3.

We show the representative s-slices of threshold aware policies and the corresponding success probabilities in Fig 2.6. Similarly to the EGT-model, we observe that the drug-on regions (shown in yellow) are strongly budget-dependent and quite different from the ones specified by $d_\star$ in Fig 2.2(b). We note that the drugs-on region generally shrinks in size (toward the $Q = 1$ line, where only S cells are present) as the budget s decreases. For even tighter budgets, this yellow region becomes disconnected, prescribing the drugs for large P values (to substantially decrease the tumor size) and in a thin layer near $\Delta_{\rm succ}$ (where a short burst of drugs is likely sufficient).

In Fig 2.7, we provide sample random trajectories and compare the performance of three different policies: the deterministically optimal $d_\star$ and the threshold-aware $d_*^{\bar{s}}$ implemented for two different thresholds $\bar{s} = 69.45$ and $\bar{s} = 60$. A suitable choice of the initial tumor configuration is less obvious for this example and deserves a separate comment. For many multi-population models, it is reasonable to assume that the system had approached some drug-free coexistence equilibrium before the tumor was detected and the therapy started. But since the model described in [30] does not include mutations, it also does not have a drug-free coexistence equilibrium. In our testing of various drug policies, we choose the initial tumor with 96% of sensitive cells and the tumor size at 90% of the carrying capacity. Since the resistant cells are much larger [30], this corresponds to initial conditions $(q_0, p_0) = (0.45, 0.9)$.

Despite the fact that all three tested policies use no drugs at the very beginning, the deterministic-optimal policy typically starts prescribing drugs much earlier. See the com-

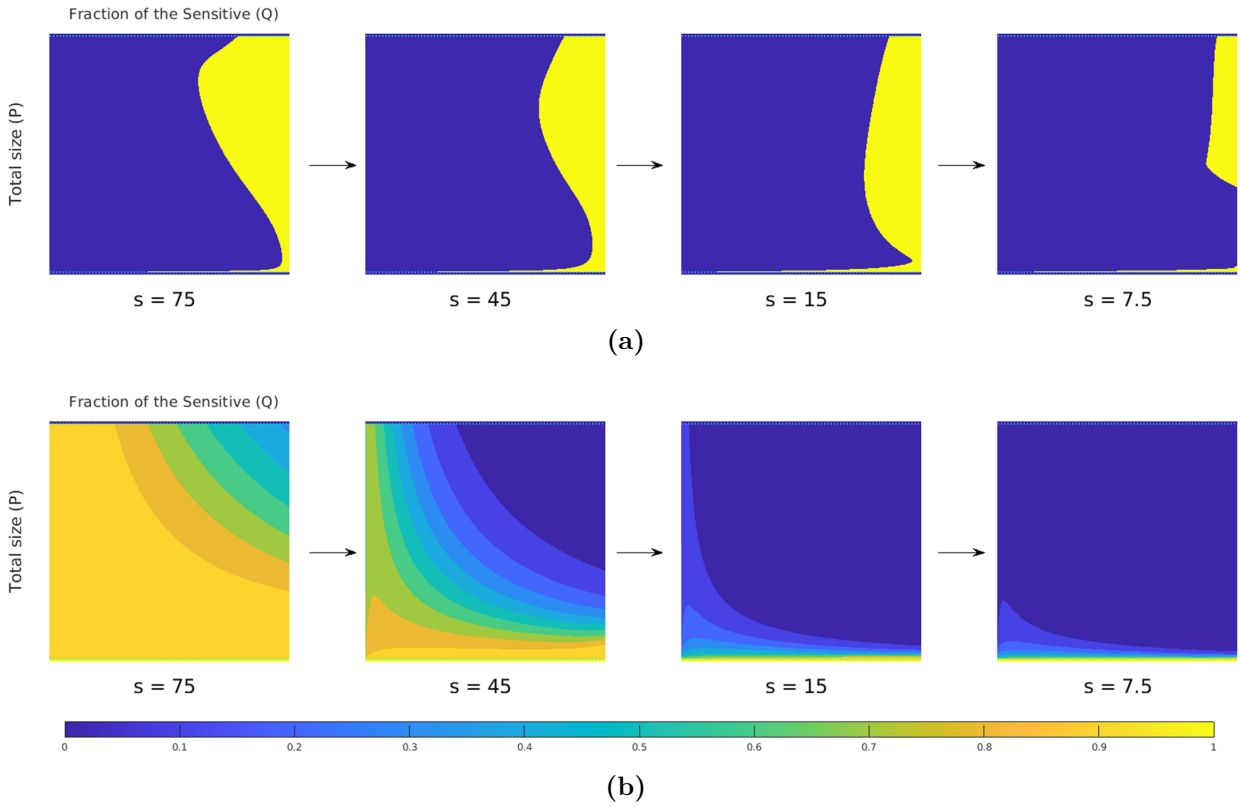


Figure 2.6. Representative slices of the threshold-aware optimal policy (top row) and the corresponding probability of success (bottom row) for the Carrère example. Each square represents all possible tumor states (sizes and compositions). The horizontal axis is the *fraction of the Sensitive (Q)* and the vertical axis is the *total population (P)*. Top row shows the policy, which prescribes the optimal instantaneous decisions on drug usage given the indicated remaining budget (s) and the current tumor state. Bottom row shows the probability of "eradication within the budget" if the optimal policy is followed from this time point and onward. Each column corresponds to a specific budget level s, which is shown below each square. The arrows indicate the natural decrease of the remaining budget while implementing the policy.

parison of sample trajectories under $d_\star$ and $d_*^{\bar{s}}$ in Fig 2.7(a) and 2.7(b). As a result, our threshold-aware policy (implemented for $\bar{s} = 69.45$, with CDF shown in pink) improves $\mathbb{P}(\mathcal{J} \leq \bar{s})$ to 67.4% from 50% produced by $d_\star$. This advantage is even more significant with lower thresholds. E.g., $\mathbb{P}(\mathcal{J}(d_\star) \leq 60)$ is only 19.6%, while our threshold-aware policy (implemented for $\bar{s} = 60$, with CDF shown in orange) more than doubles this probability of under-threshold remission to 39.8%.

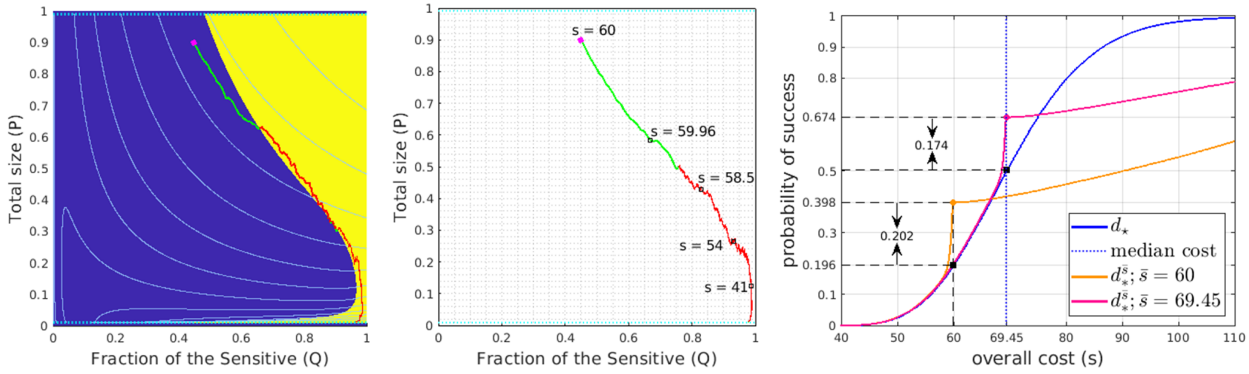


Figure 2.7. Comparison between threshold-aware policies and the deterministic-optimal policy. Starting from an initial state $(q_0, p_0) = (0.45, 0.9)$ (magenta dot): (a) a sample path with cost 57.3 under the deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 60$ with a total cost of 53.63 under the (orange) threshold-aware policy; (c) CDFs of the cumulative cost $\mathcal{J}$ with 10^5 samples. In (c), the *solid blue* curve is the CDF generated with the deterministic-optimal policy. Its median (*dashed blue* line) is 69.45 while its mean conditioning on success is 70.5. The *solid orange* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 60$; and the *solid pink* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 69.45$. See Chapter 3 §3.5.3 for time-evolution plots associated with sample paths in (a) and (b).

2.4 Discussion

That cancers evolve during therapy is now an accepted fact, and is slowly being incorporated into therapeutic decision making. In some cases, this can be implemented simply by changing from one targeted therapy to another, but in most, where tumors are a heterogeneous mixture of interacting phenotypes, this is not feasible. In these cases, ecological thinking is rising to the fore in the form of adaptive therapy. Until recently, clinical trials, and theoretical investigations, of adaptive therapy have relied on *a priori* assumptions of the underlying interactions, and their effects on tumor composition over time. Several studies, both *in vitro*[95] and *in vivo*[114, 51], however, have begun to provide methods for more rigorous quantification of these interactions. As these tools mature, the next challenges will be to understand these interactions in patients and to exploit them in improving personalized

treatment.

The presented approach is a step in this direction, aiming to limit the probability of high-cost outcomes in the presence of stochastic perturbations. It is applicable to a broad class of stochastic cancer models and therapy goals (e.g., tumor eradication or stabilization). While it is standard to tune the treatment plan to maximize the probability of reaching its goal, we go farther and maximize the probability of goal attainment without exceeding a prescribed threshold on cumulative cost (interpreted as a combination of the total drugs used, cumulative disease burden, and the time to remission/stabilization). We show that these optimal treatment policies become *threshold-aware*, with the drugs-on/drugs-off regions changing as the treatment progresses and the initial "cost budget" (for meeting the chosen threshold) gradually decreases. The comparison of CDFs generated for the deterministic-optimal policy and threshold-aware policies demonstrates clear advantages of the latter, often resulting in a significant reduction of drugs used to treat the patient.

More generally, dynamic programming provides an excellent framework for finding optimal treatment policies by solving Hamilton-Jacobi-Bellman (HJB) equations. The fact that these policies are recovered in feedback form makes this approach particularly suitable for optimization of adaptive therapies. But even though the use of general optimal control in cancer treatment is by now common [145], the same is not true for the more robust HJB-based methods, which so far have been used in only a handful of cancer-related applications [126, 1, 60, 73, 31, 103, 91]. This is partly due to the HJBs' well-known *curse of dimensionality*: the rapid increase in computational costs when the system state becomes higher-dimensional. This is a relevant limitation since our threshold-aware approach introduces the "budget" as an additional component of the state. Similarly to the presented examples, our current implementation would be easy to adopt to any cancer model based on a two-dimensional (q, p) state space, with the budget s adding the third dimension. For

cancer models with a larger number of subpopulations, the general approach would remain the same, but the approximate HJB-solver would likely need to rely on sparse grids [116], tensor decompositions [47], or deep neural networks [178].

The presented examples did not model any mutations, but we note that this is not really a limitation of the method itself. E.g., drug-usage-dependent mutations would be easy to incorporate into our EGT-based example by switching to a Replicator-Mutator ODE/SDE eco-evolutionary model [80, 85]. We did not pursue such examples here primarily to make for an easier comparison with prior work [73] and to limit the number of model parameters.

Our SDE models of global (or environmental) stochastic perturbations to subpopulation fitnesses and intrinsic growth rates are based on perspectives well-established in biological applications [65, 27]. While our focus on environmental stochasticity is motivated by "averaging-out" the variability within each subpopulation, it is worth noting that this assumption is not always justifiable. Whenever the subpopulation size is sufficiently small, the demographic stochasticity becomes crucially important. (This is also the regime in which the validity of ODE/SDE models is far less obvious.) Even though we do not deal with this important limitation here, we note that our threshold-aware approach can be used with a variety of perturbation types, including jump-diffusion processes, which could be used to build future models that account for demographic stochasticity in these special small-subpopulation regimes. Such discontinuous jump-transitions (e.g., reflecting possible subpopulation extinctions) can be naturally handled in our framework. For instance, a similar method has been developed in [32] for controlling "piecewise-deterministic" processes, where perturbations happen at discrete points in time and amount to abrupt switches in system dynamics. More recently, our framework was also used to control the hybrid dynamics of a sailboat navigating in stochastically changing wind conditions and trying to reach the destination prior to a specified deadline $\bar{s}$ [168]. We note that dynamic programming is

also used in discrete population models focused on demographic stochasticity [119]. It will be also interesting to investigate the usability of our approach in that discrete setting.

Sensitivity with respect to threshold variation can be tested by comparing CDFs of $d_*^{\bar{s}}$ for different $\bar{s}$ values. While it is also possible to perform a similar comparison under perturbation of model parameters, we believe that another approach is more promising: any bounded uncertainty in parameter values can be treated as a "game against nature," leading to a Hamilton-Jacobi-Isaacs PDE, whose solution will yield policies optimizing the threshold-performance in the "worst parameter variation" scenarios [32].

Another important extension will be to move to "partial observability" since the state of the tumor is only occasionally assessed directly through biopsies and some proxy measurements have to be used at all other times [60]. Finally, it will be also interesting to study the multiobjective control problem of optimizing threshold-aware policies for two different threshold values simultaneously.

In summary, we have presented a theoretical and computational advance for the toolbox of evolutionary therapy, a new subfield of medicine focused on using knowledge of evolutionary responses to inform therapeutic scheduling. While there are a number of cancer trials using this type of evolutionary-informed thinking, most are based on heuristic designs and are not formulated to consider the underlying stochasticities. Developing a theoretical foundation for future clinical studies requires EGT models directly grounded in objectively measurable biology [95]. Therapy optimization based on such models requires efficient computational methods, particularly in the presence of stochastic perturbations. We hope that the general approach presented here will be useful for a broad range of increasingly accurate stochastic cancer models.

CHAPTER 3

MATHEMATICAL AND NUMERICAL DETAILS FOR THRESHOLD-AWARE CANCER THERAPY

This chapter provides mathematical derivations, numerical schemes and their implementation details, and additional examples for Chapter 2 "Threshold-awareness in adaptive cancer therapy." We provide the full source code of our numerical schemes presented in §3.4 and movies for both small (§2.3.1) and large (§3.5.4) volatilities at `https://eikonal-equation.github.io/Stochastic-Cancer/`.

3.1 Derivations of controlled unperturbed/deterministic systems

3.1.1 The EGT-based competition model

We start by producing the main derivations of the controlled unperturbed system of equations (labeled (2.14) in Chapter 2) adopted from Gluzman et al. [73]. The original model from Kaznatcheev et al. [96] describes a competition of 3 types of cancer cells: Glycolytic cells (GLY) are anaerobic and produce lactic acid. The other two types are aerobic and benefit from the oxygen from vascularization. The VEGF (over)-producing cells (VOP) improve the vasculature development, while the remaining aerobic cells are defectors (DEF). As mentioned in Chapter 2, the competition of cells in the tumor is thus modeled as a "public goods" / "club goods" game: VEGF is a "club good" since it benefits only VOP and DEF cells, while the acid generated by GLY is a "public good" as it benefits all cancer cells.

The original model in [96] is based on a game of $(n+1)$ locally interacting cancer cells. The fitness of each of them depends on the proportion of all three types among these game

participants. Define $A_n(k) = \dfrac{b_a k}{n+1}$ to be the benefit to each cell due to acidity if k of $(n+1)$ participants are producing acid. Suppose $(m+1)$ of these $(n+1)$ cells are aerobic. Then $V_m(k) = \dfrac{b_v k}{m+1}$ is defined to be the benefit (to each aerobic cell) due to increased vascularization if k of these $(m+1)$ cells are (over) producing VEGF at a personal cost c.

Focusing on one specific participant, suppose $n_{\mathrm{G}}, n_{\mathrm{D}}, n_{\mathrm{V}}$ are the numbers of GLY, DEF, and VOP cells among all others, with $n_{\mathrm{G}} + n_{\mathrm{D}} + n_{\mathrm{V}} = n$. Thus, if that "focus participant" is a GLY cell, the benefit due to acidity is $A_n(n_{\mathrm{G}} + 1)$ (it is actually its payoff since GLY is the only anaerobic type). Similarly, if the focus participant is an aerobic type, the benefit due to acidity becomes $A_n(n_{\mathrm{G}})$. When the focus participant is a VOP cell, the benefit it receives due to vascularization is $V_{n-n_{\mathrm{G}}}(n_{\mathrm{V}} + 1)$ while paying a constant cost c to produce VEGF. Note that the group size of aerobic cells among n total nearby cells is $n - n_{\mathrm{G}}$. On the other hand, if the focus participant is a DEF cell, it receives a benefit of $V_{n-n_{\mathrm{G}}}(n_{\mathrm{V}})$ due to vascularization while paying no cost. We note that the benefit received by a focus participant depends heavily on that participant's type and on the types of others. To determine the fitness functions ψ, Kaznatcheev et al. further assumed that all participants are drawn uniformly at random from a large well-mixed population. They then computed the *expected benefit* for each type of focus participants, by averaging over all possible compositions of their local interaction group [96].

To simplify the notation for the EGT-based model, we will use subscripts 1, 2, and 3 to indicate GLY, DEF and VOP cells respectively. Summarizing the above in this new notation, we list the fitness function for each of the three types of cancer cells:

$$\begin{cases} \psi_1 = \langle A_n(n_1 + 1)\rangle_{n_1 \sim B_n(x_1)}, \\ \psi_2 = \langle A_n(n_1)\rangle_{n_1 \sim B_n(x_1)} + \langle V_{n-n_1}(n_3)\rangle_{(n_1,n_3)\sim M_n(x_1,x_3)}, \\ \psi_3 = \langle A_n(n_1)\rangle_{n_1 \sim B_n(x_1)} + \langle V_{n-n_1}(n_3 + 1)\rangle_{(n_1,n_3)\sim M_n(x_1,x_3)} - c, \end{cases} \tag{3.1}$$

where $B_n(x)$ is the binomial distribution with n samples and x is the probability of success, $M_n(x, y)$ is the multinomial distribution with n samples, three possible outcomes with respective probabilities $(x,\, 1-x-y,\, y)$; and $\langle f(\zeta)\rangle_{\zeta\sim\Omega}$ is the expected value of $f(\zeta)$ over a random variable ζ with distribution Ω. The expectations in (3.1) can be also computed explicitly; we refer to Section 1S of Supplementary Materials in [73] and Appendix A in [96] for details.

Let z_1, z_2, z_3 denote the absolute sizes of subpopulations of GLY, DEF, VOP cells. We assume each sub-population grows with rate equal to its respective fitness, and the GLY-targeting therapy (control) kills GLY cells only with time-dependent rate $d : \mathbb{R}_+ \to [0, d_{\max}]$. Then the controlled dynamics of sub-populations is

$$\begin{cases} \dot{z}_1(t) = \psi_1(t) z_1(t) - d(t) z_1(t), \\ \dot{z}_2(t) = \psi_2(t) z_2(t), \\ \dot{z}_3(t) = \psi_3(t) z_3(t). \end{cases}$$

By definition, the proportion of GLY cells in the entire tumor is $x_1 = z_1/(z_1 + z_2 + z_3)$. Differentiating both sides with respect to t, we obtain the dynamics for x_1:

$$\begin{aligned} \dot{x}_1 &= \frac{\dot{z}_1}{z_1+z_2+z_3} - \frac{(\dot{z}_1+\dot{z}_2+\dot{z}_3)x_1}{z_1+z_2+z_3} \\ &= \frac{z_1}{z_1+z_2+z_3}(\psi_1 - d(t)) - \left[\frac{z_1}{z_1+z_2+z_3}(\psi_1 - d(t)) + \frac{z_2}{z_1+z_2+z_3}\psi_2 + \frac{z_3}{z_1+z_2+z_3}\psi_3\right] x_1 \\ &= (\psi_1 - dd(t))x_1 - \left[x_1(\psi_1 - d(t)) + x_2\psi_2 + x_3\psi_3\right] x_1 \\ &= (\psi_1 - dd(t))x_1 + dd(t)\, x_1^2 - \langle\psi\rangle x_1 \\ &= (\psi_1 - \langle\psi\rangle)x_1 - x_1(1-x_1)d(t), \end{aligned}$$

where $\langle\psi\rangle = x_1\psi_1 + x_2\psi_2 + x_3\psi_3$ is the average fitness.

The ODEs for $\dot{x}_2$ and $\dot{x}_3$ follow a similar argument, and hence the controlled unperturbed

system (a replicator equation) is:

$$\begin{cases} \dot{x}_1 = (\psi_1 - \langle\psi\rangle)x_1 - x_1(1-x_1)d(t), \\ \dot{x}_2 = (\psi_2 - \langle\psi\rangle)x_2 + x_2x_1d(t), \\ \dot{x}_3 = (\psi_3 - \langle\psi\rangle)x_3 + x_3x_1d(t). \end{cases}$$

Since $x_1 + x_2 + x_3 = 1$, one can transform the above 3-dimensional system into 2-dimensional. Let the proportion of glycolytic cells in the tumor to be $p(t) = x_1(t)$ and the proportion of VOP among aerobic cells to be $q(t) = x_3(t)/(x_2(t) + x_3(t))$.

Hence

$$\begin{aligned} \dot{p} &= p(\psi_1 - \langle\psi\rangle) - p(1-p)d(t) \\ &= p(\psi_1 - x_1\psi_1 - x_2\psi_2 - x_3\psi_3) - p(1-p)d(t) \\ &= p(\psi_1 - p\psi_1 - x_2\psi_2 - x_3\psi_3) - p(1-p)d(t) \\ &= p(1-p)(\psi_1 - \langle\psi\rangle_{2,3} - d(t)), \end{aligned}$$

where $\langle\psi\rangle_{2,3} = \psi_2 + q(\psi_3 - \psi_2)$.

Similarly,

$$\dot{q} = \frac{\dot{x}_3x_2 - \dot{x}_2x_3}{(x_2+x_3)^2} = \frac{x_3x_2}{(x_2+x_3)^2}(\psi_3 - \psi_2) = q(1-q)(\psi_3 - \psi_2).$$

By definitions of ψ_1, ψ_2, ψ_3 from above, one can find

$$\begin{cases} \psi_3 - \psi_2 = \dfrac{b_v}{n+1}\left[\sum\limits_{k=0}^{n} p^k\right] - c, \\ \psi_1 - \langle\psi\rangle_{2,3} = \dfrac{b_a}{n+1} - q(b_v - c). \end{cases}$$

See Section 1S of Supplementary Materials of [73] for detailed derivations of the above expressions. Substituting them into equations for $\dot{p}$ and $\dot{q}$, one obtains the controlled reduced

unperturbed system (labeled (2.14) in Box 5 in Chapter 2):

$$\begin{cases} \dot{q} = q(1-q)\left(\dfrac{b_v}{n+1}\left[\sum\limits_{k=0}^{n} p^k\right] - c\right), \\ \dot{p} = p(1-p)\left(\dfrac{b_a}{n+1} - (b_v - c)q - d(t)\right). \end{cases} \tag{3.2}$$

3.1.2 The SR competition model

We now provide derivations of the controlled unperturbed system of equations (labeled (2.19) in Chapter 2) based on a model proposed by Carrère [30]. As explained in the Chapter 2, their original model describes competition between two types of lung cancer cells *in vitro*: the sensitive (S) "A549" (sensitive to the drug "Epothilene") and the resistant (R) "A549 Epo40". The model was derived based on phenotypical observations and neglected mutation events.

We denote the absolute size of their respective population as z_{S} and z_{R}. The concentration of the drug (control) has a time-dependent rate $d : \mathbb{R}_+ \to [0, d_{\max}]$. With a normalization $z_{\mathrm{S}}(t) \to z_{\mathrm{S}}(t)/C$, $z_{\mathrm{R}}(t) \to z_{\mathrm{R}}(t)/C$, we have shown in Chapter 2 that the controlled dynamics follow

$$\begin{cases} \dot{z_{\mathrm{S}}}(t) &= g_{\mathrm{S}}\Big(1 - z_{\mathrm{S}}(t) - m z_{\mathrm{R}}(t)\Big) z_{\mathrm{S}}(t) - \alpha z_{\mathrm{S}}(t) d(t), \\ \dot{z_{\mathrm{R}}}(t) &= g_{\mathrm{R}}\Big(1 - z_{\mathrm{S}}(t) - m z_{\mathrm{R}}(t)\Big) z_{\mathrm{R}}(t) - \beta C z_{\mathrm{S}}(t) z_{\mathrm{R}}(t). \end{cases} \tag{3.3}$$

Focusing on the total size of the tumor $p(t) = z_{\mathrm{S}}(t) + m z_{\mathrm{R}}(t)$ and the fraction of tumor size taken by sensitive cells $q(t) = z_{\mathrm{S}}(t)/p(t)$, we now derive the ODEs in this new coordinates $(q(t), p(t))$.

$$\begin{aligned} \dot{p} &= \dot{z_{\mathrm{S}}} + m\dot{z_{\mathrm{R}}} \\ &= g_{\mathrm{S}}\Big(1 - z_{\mathrm{S}} - m z_{\mathrm{R}}\Big) z_{\mathrm{S}} - \alpha z_{\mathrm{S}} d(t) + g_{\mathrm{R}}\Big(1 - z_{\mathrm{S}} - m z_{\mathrm{R}}\Big) z_{\mathrm{R}} - \beta C z_{\mathrm{S}} z_{\mathrm{R}} \\ &= g_{\mathrm{S}} qp(1-p) - \alpha qp\, d(t) + g_{\mathrm{R}}(1-q)p(1-p) - \beta C p^2 q(1-q) \\ &= p(1-p)[g_{\mathrm{S}} q + g_{\mathrm{R}}(1-q)] - \beta C p^2 q(1-q) - \alpha qp\, d(t) \end{aligned}$$

$$
\begin{aligned}
\dot{q} &= \frac{\dot{z}_{\mathrm{S}}}{p} - q \cdot \frac{\dot{p}}{p} \\
&= \frac{g_{\mathrm{S}}\big(1-p\big)qp - \alpha qp\, d(t)}{p} - q \cdot \frac{p(1-p)[g_{\mathrm{S}}q + g_{\mathrm{R}}(1-q)] - \beta C p^2 q(1-q) - \alpha qp\, d(t)}{p} \\
&= g_{\mathrm{S}}q(1-p) - \alpha q\, d(t) - q(1-p)[g_{\mathrm{S}}q + g_{\mathrm{R}}(1-q)] + \alpha q^2 d(t) + \beta C p q^2(1-q) \\
&= g_{\mathrm{S}}q(1-p)(1-q) - g_{\mathrm{R}}q(1-q)(1-p) + \beta C p q^2(1-q) - \alpha q(1-q)d(t) \\
&= (1-p)q(1-q)(g_{\mathrm{S}} - g_{\mathrm{R}}) + \beta C p q^2(1-q) - \alpha q(1-q)d(t)
\end{aligned}
$$

We thus obtains the controlled unperturbed system (labeled (2.19) in Box 6 in Chapter 2) in (q, p) coordinates:

$$
\begin{cases}
\dot{q} = (1-p)q(1-q)(g_{\mathrm{S}} - g_{\mathrm{R}}) + \beta C p q^2(1-q) - \alpha q(1-q)d(t), \\
\dot{p} = p(1-p)[g_{\mathrm{S}}q + g_{\mathrm{R}}(1-q)] - \beta C p^2 q(1-q) - \alpha qp\, d(t).
\end{cases}
\tag{3.4}
$$

3.2 Problem setup and derivations for stochastic models

3.2.1 Problem setup under a general stochastic optimal control framework

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ denote a complete filtered probability space. The canonical reference probability space for a standard m-dimensional Brownian motion is defined by the following 5-tuple [62, 53]

$$
\nu := \Big(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{W}\Big),
$$

where

- $\Omega := C([0,\infty), \mathbb{R}^n)$ is the space of all continuous functions from $[0,\infty)$ to $\mathbb{R}^n$ equipped with the metric of uniform convergence on compact sets;

- $\mathcal{F} := \mathcal{B}(\Omega)$ is the Borel σ-algebra on Ω;
- $\mathbb{P}$ is the Wiener measure;
- $\{\mathcal{F}_t\}_{t\geq 0}$ is the completion of the natural filtration of an m-dimensional Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$;
- $\boldsymbol{W} = (W_1, W_2, \ldots W_m) : [0,\infty) \times \Omega \to \mathbb{R}^m$ in $\mathbb{R}^m$ is an $\mathcal{F}_t$-Brownian motion.

Let $\boldsymbol{d}(\cdot) : [0,\infty) \times \Omega \to \mathcal{D}$ ($\mathcal{D}$ compact) be a $\mathcal{F}_t$-progressively measurable control process. By "$\mathcal{F}_t$-progressively measurable", we mean for each $t \in [0,\infty)$, the map $(t,\omega) \to \boldsymbol{d}(t,\omega)$ from $[0,t] \times \Omega$ into $[0, d_{\max}]$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable [62, 53]. For the sake of notational simplicity, we will suppress the ω-dependence in the rest of this chapter.

We focus on stochastic dynamics of the following form:

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{a}(\boldsymbol{X}, \boldsymbol{d})\, \mathrm{d}t + \boldsymbol{\Sigma}(\boldsymbol{X}, \boldsymbol{d})\, \mathrm{d}\boldsymbol{W}, \tag{3.5}$$

where $\boldsymbol{X} \in \mathbb{R}^n$ encodes the state variables, $\boldsymbol{a}(\boldsymbol{X}, \boldsymbol{d}) \in \mathbb{R}^n$ denotes the drift coefficient function, and $\boldsymbol{\Sigma}(\boldsymbol{X}, \boldsymbol{d}) \in \mathbb{R}^{n\times m}$ encodes the diffusion coefficient function. If $\boldsymbol{a}(Y, \boldsymbol{d})$ and $\boldsymbol{\Sigma}(Y, \boldsymbol{d})$ are *Lipschitz* in Y and *uniformly continuous* in $\boldsymbol{d}$, then Eq. (3.5) has a strong solution for a fixed initial condition [23, 52].

Consider an exit-time problem where $\Delta \subset [0,1]^n$ is the terminal set (for simplicity, we have normalized $\boldsymbol{X} \in [0,1]^n$). Furthermore, $\Delta = \Delta_{\text{succ}} \cup \Delta_{\text{fail}}$ is the union of the success region Δ_{succ} and the failure region Δ_{fail}. We define the terminal time as

$$T = T(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)) := \inf\Big\{ t \in \mathbb{R}_+ \;\mid\; \boldsymbol{X}(t) \in \Delta, \quad \boldsymbol{X}(0) = \boldsymbol{\xi} \Big\},$$

where $\boldsymbol{\xi}$ is any starting configuration.

We further define the total cost of using a policy $\boldsymbol{d}(\cdot)$ as

$$\mathcal{J}(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)) \;=\; \int_0^T K(\boldsymbol{X}(\tau), \boldsymbol{d}(\tau))\, \mathrm{d}\tau \;+\; g\Big(\boldsymbol{X}\Big(T\Big)\Big),$$

for some *strictly positive* running cost function K and a some non-negative terminal cost function g.

3.2.2 Derivation of controlled perturbed system for the EGT model

In this section, we provide detailed derivations of the controlled stochastic cancer dynamics based on the EGT model (labeled (2.15) in Chapter 2).

Derivation of controlled perturbed system for the fractions (X_1, X_2, X_3)

We start by providing detailed derivations of the controlled stochastic cancer dynamics for the fractions (X_1, X_2, X_3). As mentioned in Chapter 2, we choose the stochastic replicator equation originated from Fudenberg and Harris [65] to describe the driven dynamics.

In the ordinary replicator equation shown in §3.1.1, we assume the actual sub-population size of each type of cancer cells z_i grows with rate ψ_i. For the stochastic replicator equation, standard Brownian motion is added to the fitness function ψ_i as the source of perturbation. As a result, the driven differential equation for Z_i (we adopt the convention to use capital letters to denote random variables in this chapter as well) is modeled by geometric Brownian motion. The growth rate for Z_1 will be further affected by our control - the rate of drug administration $d(\cdot)$.

The controlled stochastic dynamics for sub-populations Z_1, Z_2, Z_3 is then:

$$\begin{cases} \mathrm{d}Z_1 = [(\psi_1 - d(t))\ \mathrm{d}t + \sigma_1\ \mathrm{d}W_1]\ Z_1, \\ \mathrm{d}Z_2 = [\psi_2\ \mathrm{d}t + \sigma_2\ \mathrm{d}W_2]\ Z_2, \\ \mathrm{d}Z_3 = [\psi_3\ \mathrm{d}t + \sigma_3\ \mathrm{d}W_3]\ Z_3, \end{cases}$$

where (W_1, W_2, W_3) is a standard 3-dimensional Brownian motion for (Z_1, Z_3, Z_3), and $(\sigma_1, \sigma_2, \sigma_3) \geq 0$ are the constant volatilities. In such a way, the expected fitness of each type is still ψ_i since standard Brownian motion has zero mean.

We first derive the Stochastic Differential Equation (SDE) for $X_1 = Z_1/(Z_1 + Z_2 + Z_3)$. By Itô's lemma, for multivariate function $f(X, Y, Z)$,

$$\begin{aligned} \mathrm{d}f = f_x\ \mathrm{d}X + f_y\ \mathrm{d}Y + f_z\ \mathrm{d}Z + \frac{1}{2}f_{xx}\ (\mathrm{d}X)^2 + \frac{1}{2}f_{yy}\ (\mathrm{d}Y)^2 + \frac{1}{2}f_{zz}\ (\mathrm{d}Z)^2 \\ + f_{xy}\ (\mathrm{d}X)(\mathrm{d}Y) + f_{xz}\ (\mathrm{d}X)(\mathrm{d}Z) + f_{yz}\ (\mathrm{d}Y)(\mathrm{d}Z). \end{aligned}$$

For X_1, we have $f(x, y, z) = \dfrac{x}{x + y + z}$, then

$$\begin{aligned} f_x &= \frac{y+z}{(x+y+z)^2},\ f_y = -\frac{x}{(x+y+z)^2},\ f_z = -\frac{x}{(x+y+z)^2} \\ f_{xx} &= -\frac{2(y+z)}{(x+y+z)^3},\ f_{yy} = \frac{2x}{(x+y+z)^3},\ f_{zz} = \frac{2x}{(x+y+z)^3} \\ f_{xy} &= \frac{x-y-z}{(x+y+z)^3},\ f_{xz} = \frac{x-y-z}{(x+y+z)^3},\ f_{yz} = \frac{2x}{(x+y+z)^3}. \end{aligned}$$

By convention, for independent Brownian motions,

$$(\mathrm{d}t)^2 = 0,\ \mathrm{d}t\ \mathrm{d}W_i = 0\ \forall i,\ \mathrm{d}W_i\ \mathrm{d}W_j = 0\ \forall i \neq j,\ (\mathrm{d}W_i)^2 = \mathrm{d}t\ \forall i.$$

Thus,

$$
\begin{aligned}
\mathrm{d}X_1 &= \mathrm{d}\left(\frac{Z_1}{Z_1+Z_2+Z_3}\right) \\
&= f_x\,\mathrm{d}Z_1 + f_y\,\mathrm{d}Z_2 + f_z\,\mathrm{d}Z_3 + \frac{1}{2}f_{xx}\,(\mathrm{d}Z_1)^2 + \frac{1}{2}f_{yy}\,(\mathrm{d}Z_2)^2 + \frac{1}{2}f_{zz}\,(\mathrm{d}Z_3)^2 \\
&\quad + f_{xy}\,(\mathrm{d}Z_1)(\mathrm{d}Z_2) + f_{xz}\,(\mathrm{d}Z_1)(\mathrm{d}Z_3) + f_{yz}\,(\mathrm{d}Z_2)(\mathrm{d}Z_3) \\
&= \frac{Z_2+Z_3}{(Z_1+Z_2+Z_3)^2}\left[(\psi_1 - d(t))\,\mathrm{d}t + \sigma_1\,\mathrm{d}W_1\right]Z_1 \\
&\quad - \frac{Z_1}{(Z_1+Z_2+Z_3)^2}\left[\psi_2\,\mathrm{d}t + \sigma_2\,\mathrm{d}W_2\right]Z_2 - \frac{Z_1}{(Z_1+Z_2+Z_3)^2}\left[\psi_3\,\mathrm{d}t + \sigma_3\,\mathrm{d}W_3\right]Z_3 \\
&\quad - \frac{1}{2}\frac{2(Z_2+Z_3)}{(Z_1+Z_2+Z_3)^3}Z_1^2\sigma_1^2\,\mathrm{d}t + \frac{1}{2}\frac{2Z_1}{(Z_1+Z_2+Z_3)^3}Z_2^2\sigma_2^2\,\mathrm{d}t + \frac{1}{2}\frac{2Z_1}{(Z_1+Z_2+Z_3)^3}Z_3^2\sigma_3^2\,\mathrm{d}t \\
&= \left[(X_2+X_3)\psi_1 - (X_2+X_3)d(t) - X_2\psi_2 - X_3\psi_3\right]X_1\,\mathrm{d}t \\
&\quad + \left[(X_2+X_3)\sigma_1\,\mathrm{d}W_1 - X_2\sigma_2\,\mathrm{d}W_2 - X_3\sigma_3\,\mathrm{d}W_3\right]X_1 \\
&\quad - \left[X_1(X_2+X_3)\sigma_1^2 - X_2^2\sigma_2^2 - X_3^2\sigma_3^2\right]X_1\,\mathrm{d}t \\
&= \left[X_1(\psi_1 - \langle\psi\rangle) - X_1(1-X_1)d(t)\right]\,\mathrm{d}t + \left(\sigma_1\,\mathrm{d}W_1 - \sum_{j=1}^{3}X_j\sigma_j\,\mathrm{d}W_j\right)X_1 \\
&\quad - \left(\sigma_1^2X_1 - \sum_{j=1}^{3}X_j^2\sigma_j^2\right)X_1\,\mathrm{d}t.
\end{aligned}
$$

For $X_2 = Z_2/(Z_1+Z_2+Z_3)$, we have $f(x,y,z) = \dfrac{y}{x+y+z}$. Then

$$
\begin{aligned}
f_x &= -\frac{y}{(x+y+z)^2},\ f_y = \frac{x+z}{(x+y+z)^2},\ f_z = -\frac{y}{(x+y+z)^2} \\
f_{xx} &= \frac{2y}{(x+y+z)^3},\ f_{yy} = -\frac{2(x+z)}{(x+y+z)^3},\ f_{zz} = \frac{2y}{(x+y+z)^3} \\
f_{xy} &= \frac{y-x-z}{(x+y+z)^3},\ f_{xz} = \frac{2y}{(x+y+z)^3},\ f_{yz} = \frac{y-x-z}{(x+y+z)^3}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{d}X_2 &= \mathrm{d}\left(\frac{Z_2}{Z_1+Z_2+Z_3}\right) \\
&= -\frac{Z_2}{(Z_1+Z_2+Z_3)^2}\left[(\psi_1 - d(t))\,\mathrm{d}t + \sigma_1\,\mathrm{d}W_1\right] Z_1 \\
&\quad + \frac{Z_1+Z_3}{(Z_1+Z_2+Z_3)^2}\left[\psi_2\,\mathrm{d}t + \sigma_2\,\mathrm{d}W_2\right] Z_2 - \frac{Z_2}{(Z_1+Z_2+Z_3)^2}\left[\psi_3\,\mathrm{d}t + \sigma_3\,\mathrm{d}W_3\right] Z_3 \\
&\quad + \frac{1}{2}\frac{2Z_2}{(Z_1+Z_2+Z_3)^3} Z_1^2\sigma_1^2\,\mathrm{d}t - \frac{1}{2}\frac{2(Z_1+Z_3)}{(Z_1+Z_2+Z_3)^3} Z_2^2\sigma_2^2\,\mathrm{d}t + \frac{1}{2}\frac{2Z_2}{(Z_1+Z_2+Z_3)^3} Z_3^2\sigma_3^2\,\mathrm{d}t \\
&= \left[-X_1\psi_1 + X_1 d(t) + (X_1+X_3)\psi_2 - X_3\psi_3\right] X_2\,\mathrm{d}t \\
&\quad + \left[-X_1\sigma_1\,\mathrm{d}W_1 + (X_1+X_3)\sigma_2\,\mathrm{d}W_2 - X_3\sigma_3\,\mathrm{d}W_3\right] X_2 \\
&\quad - \left[-X_1^2\sigma_1^2 + (X_1+X_3)X_2\sigma_2^2 - X_3^2\sigma_3^2\right] X_2\,\mathrm{d}t \\
&= \left[X_2(\psi_2 - \langle\psi\rangle) + X_2X_1 d(t)\right]\,\mathrm{d}t + \left(\sigma_2\,\mathrm{d}W_2 - \sum_{j=1}^{3} X_j\sigma_j\,\mathrm{d}W_j\right) X_2 \\
&\quad - \left(\sigma_2^2 X_2 - \sum_{j=1}^{3} X_j^2\sigma_j^2\right) X_2\,\mathrm{d}t.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{d}X_3 &= \left[X_3(\psi_3 - \langle\psi\rangle) + X_3X_1 d(t)\right]\,\mathrm{d}t + \left(\sigma_3\,\mathrm{d}W_3 - \sum_{j=1}^{3} X_j\sigma_j\,\mathrm{d}W_j\right) X_3 \\
&\quad - \left(\sigma_3^2 X_3 - \sum_{j=1}^{3} X_j^2\sigma_j^2\right) X_3\,\mathrm{d}t.
\end{aligned}
$$

Hence the controlled stochastic replicator dynamics is:

$$
\begin{cases}
\mathrm{d}X_1 = & \left[X_1(\psi_1 - \langle\psi\rangle) - X_1(1-X_1)d(t)\right]\,\mathrm{d}t + \left(\sigma_1\,\mathrm{d}W_1 - \sum\limits_{j=1}^{3} X_j\sigma_j\,\mathrm{d}W_j\right) X_1 \\
& - \left(\sigma_1^2 X_1 - \sum\limits_{j=1}^{3} X_j^2\sigma_j^2\right) X_1\,\mathrm{d}t, \\
\mathrm{d}X_2 = & \left[X_2(\psi_2 - \langle\psi\rangle) + X_2X_1 d(t)\right]\,\mathrm{d}t + \left(\sigma_2\,\mathrm{d}W_2 - \sum\limits_{j=1}^{3} X_j\sigma_j\,\mathrm{d}W_j\right) X_2 \\
& - \left(\sigma_2^2 X_2 - \sum\limits_{j=1}^{3} X_j^2\sigma_j^2\right) X_2\,\mathrm{d}t, \\
\mathrm{d}X_3 = & \left[X_3(\psi_3 - \langle\psi\rangle) + X_3X_1 d(t)\right]\,\mathrm{d}t + \left(\sigma_3\,\mathrm{d}W_3 - \sum\limits_{j=1}^{3} X_j\sigma_j\,\mathrm{d}W_j\right) X_3 \\
& - \left(\sigma_3^2 X_3 - \sum\limits_{j=1}^{3} X_j^2\sigma_j^2\right) X_3\,\mathrm{d}t.
\end{cases}
\tag{3.6}
$$

Derivation of SDEs in reduced coordinates

We now derive the SDEs in the reduced coordinates $(Q,\ P)$. Since $X_1 + X_2 + X_3 = 1$ in the stochastic case as well, we can again take advantage of this and transform the 3-dimensional system into a 2-dimensional one. Let $P(t) = X_1(t)$ and $Q(t) = \dfrac{X_3(t)}{X_2(t) + X_3(t)}$ as in §3.1.1.

Then with $X_2 = (1-P)(1-Q)$ and $X_3 = (1-P)Q$, we have

$$\begin{aligned}
\mathrm{d}P = &\Big[P(\psi_1 - \langle\psi\rangle) - P(1-P)d(t)\Big]\ \mathrm{d}t \\
&+ \left(\sigma_1\ \mathrm{d}W_1 - \Big\{P\sigma_1\ \mathrm{d}W_1 + (1-P)(1-Q)\sigma_2\ \mathrm{d}W_2 + (1-P)Q\sigma_3\ \mathrm{d}W_3\Big\}\right)P \\
&- \Big[\sigma_1^2 P - (\sigma_1^2 P^2 + \sigma_2^2(1-P)^2(1-Q)^2 + \sigma_3^2(1-P)^2Q^2)\Big]P\ \mathrm{d}t.
\end{aligned}$$

Let

$$\begin{cases}
\phi(Q,P;\mathrm{d}W_1,\mathrm{d}W_2,\mathrm{d}W_3) &= P\sigma_1\ \mathrm{d}W_1 + (1-P)(1-Q)\sigma_2\ \mathrm{d}W_2 + (1-P)Q\sigma_3\ \mathrm{d}W_3, \\
\lambda(Q,P) &= \sigma_1^2P^2 + \sigma_2^2(1-P)^2(1-Q)^2 + \sigma_3^2(1-P)^2Q^2.
\end{cases}$$

Then $\phi^2 = \lambda(Q,P)\ \mathrm{d}t$. Therefore, the SDE for $P(t)$ can be rewritten as

$$\mathrm{d}P = [P(\psi_1 - \langle\psi\rangle) - P(1-P)d(t)]\ \mathrm{d}t + \Big[\sigma_1\ \mathrm{d}W_1 - \phi\Big]P - \Big[\sigma_1^2P - \lambda\Big]\ P\ \mathrm{d}t. \tag{3.7}$$

By Itô's lemma, for multivariate function $f(X,Y)$,

$$\mathrm{d}f = f_x\ \mathrm{d}X + f_y\ \mathrm{d}Y + \frac{1}{2}f_{xx}\ (\mathrm{d}X)^2 + \frac{1}{2}f_{yy}\ (\mathrm{d}Y)^2 + f_{xy}\ (\mathrm{d}X)(\mathrm{d}Y)$$

For $Q = X_3/(X_2+X_3)$, we have $f(x,y) = \dfrac{x}{x+y}$, then

$$f_x = \frac{y}{(x+y)^2},\ f_y = -\frac{x}{(x+y)^2},\ f_{xx} = -\frac{2y}{(x+y)^3},\ f_{yy} = \frac{2x}{(x+y)^3},\ f_{xy} = \frac{x-y}{(x+y)^3}.$$

Hence,

$$
\begin{aligned}
\mathrm{d}Q &= \mathrm{d}\left(\frac{X_3}{X_2+X_3}\right) \\
&= \frac{X_2}{(X_2+X_3)^2}\,\mathrm{d}X_3 - \frac{X_3}{(X_2+X_3)^2}\,\mathrm{d}X_2 - \frac{1}{2}\frac{2X_2}{(X_2+X_3)^3}\,(\mathrm{d}X_3)^2 + \frac{1}{2}\frac{2X_2}{(X_2+X_3)^3}\,(\mathrm{d}X_2)^2 \\
&\quad + \frac{X_3-X_2}{(X_2+X_3)^3}\,(\mathrm{d}X_2)(\mathrm{d}X_3) \\
&= \frac{(1-P)(1-Q)}{(1-P)^2}\,\mathrm{d}X_3 - \frac{(1-P)Q}{(1-P)^2}\,\mathrm{d}X_2 - \frac{(1-P)(1-Q)}{(1-P)^3}\,(\mathrm{d}X_3)^2 + \frac{(1-P)Q}{(1-P)^3}\,(\mathrm{d}X_2)^2 \\
&\quad + \frac{(1-P)Q-(1-P)(1-Q)}{(1-P)^3}\,(\mathrm{d}X_2)(\mathrm{d}X_3) \\
&= \frac{1-Q}{1-P}\,\mathrm{d}X_3 - \frac{Q}{1-P}\,\mathrm{d}X_2 - \frac{1-Q}{(1-P)^2}\,(\mathrm{d}X_3)^2 + \frac{Q}{(1-P)^2}\,(\mathrm{d}X_2)^2 + \frac{2Q-1}{(1-P)^2}\,(\mathrm{d}X_2)(\mathrm{d}X_3).
\end{aligned}
$$

Now substituting (3.6) and (3.7) into the above equation, we have

$$
\begin{aligned}
\mathrm{d}Q = {} & \frac{1-Q}{1-P}\Big[(1-P)Q(\psi_3-\langle\psi\rangle)\,\mathrm{d}t + d(t)PQ(1-P)\,\mathrm{d}t \\
& \qquad + (1-P)Q(\sigma_3\,\mathrm{d}W_3-\phi) - (1-P)Q(\sigma_3^2(1-P)Q-\lambda)\,\mathrm{d}t\Big] \\
& - \frac{Q}{1-P}\Big[(1-P)(1-Q)(\psi_2-\langle\psi\rangle)\,\mathrm{d}t + d(t)P(1-P)(1-Q)\,\mathrm{d}t \\
& \qquad + (1-P)(1-Q)(\sigma_2\,\mathrm{d}W_2-\phi) - (1-P)(1-Q)(\sigma_2^2(1-P)(1-Q)-\lambda)\mathrm{d}t\Big] \\
& - \frac{1-Q}{(1-P)^2}\left[(1-P)^2Q^2(\sigma_3^2\,\mathrm{d}t+\phi^2-2(1-P)Q\sigma_3^2\,\mathrm{d}t)\right] \\
& + \frac{Q}{(1-P)^2}\left[(1-P)^2(1-Q)^2(\sigma_2^2\,\mathrm{d}t+\phi^2-2(1-P)(1-Q)\sigma_2^2\,\mathrm{d}t)\right] \\
& + \frac{2Q-1}{(1-P)^2}\left[(1-P)^2Q(1-Q)(\phi^2-(1-P)(1-Q)\sigma_2^2\,\mathrm{d}t-(1-P)Q\sigma_3^3\,\mathrm{d}t)\right] \\
= {} & Q(1-Q)(\psi_3-\psi_2)\,\mathrm{d}t + \cancel{d(t)PQ(1-Q)\,\mathrm{d}t} - \cancel{d(t)PQ(1-Q)\,\mathrm{d}t} \\
& + Q(1-Q)(\sigma_3\,\mathrm{d}W_3-\sigma_2\,\mathrm{d}W_2) - Q^2(1-Q)\sigma_3^2\,\mathrm{d}t + Q(1-Q)^2\sigma_2^2\,\mathrm{d}t \\
= {} & Q(1-Q)(\psi_3-\psi_2)\,\mathrm{d}t + Q(1-Q)(\sigma_3\,\mathrm{d}W_3-\sigma_2\,\mathrm{d}W_2) + Q(1-Q)\left[(1-Q)\sigma_2^2-Q\sigma_3^2\right]\,\mathrm{d}t.
\end{aligned}
$$

Recall from §3.1.1 that

$$
\begin{cases}
P(\psi_1-\langle\psi\rangle) = P(1-P)\left(\dfrac{b_a}{n+1}-(b_v-c)Q\right), \\
\psi_3-\psi_2 = \dfrac{b_v}{n+1}\left[\sum\limits_{k=0}^{n}P^k\right]-c.
\end{cases}
$$

Thus, the controlled reduced stochastic system provided in Chapter 2 Box 5 is

$$\begin{cases} \mathrm{d}Q = Q(1-Q)\left(\dfrac{b_v}{n+1}\left[\sum\limits_{k=0}^{n} P^k\right] - c\right)\,\mathrm{d}t + Q(1-Q)\Big(\sigma_3\,\mathrm{d}W_3 - \sigma_2\,\mathrm{d}W_2\Big) \\ \qquad\quad +\Big[(1-Q)\sigma_2^2 - Q\sigma_3^2\Big]Q(1-Q)\,\mathrm{d}t, \\ \mathrm{d}P = P(1-P)\left(\dfrac{b_a}{n+1} - (b_v-c)Q - d(t)\right)\,\mathrm{d}t \\ \qquad\quad +\sigma_1 P(1-P)\,\mathrm{d}W_1 + \sigma_2 P(1-P)(1-Q)\,\mathrm{d}W_2 + \sigma_3 P(1-P)Q\,\mathrm{d}W_3 \\ \qquad\quad -\Big[\sigma_1^2 P - \sigma_2^2(1-P)(1-Q)^2 - \sigma_3^2(1-P)Q^2\Big]P(1-P)\,\mathrm{d}t. \end{cases} \tag{3.8}$$

Since $Q,\ P \in [0,1]$ and $d \in [0, d_{\max}]$, the drift coefficient function of the above system is Lipschitz with respect to Q, P, and d, and the diffusion coefficient function of the above system is Lipschitz with respect to Q and P. Consequently, with any fixed initial data (q_0, p_0), the above system has a unique strong solution on our reference probability space ν [52, 23].

3.2.3 Derivation of controlled perturbed system for the SR model

Next, we provide detailed derivations for the controlled stochastic cancer dynamics based on the Sensitive-Resistant competition model (labeled (2.20) in Chapter 2).

As mentioned in Chapter 2, for this example we assume a single (1D) Brownian motion B_t affecting the intrinsic growth rates of both S and R simultaneously, with expectation $(g_{\mathrm{S}}, g_{\mathrm{R}})$ and volatilities $(\sigma_{\mathrm{S}}, \sigma_{\mathrm{R}})$. This results in the following SDEs for $(Z_{\mathrm{S}}, Z_{\mathrm{R}})$:

$$\begin{aligned} \mathrm{d}Z_{\mathrm{S}} &= (g_{\mathrm{S}}\mathrm{d}t + \sigma_{\mathrm{S}}\,\mathrm{d}B_t)(1 - Z_{\mathrm{S}} - mZ_{\mathrm{R}})Z_{\mathrm{S}} - \alpha d(t) Z_{\mathrm{S}}\mathrm{d}t \\ &= [g_{\mathrm{S}}(1 - Z_{\mathrm{S}} - mZ_{\mathrm{R}}) - \alpha d(t)]Z_{\mathrm{S}}\mathrm{d}t + \sigma_{\mathrm{S}} Z_{\mathrm{S}}(1 - Z_{\mathrm{S}} - mZ_{\mathrm{R}})\mathrm{d}B_t. \\ \mathrm{d}Z_{\mathrm{R}} &= (g_{\mathrm{R}}\mathrm{d}t + \sigma_{\mathrm{R}}\,\mathrm{d}B_t)(1 - Z_{\mathrm{S}} - mZ_{\mathrm{R}})Z_{\mathrm{R}} - \beta C Z_{\mathrm{S}} Z_{\mathrm{R}}\mathrm{d}t \\ &= [g_{\mathrm{R}}(1 - Z_{\mathrm{S}} - mZ_{\mathrm{R}}) - \beta C Z_{\mathrm{S}}]Z_{\mathrm{R}}\mathrm{d}t + \sigma_{\mathrm{R}} Z_{\mathrm{R}}(1 - Z_{\mathrm{S}} - mZ_{\mathrm{R}})\mathrm{d}B_t. \end{aligned}$$

Let $P(t) = Z_{\rm S}(t) + mZ_{\rm R}(t)$ and $Q(t) = \dfrac{Z_{\rm S}(t)}{P(t)}$ as in §3.1.2. It follows that $Z_{\rm S} = QP$, $mZ_{\rm R} = (1-Q)P$, and hence

$$\begin{aligned}
\mathrm{d}P &= \mathrm{d}Z_{\rm S} + m\mathrm{d}Z_{\rm R} \\
&= (g_{\rm S}\mathrm{d}t + \sigma_{\rm S}\,\mathrm{d}B_t)(1 - Z_{\rm S} - mZ_{\rm R})Z_{\rm S} - \alpha d(t) Z_{\rm S}\mathrm{d}t \\
&= [g_{\rm S}(1 - Z_{\rm S} - mZ_{\rm R}) - \alpha d(t)]Z_{\rm S}\mathrm{d}t + \sigma_{\rm S}Z_{\rm S}(1 - Z_{\rm S} - mZ_{\rm R})\mathrm{d}B_t \\
&\quad + [g_{\rm R}(1 - Z_{\rm S} - mZ_{\rm R}) - \beta C Z_{\rm S}]Z_{\rm R}\mathrm{d}t + \sigma_{\rm R}Z_{\rm R}(1 - Z_{\rm S} - mZ_{\rm R})\mathrm{d}B_t \\
&= \Big[g_{\rm S}QP(1-P) - \alpha QPd(t) + g_{\rm R}(1-Q)P(1-P) - \beta C P^2 Q(1-Q)\Big]\mathrm{d}t \\
&\quad + \Big[\sigma_{\rm S}QP(1-P) + \sigma_{\rm R}(1-Q)P(1-P)\Big]\mathrm{d}B_t \\
&= \Big[P(1-P)(g_{\rm S}Q + g_{\rm R}(1-Q)) - \alpha QPd(t) - \beta C P^2 Q(1-Q)\Big]\mathrm{d}t \\
&\quad + P(1-P)\Big[\sigma_{\rm S}Q + \sigma_{\rm R}(1-Q)\Big]\mathrm{d}B_t
\end{aligned}$$

By Itô's lemma, for multivariate function $f(X, Y)$,

$$\mathrm{d}f = f_x\,\mathrm{d}X + f_y\,\mathrm{d}Y + \frac{1}{2}f_{xx}\,(\mathrm{d}X)^2 + \frac{1}{2}f_{yy}\,(\mathrm{d}Y)^2 + f_{xy}\,(\mathrm{d}X)(\mathrm{d}Y)$$

For $Q = Z_{\rm S}/P$, we have $f(x, y) = \dfrac{x}{y}$, then

$$f_x = \frac{1}{y},\ f_y = -\frac{x}{y^2},\ f_{xx} = 0,\ f_{yy} = \frac{2x}{y^3},\ f_{xy} = -\frac{1}{y^2}.$$

Hence,

$$
\begin{aligned}
\mathrm{d}Q &= \mathrm{d}\left(\frac{Z_{\mathrm{S}}}{P}\right) \\
&= \frac{1}{P}\mathrm{d}Z_{\mathrm{S}} - \frac{Z_{\mathrm{S}}}{P^2}\mathrm{d}P + \frac{Z_{\mathrm{S}}}{P^3}(\mathrm{d}P)^2 - \frac{1}{P^2}(\mathrm{d}Z_{\mathrm{S}})(\mathrm{d}P) \\
&= \frac{1}{P}\Big\{\Big[g_{\mathrm{S}}QP(1-P) - \alpha QPd(t)\Big]\mathrm{d}t + \sigma_{\mathrm{S}}QP(1-P)\mathrm{d}B_t\Big\} \\
&\quad - \frac{Q}{P}\Big\{\Big[P(1-P)(g_{\mathrm{S}}Q + g_{\mathrm{R}}(1-Q)) - \alpha QPd - \beta CP^2Q(1-Q)\Big]\mathrm{d}t \\
&\qquad\qquad + P(1-P)\Big[\sigma_{\mathrm{S}}Q + \sigma_{\mathrm{R}}(1-Q)\Big]\mathrm{d}B_t\Big\} \\
&\quad + \frac{Q}{P^2}\Big\{P^2(1-P)^2\Big[\sigma_{\mathrm{S}}Q + \sigma_{\mathrm{R}}(1-Q)\Big]^2\Big\}\mathrm{d}t \\
&\quad - \frac{1}{P^2}\Big\{\sigma_{\mathrm{S}}QP(1-P)\cdot P(1-P)\Big[\sigma_{\mathrm{S}}Q + \sigma_{\mathrm{R}}(1-Q)\Big]\Big\}\mathrm{d}t \\
&= \Big[(1-Q)Q(1-P)(g_{\mathrm{S}} - g_{\mathrm{R}}) + \beta CPQ^2(1-Q) - \alpha Q(1-Q)d(t)\Big]\mathrm{d}t \\
&\quad + Q(1-Q)(1-P)(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}})\mathrm{d}B_t \\
&\quad + Q(1-Q)(1-P)^2\Big[\sigma_{\mathrm{R}}^2(1-Q) - \sigma_{\mathrm{S}}^2Q + \sigma_{\mathrm{S}}\sigma_{\mathrm{R}}\Big]\mathrm{d}t
\end{aligned}
$$

Thus, the controlled reduced stochastic system provided in Chapter 2 Box 6 is

$$
\begin{cases}
\mathrm{d}Q = Q(1-Q)\Big\{(1-P)(g_{\mathrm{S}} - g_{\mathrm{R}}) - \alpha d(t) + \beta CQP + (1-P)^2\Big[\sigma_{\mathrm{R}}^2(1-Q) - \sigma_{\mathrm{S}}^2Q + \sigma_{\mathrm{S}}\sigma_{\mathrm{R}}\Big]\Big\}\mathrm{d}t \\
\qquad\quad +Q(1-Q)(1-P)(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}})\,\mathrm{d}B_t, \\
\mathrm{d}P = \Big[P(1-P)\Big(g_{\mathrm{S}}Q + g_{\mathrm{R}}(1-Q)\Big) - \alpha QPd(t) - \beta CP^2Q(1-Q)\Big]\mathrm{d}t \\
\qquad\quad +P(1-P)\Big[\sigma_{\mathrm{S}}Q + \sigma_{\mathrm{R}}(1-Q)\Big]\,\mathrm{d}B_t.
\end{cases}
\tag{3.9}
$$

By the same argument as in §3.2.2, with any fixed initial data (q_0, p_0), the above system has a unique strong solution on our reference probability space ν [52, 23].

3.3 Derivation of Hamilton-Jacobi-Bellman equations

We derive the threshold-awareness HJB equation (labeled (2.11) in Chapter 2 Box 4) via tools of dynamic programming in §3.3.2. We also summarize the objective and the first-order HJB equation in the deterministic case since we extensively compare our threshold-aware optimal policy to the deterministic-optimal policy in Chapter 2. (The derivation of PDEs in Box 3 is omitted; it follows the standard methods in [62].)

3.3.1 The first-order HJB equation in the deterministic case

In this section, we derive the first-order HJB PDE (2.4) provided in Chapter 2 Box 1. As mentioned there, we consider a general ODE describing the deterministic cancer dynamics:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{d}), \qquad \boldsymbol{x}(0) = \boldsymbol{\xi}. \tag{3.10}$$

We follow the same definitions (but now in the deterministic setting) of the terminal time, terminal cost function g, running cost function K and the total cost function $\mathcal{J}$ as presented in §3.2.1. The objective (value function) minimizing the (deterministic) total cost $\mathcal{J}$ over all available (deterministic) policies is now defined as

$$u(\boldsymbol{\xi}) = \inf_{\boldsymbol{d}(\cdot)} \mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)\Big), \tag{3.11}$$

and a policy $\boldsymbol{d}_\star(\cdot)$ is optimal if $u(\boldsymbol{\xi}) = \mathcal{J}\Big(\boldsymbol{\xi}, \boldsymbol{d}_\star(\cdot)\Big)$. To simplify the derivation, we assume that such $\boldsymbol{d}_\star$ exists. (Otherwise, a similar argument can be built using ϵ-suboptimal policies.) For a sufficiently small $\theta > 0$, by Bellman's Optimality Principle we have

$$\begin{aligned} u(\boldsymbol{\xi}) &= \int_0^\theta K(\boldsymbol{x}(\tau), \boldsymbol{d}_\star(\tau))\, \mathrm{d}\tau + u(\boldsymbol{x}(\theta)) \\ &= \theta K(\boldsymbol{\xi}, \boldsymbol{d}_\star(0)) + \Big[u(\boldsymbol{\xi}) + \theta \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{d}_\star(0)) + \Big] + o(\theta); \\ 0 &= \theta K(\boldsymbol{\xi}, \boldsymbol{d}_\star(0)) + \theta \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{d}_\star(0)) + o(\theta). \end{aligned}$$

Now dividing both sides by θ and sending θ to 0, we obtain

$$0 = K(\boldsymbol{\xi}, \boldsymbol{d}_\star(0)) + \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{d}_\star(0)).$$

Notice that the above equation involves $\boldsymbol{d}_\star(0)$ only. It is then natural to switch to a state-dependent optimal control in feedback form. The HJB equation for (3.11) is then obtained by maximizing over $\boldsymbol{d} = \boldsymbol{d}_\star(0) \in \mathcal{D}$. By demanding the above equation holds for all $\boldsymbol{\xi} \in [0,1]^n \setminus \Delta$, the PDE can be written as:

$$0 = \min_{\boldsymbol{d} \in \mathcal{D}} \Big\{ K(\boldsymbol{\xi}, \boldsymbol{d}) + \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{d}) \Big\}. \tag{3.12}$$

Recall from (3.2) that Eq. (3.10) for our Example 1 in component-wise form is

$$\begin{cases} \dot{q} = q(1-q)\left(\dfrac{b_v}{n+1} \left[\sum\limits_{k=0}^{n} p^k \right] - c \right), \\ \dot{p} = p(1-p)\left(\dfrac{b_a}{n+1} - (b_v - c)q - d \right). \end{cases}$$

It follows that Eq. (3.12) in component-wise form for Example 1 is

$$\begin{aligned} \min_{d \in [0, d_{\max}]} \Big[\Big(1 - u_p p(1-p) \Big) d \Big] \; &+ \; u_q q(1-q) \Big(\frac{b_v}{n+1} \sum_{k=0}^{n} p^k - c \Big) \\ &+ \; u_p p(1-p) \Big(\frac{b_a}{n+1} - q(b_v - c) \Big) + \delta \;\; = \;\; 0. \end{aligned} \tag{3.13}$$

The linear dependence on d of the minimized expression yields the *bang-bang* property:

$$d_\star(q,p) = \begin{cases} d_{\max}, & \text{if } \Big(1 - u_p p(1-p) \Big) < 0; \\ 0, & \text{otherwise.} \end{cases} \tag{3.14}$$

Similarly, Eq. (3.10) for our Example 2 from (3.4) in component-wise form is

$$\begin{cases} \dot{q} = (1-p)q(1-q)(g_{\mathrm{S}} - g_{\mathrm{R}}) + \beta C p q^2 (1-q) - \alpha q(1-q)d, \\ \dot{p} = p(1-p)[g_{\mathrm{S}} q + g_{\mathrm{R}}(1-q)] - \beta C p^2 q(1-q) - \alpha q p \, d. \end{cases}$$

Therefore, Eq. (3.12) in component-wise form for Example 2 is

$$\min_{d\in[0,d_{\max}]}\Big[\ \Big(1-u_q\alpha q(1-q)-u_p\alpha qp\Big)d\Big]\ +\ u_q q(1-q)\Big[(1-p)(g_{\rm S}-g_{\rm R})+\beta Cqp\Big]+ \\ u_p\Big[p(1-p)\Big(g_{\rm S}q+g_{\rm R}(1-q)\Big)-\beta Cp^2q(1-q)\Big]\ =\ 0. \quad (3.15)$$

Again, the linear dependence on d yields the *bang-bang* property:

$$d_\star(q,p)=\begin{cases} d_{\max}, & \text{if } \Big(1-u_q\alpha q(1-q)-u_p\alpha qp\Big)<0;\\ 0, & \text{otherwise.}\end{cases} \quad (3.16)$$

3.3.2 Derivation of the threshold-awareness HJB equation

Given the fixed reference probability space ν defined in §3.2.1, we define the set of all *admissible* controls as

$$\mathcal{A}_\nu:=\Big\{\boldsymbol{d}(\cdot):[0,\infty)\times\Omega\to\mathcal{D}\ \mid\ \boldsymbol{d}(\cdot)\text{ is }\mathcal{F}_t-\text{progressively measurable}\Big\}.$$

Ideally, we would want to consider controls in *feedback form*; i.e., $\boldsymbol{d}$ would be determined based on the current tumor configuration and perhaps the amount of drugs administered so far. But for technical reasons, we will first use *progressively measurable* open-loop controls, which we define below, to derive dynamic programming equations, and only then show that an optimal control can be found in feedback form.

To work under the framework of dynamic programming, we define a *value function* $v(\boldsymbol{\xi},\bar{s})$ encoding the maximal probability of reaching $\Delta_{\rm succ}$ while keeping $\mathcal{J}$ from exceeding the threshold/budget value $\bar{s}$:

$$v(\boldsymbol{\xi},\bar{s})=\sup_{\boldsymbol{d}(\cdot)\in\mathcal{A}_\nu}\mathbb{P}\Big(\mathcal{J}(\boldsymbol{\xi},\boldsymbol{d}(\cdot))\le\bar{s}\Big). \quad (3.17)$$

Any policy $\boldsymbol{d}_*(\cdot)$ is called optimal if $v(\boldsymbol{\xi},\bar{s})=\mathbb{P}\Big(\mathcal{J}(\boldsymbol{\xi},\boldsymbol{d}_*(\cdot))\le\bar{s}\Big)$. Notice that $\mathbb{P}\Big(\mathcal{J}(\boldsymbol{\xi},\boldsymbol{d}(\cdot))\le\bar{s}\Big)$ can be treated as $\mathbb{E}\Big[\mathbf{1}_{\{\mathcal{J}(\boldsymbol{\xi},\boldsymbol{d}(\cdot))\le\bar{s}\}}\Big]$, where $\mathbf{1}_{\{Y\le y\}}$ is the indicator function such that it has value 1 if $Y\le y$ and 0 otherwise.

As explained in Chapter 2, we introduce a new component of the state, the budget variable s. The (random) ODE describing its rate of change is

$$\dot{s} = -K(\boldsymbol{X}(t), d(t)), \qquad s(0) = \bar{s}.$$

Notice that $s(t)$ is strictly decreasing as time progresses. Thus, to ensure a nonnegative budget, we define a new terminal set

$$\widehat{\Delta} = \Big\{(x, s) \in [0,1]^n \times [0, \bar{\mathcal{S}}] \;\mid\; x \in \Delta \;\text{ or }\; s = 0\Big\}$$

and correspondingly a new terminal time

$$\widehat{T} = \widehat{T}(\boldsymbol{\xi}, \bar{s}, \boldsymbol{d}(\cdot)) := \inf\Big\{t \in \mathbb{R}_+ \;\mid\; \Big(\boldsymbol{X}(t),\; s(t)\Big) \in \widehat{\Delta}, \quad \boldsymbol{X}(0) = \boldsymbol{\xi},\; s(0) = \bar{s}\Big\}.$$

We define

$$\Psi\Big(\boldsymbol{X}(\widehat{T})\Big) = \begin{cases} 1, \text{ if } \boldsymbol{X}(\widehat{T}) \in \Delta_{\text{succ}}, \\ 0, \text{ otherwise.} \end{cases}$$

as the last step of transforming our problem into a Mayer form [61].

Notice that regardless of random realizations, the largest possible $\widehat{T}$ is $\bar{s}/\delta$. In principle, we can recast it as a finite-horizon problem with $[0, \bar{s}/\delta]$ representing the horizon.

Now the original problem is equivalent to

$$v(\boldsymbol{\xi}, \bar{s}) = \max_{\boldsymbol{d}(\cdot)} \; \mathbb{E}^0\Big[\Psi(\boldsymbol{X}(\widehat{T}))\Big], \tag{3.18}$$

where

$$\mathbb{E}^0\big[\;\cdot\;\big] = \mathbb{E}\big[\;\cdot\;\mid \text{Initial Conditions}\big].$$

This is now an exit-time problem in Mayer form over the domain $(\boldsymbol{\xi}, \bar{s}) \in [0,1]^n \times [0, \bar{\mathcal{S}}]$. We can therefore apply the stochastic dynamic programming principle [62] to derive the HJB PDE.

Assume the optimal $\boldsymbol{d}_*(\cdot)$ exists, and let $\widehat{T}_* = \widehat{T}(\boldsymbol{\xi}, \bar{s}, \boldsymbol{d}_*)$. The stochastic dynamic programming principle (Chapter V.2 in [62]) states for any (nonrandom) $\theta \in (0, \bar{s}/\delta)$,

$$v((\boldsymbol{\xi}, \bar{s}) = \mathbb{E}^0 \left[v(\boldsymbol{X}(\widehat{T}_* \wedge \theta), s(\widehat{T}_* \wedge \theta)) \right], \tag{3.19}$$

where $a \wedge b = \min(a, b)$.

We now provide a formal derivation of the PDE that v has to satisfy if it is sufficiently smooth. Notice that Eq. (3.19) can be rewritten as

$$v((\boldsymbol{\xi}, \bar{s}) = \mathbb{E}^0 \left[\mathbf{1}_{\{\widehat{T}_* < \theta\}} v(\boldsymbol{X}(\widehat{T}_*), s(\widehat{T}_*)) + \mathbf{1}_{\{\widehat{T}_* \geq \theta\}} v(\boldsymbol{X}(\theta), s(\theta)) \right].$$

Seeking stochastic Taylor expansion of $v(\boldsymbol{X}(\theta), s(\theta))$ around $\theta = 0$, we have

$$\begin{aligned} v(\boldsymbol{X}(\theta), s(\theta)) - v((\boldsymbol{\xi}, \bar{s}) = &\int_0^\theta \nabla v(\boldsymbol{X}(\tau), s(\tau)) \cdot \mathrm{d}\boldsymbol{X}(\tau) + \int_0^\theta \frac{\partial}{\partial s} v(\boldsymbol{X}(\tau), s(\tau)) \, \mathrm{d}s(\tau) \\ &+ \frac{1}{2} \sum_{i,j=1}^n \int_0^\theta \frac{\partial^2}{\partial \xi_i \partial \xi_j} v(\boldsymbol{X}(\tau), s(\tau)) \Big(\mathrm{d}X_i(\tau) \mathrm{d}X_j(\tau) \Big) + o(\theta). \end{aligned}$$

Let $\boldsymbol{B} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$. It follows that

$$\begin{aligned} & v(\boldsymbol{X}(\theta), s(\theta)) - v(\boldsymbol{\xi}, \bar{s}) \\ = &\int_0^\theta \nabla v(\boldsymbol{X}(\tau), s(\tau)) \cdot [\boldsymbol{a}(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau)) \, \mathrm{d}\tau + \boldsymbol{\Sigma}(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau)) \, \mathrm{d}\boldsymbol{W}(\tau)] \\ &- \int_0^\theta K(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau)) \frac{\partial}{\partial s} v(\boldsymbol{X}(\tau), s(\tau)) \, \mathrm{d}\tau \\ &+ \frac{1}{2} \sum_{i,j=1}^n \int_0^\theta \frac{\partial^2}{\partial \xi_i \partial \xi_j} v(\boldsymbol{X}(\tau), s(\tau)) \boldsymbol{B}(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau))_{i,j} \, \mathrm{d}\tau + o(\theta). \end{aligned}$$

Since Itô integrals with bounded functions have zero mean, we have

$$\begin{aligned} & \mathbb{E}^0 \left[\mathbf{1}_{\{\widehat{T}_* \geq \theta\}} v(\boldsymbol{X}(\theta), s(\theta)) \right] \\ = \, & v(\boldsymbol{\xi}, \bar{s}) \mathbb{P}(\widehat{T}_* \geq \theta) \\ + \, & \mathbb{E}^0 \left[\mathbf{1}_{\{\widehat{T}_* \geq \theta\}} \left\{ \int_0^\theta \nabla v(\boldsymbol{X}(\tau), s(\tau)) \cdot \boldsymbol{a}(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau)) \, \mathrm{d}\tau \right. \right. \\ & \qquad - \int_0^\theta K(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau)) \frac{\partial}{\partial s} v(\boldsymbol{X}(\tau), s(\tau)) \, \mathrm{d}\tau \\ & \qquad \left. \left. + \frac{1}{2} \sum_{i,j=1}^n \int_0^\theta \frac{\partial^2}{\partial \xi_i \partial \xi_j} v(\boldsymbol{X}(\tau), s(\tau)) \boldsymbol{B}(\boldsymbol{X}(\tau), \boldsymbol{d}_*(\tau))_{i,j} \, \mathrm{d}\tau + o(\theta) \right\} \right]. \end{aligned}$$

Therefore,

$$\begin{aligned}v(\boldsymbol{\xi},\bar{s}) = \mathbb{E}^0\left[\mathbf{1}_{\{\widehat{T}_*<\theta\}}v(\boldsymbol{X}(\widehat{T}),s(\widehat{T}))\right] + v(\boldsymbol{\xi},\bar{s})\mathbb{P}(\widehat{T}_*\geq\theta)\\ + \mathbb{E}^0\left[\mathbf{1}_{\{\widehat{T}_*\geq\theta\}}\left\{\int_0^\theta \nabla v(\boldsymbol{X}(\tau),s(\tau))\cdot\boldsymbol{a}(\boldsymbol{X}(\tau),\boldsymbol{d}_*(\tau))\,\mathrm{d}\tau\right.\right.\\ -\int_0^\theta K(\boldsymbol{X}(\tau),\boldsymbol{d}_*(\tau))\frac{\partial}{\partial s}v(\boldsymbol{X}(\tau),s(\tau))\,\mathrm{d}\tau\\ \left.\left.+\frac{1}{2}\sum_{i,j=1}^n\int_0^\theta\frac{\partial^2}{\partial\xi_i\partial\xi_j}v(\boldsymbol{X}(\tau),s(\tau))\boldsymbol{B}(\boldsymbol{X}(\tau),\boldsymbol{d}_*(\tau))_{i,j}\,\mathrm{d}\tau\right\}\right] + o(\theta).\end{aligned}$$

Now dividing both sides by θ and sending $\theta\to 0$, we have

$$\begin{aligned}v(\boldsymbol{\xi},\bar{s})\lim_{\theta\to0}\frac{1-\mathbb{P}(\widehat{T}_*\geq\theta)}{\theta} = \lim_{\theta\to0}\mathbb{E}^0\left[\frac{\mathbf{1}_{\{\widehat{T}_*<\theta\}}}{\theta}v(\boldsymbol{X}(\widehat{T}_*),s(\widehat{T}_*))\right]\\ +\lim_{\theta\to0}\mathbb{E}^0\left[\frac{\mathbf{1}_{\{\widehat{T}_*\geq\theta\}}}{\theta}\left\{\int_0^\theta\nabla v(\boldsymbol{X}(\tau),s(\tau))\cdot\boldsymbol{a}(\boldsymbol{X}(\tau),\boldsymbol{d}_*(\tau))\,\mathrm{d}\tau\right.\right.\\ -\int_0^\theta K(\boldsymbol{X}(\tau),\boldsymbol{d}_*(\tau))\frac{\partial}{\partial s}v(\boldsymbol{X}(\tau),s(\tau))\,\mathrm{d}\tau\\ \left.\left.+\frac{1}{2}\sum_{i,j=1}^n\int_0^\theta\frac{\partial^2}{\partial\xi_i\partial\xi_j}v(\boldsymbol{X}(\tau),s(\tau))\boldsymbol{B}(\boldsymbol{X}(\tau),\boldsymbol{d}_*(\tau))_{i,j}\,\mathrm{d}\tau\right\}\right]\end{aligned}$$

From [62, Chapter V], we know

$$\lim_{\theta\to0}\frac{\mathbb{P}(\widehat{T}_*<\theta)}{\theta}=0,$$

and hence

$$v(\boldsymbol{\xi},\bar{s})\lim_{\theta\to0}\frac{1-\mathbb{P}(\widehat{T}_*\geq\theta)}{\theta} = v(\boldsymbol{\xi},\bar{s})\lim_{\theta\to0}\frac{\mathbb{P}(\widehat{T}_*<\theta)}{\theta}=0;$$

$$\lim_{\theta\to0}\mathbb{E}^0\left[\frac{\mathbf{1}_{\{\widehat{T}_*<\theta\}}}{\theta}v(\boldsymbol{X}(\widehat{T}_*),s(\widehat{T}_*))\right]\leq\lim_{\theta\to0}\frac{\mathbb{P}(\widehat{T}_*<\theta)}{\theta}=0.$$

Since K and all components of $\boldsymbol{a}$ and $\boldsymbol{B}$ are bounded, v is assumed to be sufficiently smooth, and $\mathbf{1}_{\{\widehat{T}\geq\theta\}}\to1$ as $\theta\to0$, by the dominated convergence theorem we have

$$\begin{aligned}0=\mathbb{E}^0\Big[\nabla v(\boldsymbol{\xi},\bar{s})\cdot\boldsymbol{a}(\boldsymbol{\xi},\boldsymbol{d}_*(0,\omega)) - \frac{\partial}{\partial s}v(\boldsymbol{\xi},\bar{s})K(\boldsymbol{\xi},\boldsymbol{d}_*(0,\omega))\\ +\frac{1}{2}\sum_{i,j=1}^n\frac{\partial^2}{\partial\xi_i\partial\xi_j}v(\boldsymbol{\xi},\bar{s})\boldsymbol{B}(\boldsymbol{\xi},\boldsymbol{d}_*(0,\omega))_{i,j}\Big]\end{aligned}$$

Notice that the above equation only involves $\boldsymbol{d}_*(0)$ for every $\omega \in \Omega$. Namely, the starting optimal control value depends only on the initial state $(\boldsymbol{\xi}, \bar{s})$ rather than a specific ω. We then switch to a budget-dependent optimal control $\boldsymbol{d}_*(0)(\boldsymbol{\xi}, \bar{s})$ in feedback form. The HJB equation for (3.17) is then obtained by maximizing over $d \in [0, d_{\max}]$. By demanding the above equation holds for all $(\boldsymbol{\xi}, \bar{s}) \in ([0,1]^n \setminus \Delta) \times [0, +\infty)$, the PDE can be written as:

$$0 = \max_{\boldsymbol{d} \in \mathcal{D}} \left\{ -\frac{\partial}{\partial s} v(\boldsymbol{\xi}, s) K(\boldsymbol{\xi}, \boldsymbol{d}) + \nabla v(\boldsymbol{\xi}, s) \cdot \boldsymbol{a}(\boldsymbol{\xi}, \boldsymbol{d}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial \xi_i \partial \xi_j} v(\boldsymbol{\xi}, s) \boldsymbol{B}(\boldsymbol{\xi}, \boldsymbol{d})_{i,j} \right\}. \quad (3.20)$$

In both examples considered in Chapter 2

$$K(\boldsymbol{X}, d) = d + \delta, \text{ where } d \in [0, d_{\max}],$$
$$T = T(q_0, p_0, d(\cdot)) := \inf \Big\{ t \in \mathbb{R}_+ \ \mid \ \Big(Q(t), P(t)\Big) \in \Delta, \quad Q(0) = q_0, \ P(0) = p_0 \Big\},$$

where $\Delta = \Big\{(q,p) \in [0,1]^2 \ \mid \ p < \gamma_{\mathrm{r}} \ \text{ or } \ p > \gamma_{\mathrm{f}} \Big\}$. The terminal function is defined as

$$g\Big(\boldsymbol{X}(T)\Big) = \begin{cases} +\infty, & \text{if } P(T) > \gamma_{\mathrm{f}}, \\ 0, & \text{if } P(T) < \gamma_{\mathrm{r}}. \end{cases}$$

With the same extension to $\widehat{\Delta}$ and $\widehat{T}$, we have

$$\Psi\Big(Q(\widehat{T}), P(\widehat{T})\Big) = \begin{cases} 1, \text{ if } P(\widehat{T}) \le \gamma_{\mathrm{r}}, \\ 0, \text{ otherwise.} \end{cases}$$

Specific formulations for the EGT-based model

Recall from Chapter 2 Box 5 and derivations for Eq. (3.8) in §3.2.2 that the components for Eq. (3.5) are

$$\boldsymbol{X} = \begin{bmatrix} Q \\ P \end{bmatrix},$$

$$\boldsymbol{a}(\boldsymbol{X}, d) = \begin{bmatrix} Q(1-Q)\left\{\left(\frac{b_v}{n+1}\left[\sum_{k=0}^{n} P^k\right] - c\right) + \left[(1-Q)\sigma_2^2 - Q\sigma_3^2\right]\right\} \\ P(1-P)\left\{\left(\frac{b_a}{n+1} - (b_v - c)Q - d\right) - \left[\sigma_1^2 P - \sigma_2^2(1-P)(1-Q)^2 - \sigma_3^2(1-P)Q^2\right]\right\} \end{bmatrix},$$

$$\boldsymbol{\Sigma}(\boldsymbol{X}, d) = \begin{bmatrix} 0 & -\sigma_2 Q(1-Q) & \sigma_3 Q(1-Q) \\ \sigma_1 P(1-P) & \sigma_2 P(1-P)(1-Q) & \sigma_3 P(1-P)Q \end{bmatrix}.$$

Thus, Eq. (3.20) in component-wise form is

$$\begin{aligned} 0 = \max_{d \in [0, d_{\max}]} \Bigg\{ & -\left[\frac{\partial v}{\partial p} p(1-p) + \frac{\partial v}{\partial s}\right] d \Bigg\} - \delta \frac{\partial v}{\partial s} \\ & + \frac{\partial v}{\partial q}\left[\left(\frac{b_v}{n+1}\sum_{k=0}^{n} p^k - c\right) - q\sigma_3^2 + (1-q)\sigma_2^2\right] q(1-q) \\ & + \frac{\partial v}{\partial p}\left[\left(\frac{b_a}{n+1} - q(b_v - c)\right) - [\sigma_1^2 p - \sigma_2^2(1-p)(1-q)^2 - \sigma_3^2(1-p)q^2]\right] p(1-p) \\ & + \frac{1}{2}\frac{\partial^2 v}{\partial q^2}\left[q^2(1-q)^2(\sigma_2^2 + \sigma_3^2)\right] + \frac{1}{2}\frac{\partial^2 v}{\partial p^2}\left[\sigma_1^2 + (1-q)^2\sigma_2^2 + q^2\sigma_3^2\right] p^2(1-p)^2 \\ & + \frac{\partial^2 v}{\partial q \partial p}\left[q\sigma_3^2 - (1-q)\sigma_2^2\right] pq(1-p)(1-q). \end{aligned} \tag{3.21}$$

With degenerate parabolicity present in the problem, the value function v does not have to be smooth, and there may not be a *classical solution* to the above PDE. Thus, v has to be interpreted as a (possibly discontinuous) *viscosity solution* of this equation [8, Chapter 5].

The linear dependence on d yields the *bang-bang* property:

$$d_*(q,p,s) = \begin{cases} d_{\max}, & \text{if } \left(\dfrac{\partial v}{\partial p}p(1-p) + \dfrac{\partial v}{\partial s}\right) < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.22}$$

Specific formulations for the SR competition model

Recall from Chapter 2 Box 6 and derivations for Eq. (3.9) in §3.2.3 that the components for Eq. (3.5) are

$$\boldsymbol{X} = \begin{bmatrix} Q \\ P \end{bmatrix} := \begin{bmatrix} \dfrac{Z_{\mathrm{S}}}{Z_{\mathrm{S}} + mZ_{\mathrm{R}}} \\ Z_{\mathrm{S}} + mZ_{\mathrm{R}} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} B_t \end{bmatrix} \text{ (where } B_t \text{ is a standard 1D Brownian motion),}$$

$$\boldsymbol{a}(\boldsymbol{X}, d) = \begin{bmatrix} Q(1-Q)\Big\{(1-P)(g_{\mathrm{S}} - g_{\mathrm{R}}) - \alpha d + \beta CQP + (1-P)^2[\sigma_{\mathrm{R}}^2(1-Q) - \sigma_{\mathrm{S}}^2 Q + \sigma_{\mathrm{S}}\sigma_{\mathrm{R}}]\Big\} \\ P(1-P)(g_{\mathrm{S}}Q + g_{\mathrm{R}}(1-Q)) - \alpha QPd - \beta CP^2Q(1-Q) \end{bmatrix},$$

$$\boldsymbol{\Sigma}(\boldsymbol{X}, d) = \begin{bmatrix} (1-P)Q(1-Q)(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}}) \\ P(1-P)[\sigma_{\mathrm{S}}Q + \sigma_{\mathrm{R}}(1-Q)] \end{bmatrix}.$$

Thus, Eq. (3.20) in component-wise form is

$$\begin{aligned} 0 = \max_{d\in[0,d_{\max}]} \Bigg\{ & -\left[\frac{\partial v}{\partial q}\alpha q(1-q) + \frac{\partial v}{\partial p}\alpha qp + \frac{\partial v}{\partial s}\right] d \Bigg\} - \delta\frac{\partial v}{\partial s} \\ & + \frac{\partial v}{\partial q}\left[(1-p)(g_{\mathrm{S}} - g_{\mathrm{R}}) + \beta Cpq + (1-p)^2\{\sigma_{\mathrm{R}}^2(1-q) - \sigma_{\mathrm{S}}^2 q + \sigma_{\mathrm{S}}\sigma_{\mathrm{R}}\}\right] q(1-q) \\ & + \frac{\partial v}{\partial p}\left[(1-p)[g_{\mathrm{S}}q + g_{\mathrm{R}}(1-q)] - \beta Cpq(1-q)\right] p \\ & + \frac{1}{2}\frac{\partial^2 v}{\partial q^2}(1-p)^2q^2(1-q)^2(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}})^2 + \frac{1}{2}\frac{\partial^2 v}{\partial p^2}p^2(1-p)^2[\sigma_{\mathrm{S}}q + \sigma_{\mathrm{R}}(1-q)]^2 \\ & + \frac{\partial^2 v}{\partial q \partial p}(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}})(1-p)^2pq(1-q)[\sigma_{\mathrm{S}}q + \sigma_{\mathrm{R}}(1-q)]. \end{aligned} \tag{3.23}$$

Again, the linear dependence on d yields the *bang-bang* property:

$$d_*(q,p,s) = \begin{cases} d_{\max}, & \text{if } \left(\dfrac{\partial v}{\partial q}\alpha q(1-q) + \dfrac{\partial v}{\partial p}\alpha qp + \dfrac{\partial v}{\partial s}\right) < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.24}$$

3.4 Numerical methods and implementation details

We provide the numerical schemes and implementation details of: (i) solving for $v(\boldsymbol{\xi}, s)$; (ii) solving for $u(q, p)$; and (iii) generating CDFs in §3.4.1, §3.4.2, and §3.4.3 respectively. In both (i) and (ii), the optimal policy is found by numerically solving the corresponding HJB equation.

3.4.1 Threshold-aware optimal case

We approximate the solution to (3.20) by a first-order accurate semi-Lagrangian discretization [54] over the $(\boldsymbol{\xi}, s)$ space in standard $(n+1)$-dimensional Cartesian coordinates. Since we are really only interested in the expected value, it suffices to consider weak approximations [58, 99] of (3.5). For any small $\tau > 0$, a first order weak approximation of $\boldsymbol{X}^{\tau,\boldsymbol{d}} \approx \boldsymbol{X}(\tau; \boldsymbol{d})$ starting from $\boldsymbol{X}(0) = \boldsymbol{\xi}$ with any control value $d \in \mathcal{D}$ is

$$\boldsymbol{X}^{\tau,\boldsymbol{d}} = \boldsymbol{\xi} + \tau \boldsymbol{a}(\boldsymbol{\xi}, \boldsymbol{d}) + \boldsymbol{\Sigma}(\boldsymbol{\xi}, \boldsymbol{d}) \Delta \boldsymbol{W}^{\tau},$$

where $\Delta \boldsymbol{W}^{\tau} = (\Delta W_1^{\tau}, \ldots, \Delta W_m^{\tau})$ with ΔW_j^{τ} representing a two-point distributed variable with the distribution

$$\mathbb{P}\left(\Delta W_j^{\tau} = \pm \sqrt{\tau}\right) = \frac{1}{2}.$$

Recall $\boldsymbol{W}$ is a standard m-dimensional Brownian motion. It follows from above that $\boldsymbol{X}^{\tau,\boldsymbol{d}}$ will have 2^m possible locations to "land" with equal probability. Let $\tilde{\boldsymbol{X}}_{\ell}^{\boldsymbol{d}}$, $\ell \in \{1, 2, \ldots, 2^m\}$, denote the 2^m possible locations of $\boldsymbol{X}^{\tau,\boldsymbol{d}}$. Assuming that $\mu_{j,\ell}$ is the j-th bit in a binary representation of $(\ell - 1)$ and $\Delta \boldsymbol{W}_{\ell}^{\tau} = (\Delta W_{1,\ell}^{\tau}, \ldots, \Delta W_{m,\ell}^{\tau})$ with $\Delta W_{j,\ell}^{\tau} := (-1)^{\mu_{j,\ell}} \sqrt{\tau}$, we can now express these possible locations explicitly as

$$\boldsymbol{X}_{\ell}^{\tau,\boldsymbol{d}} = \boldsymbol{\xi} + \tau \boldsymbol{a}(\boldsymbol{\xi}, \boldsymbol{d}) + \boldsymbol{\Sigma}(\boldsymbol{\xi}, \boldsymbol{d}) \Delta \boldsymbol{W}_{\ell}^{\tau}.$$

Assume the running cost is constant over τ units of time. The dynamic programming equation obtained by weak approximations is then

$$v(\boldsymbol{\xi}, \bar{s}) = \max_{\boldsymbol{d}\in\mathcal{D}} \left\{ \frac{1}{2^m} \sum_{\ell=1}^{2^m} v\Big(\tilde{\boldsymbol{X}}_\ell^{\boldsymbol{d}}, \bar{s} - \tau K(\boldsymbol{\xi}, \boldsymbol{d}) \Big) \right\} + o(\tau). \tag{3.25}$$

To solve Eq. (3.25) numerically, we discretize $[0, \bar{\mathcal{S}}]$ into $(M+1)$ equidistant slices, with $s_k = k\Delta s$ for $k = 0, 1, 2, \ldots, M$. Each s-slice consists of a uniform rectangular grid on $[0,1]^n$, with $\boldsymbol{x}_i = i\Delta x$ for a multi-index $i = (i_1, i_2, \ldots, i_n)$ where $i_j = 0, 1, 2, \ldots, N_x$ for all $j \in \{1, 2, \ldots, n\}$.

Suppose starting from a gridpoint $\boldsymbol{x}_i$ on an s_k-slice, We have

$$v(\boldsymbol{x}_i, s_k) = \max_{\boldsymbol{d}\in\mathcal{D}} \left\{ \frac{1}{2^m} \sum_{\ell=1}^{2^m} v\Big(\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}},\ s_k - \tau K(\boldsymbol{x}_i, \boldsymbol{d}) \Big) \right\} + o(\tau),$$

where $\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}} := \boldsymbol{x}_i + \tau\ \boldsymbol{a}(\boldsymbol{x}_i, \boldsymbol{d}) + \boldsymbol{\Sigma}(\boldsymbol{x}_i, \boldsymbol{d})\Delta \boldsymbol{W}_\ell^\tau$. We choose an explicit "s-marching" scheme, in which the causality is maintained in threshold variable s since it is strictly decreasing along a path. By "causality", we mean the scheme uses the values from the previous s-slice (known) to approximate the values for the current s-slice (unknown). Namely, the above equation is exactly

$$v(\boldsymbol{x}_i, s_k) = \max_{\boldsymbol{d}\in\mathcal{D}} \left\{ \frac{1}{2^m} \sum_{\ell=1}^{2^m} v\Big(\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}}, s_{k-1} \Big) \right\} + o(\tau). \tag{3.26}$$

As a result, τ must be chosen so that $\tau K(\boldsymbol{x}_i, \boldsymbol{d}) = \Delta s$, which makes it necessary to use a control-dependent τ_d, yielding $\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}} := \boldsymbol{x}_i + \tau_{\boldsymbol{d}}\ \boldsymbol{a}(\boldsymbol{x}_i, \boldsymbol{d}) + \boldsymbol{\Sigma}(\boldsymbol{x}_i, \boldsymbol{d})\Delta \boldsymbol{W}_\ell^{\tau_{\boldsymbol{d}}}$. Let $V_i^k \approx v(\boldsymbol{x}_i, s_k)$ denote the discretized approximate solution at $(\boldsymbol{x}_i, s_k)$ and $\widetilde{V}_{i,\ell}^{k-1,\boldsymbol{d}} \approx v(\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}}, s_{k-1})$. The fully discretized equation is then

$$V_i^k = \max_{\boldsymbol{d}\in\mathcal{D}} \left\{ \frac{1}{2^m} \sum_{\ell=1}^{2^m} \widetilde{V}_{i,\ell}^{k-1,\boldsymbol{d}} \right\}. \tag{3.27}$$

Let $D_i^k \approx \boldsymbol{d}_*(\boldsymbol{x}_i, s_k)$ denote the discretized approximate optimal feedback policy at $(\boldsymbol{x}_i, s_k)$. We note that it is constructed as a by-product of solving (3.27) while determining the maximizing d at each gridpoint.

It is worth noting that, regardless of $\boldsymbol{d}$ value, $\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}}$, $\ell \in \{1, 2, \ldots, 2^m\}$, is generally not a gridpoint on the s_{k-1}-slice. Thus, each $\widetilde{V}_{i,\ell}^{k-1,\boldsymbol{d}}$ is computed by interpolating the values from the neighboring gridpoints of $\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}}$, which is the essence of semi-Lagrangian schemes. In our implementation, we have chosen the fourth-order accurate ENO cubic interpolation [152, 58] to reduce the numerical diffusion. Our full method is summarized in Algorithm 1 where $\Xi = \{i\Delta x \mid i = (i_1, i_2, \ldots, i_n)$ where $i_j = 0, 1, 2, \ldots, N_x \; \forall j \in \{1, 2, \ldots, n\}\}$.

Algorithm 1: Threshold(risk)-aware value function/policy computation

Initialize V, D at $s = 0$ using the initial & boundary conditions;
for $s_k = k\Delta s,\ k = 1, \ldots M$ **do**
  for *every* $\boldsymbol{x}_i \in \Xi$ **do**
    if $\boldsymbol{x}_i \in \Delta_{\text{succ}}$ **then**
      $V_i^k \leftarrow 1$;
    else if $\boldsymbol{x}_i \in \Delta_{\text{fail}}$ **then**
      $V_i^k \leftarrow 0$;
    else
      for *every* $\boldsymbol{d} \in \mathcal{D}$ **do**
        for $\ell = 1, 2, \ldots, 2^m$ **do**
          $\tilde{\boldsymbol{x}}_{i,\ell}^{\boldsymbol{d}} \leftarrow \boldsymbol{x}_i + \tau_{\boldsymbol{d}}\, \boldsymbol{a}(\boldsymbol{x}_i, \boldsymbol{d}) + \boldsymbol{\Sigma}(\boldsymbol{x}_i, \boldsymbol{d}) \Delta \boldsymbol{W}_\ell^{\tau_{\boldsymbol{d}}}$;
          Compute $\widetilde{V}_{i,\ell}^{k-1,\boldsymbol{d}}$ by interpolation;
        $\bar{V}_i^{k,\boldsymbol{d}} \leftarrow \dfrac{1}{2^m} \sum_{\ell=1}^{2^m} \widetilde{V}_{i,\ell}^{k-1,\boldsymbol{d}}$;
      $V_i^k \leftarrow \max_{\boldsymbol{d}} \left\{ \bar{V}_i^{k,\boldsymbol{d}} \right\}$;
      $D_i^k \leftarrow \arg\max_{\boldsymbol{d}} \left\{ \bar{V}_i^{k,\boldsymbol{d}} \right\}$;

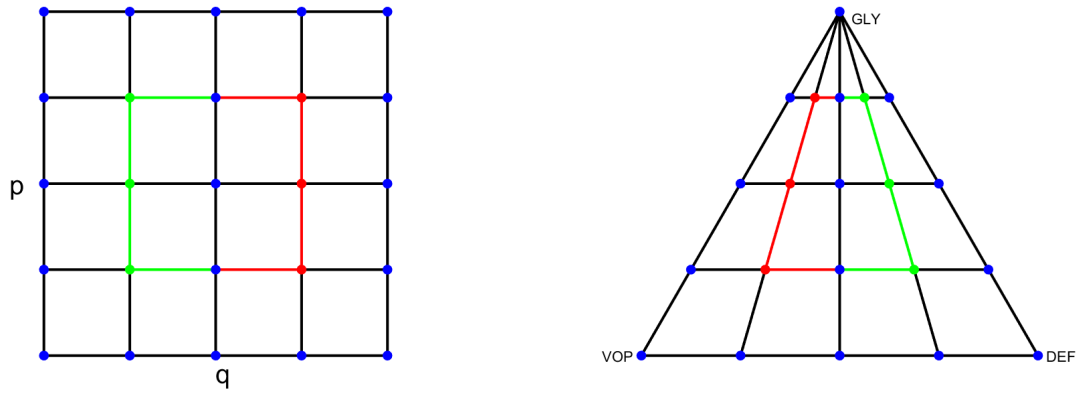
In accordance with [58], $\Delta x = o(\Delta s^{1/r})$ is needed to guarantee the convergence of this numerical scheme to the viscosity solution. Here, $\Delta x = \Delta x_j$ for all $j \in \{1, 2, \ldots, n\}$ is the space discretization step and $r = m + 1$ with m being the order of the interpolating polynomial. Since we use cubic ENO interpolants, this requires $\Delta x = o(\Delta s^{1/4})$.

To obtain accurate optimal threshold-aware policies and optimal probability of success shown in Figs 2.3&2.6 in Chapter 2, we have used $N = 1600$ on each side of the *qp* square and $M = 6000$ slices along the positive s-axis in Example 1. We use $N = 3200$ and $M = 48000$

in Example 2. With these choices of N, the success and failure barriers are just grid lines (recall that we choose $\gamma_{\mathrm{r}} = 1 - \gamma_{\mathrm{f}} = 10^{-2}$), and it is easy to apply the boundary conditions.

Additionally, since we have causality in s variable, we have further taken advantage of loop parallelism on the inner loops (iterating over spatial variables q and p) with the aid of `OpenMP` in `C++`.

For visualization purposes, for all figures related to Example 1, we transform the approximate solution V on each s-slice from the qp square into the GLY-DEF-VOP triangle. See Fig 3.1 for the specific geometric transformation.



(a) **(b)**

Figure 3.1. Geometric transformation from the qp squre to the GLY-DEF-VOP triangle in Cartesian coordinates. The subfigures show (a) grid on qp square; (b) grid on GLY-DEF-VOP triangle. See how the shape and the colors of boundary of the region enclosed by *green-red* lines change.

Remark 3.1: For both of our examples in Chapter 2, $\mathcal{D} = \{0, d_{\max}\}$ and $K(\boldsymbol{x}, d) = d + \delta$. Thus, the d-dependent τ has only two possible values

$$\begin{cases} \tau_{\{d=0\}} = \dfrac{\Delta s}{\delta}, \\ \tau_{\{d=d_{\max}\}} = \dfrac{\Delta s}{d_{\max} + \delta}, \end{cases} \tag{3.28}$$

and we can significantly improve the computational efficiency by pre-computing all $\tilde{\boldsymbol{x}}_{i,l}^{d}$.

We note that the disparity in d-dependent τ values described in (3.28) leads to much larger spatial steps with $d = 0$. This does not prevent the convergence under the grid refinement, but does contribute to larger local truncation errors on a fixed grid. In the future we hope to investigate alternative s-implicit semi-Lagrangian discretizations, which avoid such τ-disparities, but make equations coupled in each s-slice. A similar approach has already proven to be advantageous in solving first-order HJB equations with rapid transition layers [167].

Remark 3.2: Recall the random ODE satisfied by the budget variable s :

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -K(\boldsymbol{x}(t), \boldsymbol{d}(t)), \qquad s(0) = \bar{s}.$$

If we rescale the time as $\tau = t/L$ (for some fixed $L > 0$) and denote $\tilde{s}(\tau) = s(t) = s(\tau L)$, we obtain the following relation

$$\begin{aligned} \frac{\mathrm{d}\tilde{s}}{\mathrm{d}\tau} &= -LK(\tau), \\ \tilde{s}(\tau) = \bar{s} - \int_0^\tau LK(\theta)\mathrm{d}\theta &= L\left(\frac{\bar{s}}{L} - \int_0^\tau K(\theta)\mathrm{d}\theta\right). \end{aligned}$$

Consequently, the solution $v(\xi, s)$ on $s \in [0, \bar{\mathcal{S}}]$ is equivalent to $v(\xi, \tilde{s})$ on $\tilde{s} \in [0, \bar{\mathcal{S}}/L]$ with $v(\xi, s) = v(\xi, \tilde{s}L)$ for all s and $\tilde{s}$. This enables us to reduce the size of $\bar{\mathcal{S}}$ if it is too large when implementing Algorithm 1 and makes it easier to fine-tune the value of Δs. In our implementations in Chapter 2, we have chosen $L = 1$ for Example 1 and $L = 15$ for Example 2.

3.4.2 Deterministic-optimal case

In this section, we describe a numerical scheme for approximating a two-dimensional value function $u(q, p)$ in deterministic optimal control problems. This scheme has been applied to both Example 1 and Example 2 to solve (3.13) and (3.15).

We use a first-order accurate semi-Lagrangian discretization [54] over the (q, p) plane in standard two-dimensional Cartesian coordinates. We again discretize $(q, p) \in [0, 1]^2$ using a uniform $(N + 1) \times (N + 1)$ rectangular grid, on which the value function is approximated by $U_{i,j} \approx u(q_i, p_j)$.

Note that the semi-Lagrangian discretization presented here is similar to (but not quite the same as) the one used in [73]. In particular, Gluzman et al. used a *linear* interpolation on a *triangular* mesh with *adaptive* time step τ, while our current version uses a *bi-linear* interpolation on a *rectangular* grid with a *fixed* time step τ. While different, these two choices have the same formal order of accuracy and converge to the same result as the discretization parameters approach zero.

To be more compact, we further denote the deterministic dynamics (for both Example 1 and Example 2) as

$$\begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} f_q(q, p, d) \\ f_p(q, p, d) \end{bmatrix} \tag{3.29}$$

Assuming the rate of change is constant for a small amount of time $\tau > 0$, the foot of the characteristics starting from a gridpoint (q_i, p_j) lands at a new state

$$\tilde{q}_i^d = q_i + \tau f_q(q_i, p_j, d),$$
$$\tilde{p}_j^d = p_j + \tau f_p(q_i, p_j, d).$$

From Bellman's Optimality Principle [16], we have

$$u(q_i, p_j) = \min_{dd \in \{0,\, d_{\max}\}} \left\{ \tau(d + \delta) + u\left(\tilde{q}_i^d, \tilde{p}_j^d\right) \right\} + o(\tau), \tag{3.30}$$

yielding the discretized version

$$U_{i,j} = \min_{d \in \{0,\, d_{\max}\}} \left\{ \tau(d + \delta) + \widetilde{U}_{i,j}^d \right\}, \tag{3.31}$$

where $\widetilde{U}^d_{i,j} \approx u(\tilde{q}^d_i, \tilde{p}^d_j)$ is computed through a *bi-linear* interpolation of the U values from the four neighboring gridpoints surrounding $(\tilde{q}^d_i, \tilde{p}^d_j)$. The optimal feedback policy is recovered as an argmin in (3.31).

Since the discretization (3.31) is not explicitly causal, an iterative method is needed to solve the resulting system. This is often done through standard Value Iterations (VI) [19, 18, 20], but in our case this approach results in a very slow convergence on a fixed grid. To address this, we have implemented a version of the "*hybrid Value-Policy*" Iteration (VPI) algorithm [78].

That is, we start with value iterations where we solve the nonlinear Eq. (3.31) by Gauss-Seidel iterations with "Fast Sweeping" orderings [24, 180, 137]. We initialize `err` to be the L_∞-norm of U-change in the very first value iteration, and use ϵ to denote the L_∞-norm of U-change in the *last* value iteration performed so far. Whenever ϵ falls below $\rho * \texttt{err}$, where $\rho \in (0, 1)$ is a preset hyperparameter, we store this ϵ as the new `err` and proceed to the "*policy-evaluation*" (PE) step.

In the PE step, we compute the value function by solving a system of linear equations with a fixed policy $\hat{d}$ (recovered from the most recent value iteration). We thus solve a linear system of equations

$$U_{i,j} = \tau\Big(\widehat{D}_{i,j} + \delta\Big) + \widetilde{U}^{\widehat{D}_{i,j}}, \tag{3.32}$$

where $\widehat{D}_{i,j} = \hat{d}(q_i, p_j)$ and $\widetilde{U}^{\widehat{D}_{i,j}} \approx u\left(\tilde{q}_i, \tilde{p}_j^{\widehat{D}_{i,j}}\right)$ is again computed through a bi-linear interpolation.

After obtaining the solution to Eq. (3.32), we return to the value iteration part and repeat the process until $\epsilon < \texttt{tol}$, where `tol` is a preset tolerance of convergence. In both Example 1 and Example 2, to obtain an accurate deterministic-optimal policy, we have used $N = 3200$ on each side of the unit qp-square, $\texttt{tol} = 10^{-6}$, and $\rho = 0.7$.

3.4.3 Generating CDFs

To generate the CDFs we provide in both Chapter 2 and §3.5.4, we have used a Monte Carlo (MC) method with 10^5 samples. That is, for each sample, we start with a fixed initial tumor configuration (q_0, p_0), and then apply a strong approximation [99] of $(Q(t), P(t))$ to simulate a sample path till either stabilization/remission or death. We store the random cost $\mathcal{J}$ incurred along each sample path as a data point, and then generate the (empirical) CDF based on 10^5 data points. In particular, we compute the Kaplan-Meier estimate [93] (empirical) CDF by `Matlab`'s built-in function `ecdf()`.

Let Δt denote the uniform time step. The cost incurred for each time step along the sample path is

$$\Delta J = \begin{cases} \Big(d_{\max} + \delta\Big)\Delta t, & \text{if } d_* = d_{\max} \text{ at this step,} \\ \delta\ \Delta t, & \text{otherwise.} \end{cases}$$

Thus, the cost accumulated at the n-th time step can be defined recursively as

$$J^n = J^{n-1} + \Delta J, \qquad J_0 = 0.$$

Suppose we start with an initial budget value $\bar{s}$ at $t = 0$. Then the remaining budget at the n-th time step is

$$S^n = S^{n-1} - \Delta J, \qquad S_0 = \bar{s}.$$

As mentioned in §3.4.1, we discretize $[0, \bar{s}]$ into equidistant slices, and $(q, p) \in [0, 1]^2$ into a uniform rectangular grid. Among many choices of strong approximations of SDEs, we choose the Euler-Maruyama scheme [99].

Let us denote the approximate solution to (3.8) (or (3.9)) at the n-th time step ($t_n = n\Delta t$) as $(Q^n, P^n) \approx (Q(t_n), P(t_n))$, and the optimal feedback policy at (Q^n, P^n, S^n) as D^n. Then

the Euler-Maruyama scheme defines (Q^n, P^n) recursively by

$$\begin{bmatrix} Q^n \\ P^n \end{bmatrix} = \begin{bmatrix} Q^{n-1} \\ P^{n-1} \end{bmatrix} + \boldsymbol{a}(Q^{n-1}, P^{n-1}, D^{n-1})\Delta t + \boldsymbol{\Sigma}(Q^{n-1}, P^{n-1})\Delta W^n. \tag{3.33}$$

where $(Q_0, P_0) = (q_0, p_0)$. Unlike in the weak approximation mentioned in §3.4.1, this time the ΔW_j^n's are *independent and identically distributed* (i.i.d.) normal random variables with mean zero and variance Δt for all $j \in \{1, 2, \ldots, m\}$.

Due to memory limitations, we only store the policy for a subsample of s-slices represented in the PDE discretization grid. The policy is stored for $s = 0$, $\Delta\hat{s}$, $2\Delta\hat{s}$, $3\Delta\hat{s}$, ... (We use $\Delta\hat{s} = 0.005$ for Example 1 and $\Delta\hat{s} = 0.00125$ for Example 2). To obtain an accurate sample path approximation, we need a sufficiently small Δt, which often means that (Q^n, P^n, S^n) is not on any stored s-slice. As a consequence, we have to determine whether to use drugs or not based on the *data cube* surrounding (Q^n, P^n, S^n) (the cube is formed by the 4 data points on the s-slice above (Q^n, P^n, S^n) and another 4 data points on the s-slice below it).

For all the figures related to Example 1 we provide here and in Chapter 2, we choose a "conservative" drugs-on/off determination strategy for Example 1. That is, we decide to use drugs at (Q^n, P^n, S^n) if all of the 8 data points of the cube have the value $d_* = d_{\max}$. On the other hand, we used a "Majority" strategy for Example 2.[1] That is, we decide to use drugs at (Q^n, P^n, S^n) if 5 of the 8 data points of the cube have the value $d_* = d_{\max}$. When the budget runs out, i.e., $S_n = 0$, we switch to the deterministic-optimal policy as mentioned in Chapter 2 §2.3.1. Notice that the cost will not stop accumulating until we cross either γ_{r} or γ_{f}. In the deterministic-optimal case, we only need to consider the *data square* surrounding (Q^n, P^n) since it is now s-independent.

[1] Alternatively, one can also try define an "aggressive" strategy in either example; i.e., to use drugs as long as one of the 8 data points has $d_* = d_{\max}$. We have considered this and other determination strategies and they all yield results within the 95% confidence interval of each other.

In our implementations, we have used the following adaptive choice of Δt:

$$\text{Example 1: } \Delta t = \begin{cases} 2.5 \times 10^{-3}, & \text{if } d_* = 0, \\ 1.64 \times 10^{-4}, & \text{if } d_* = d_{\max}. \end{cases}$$

$$\text{Example 2: } \Delta t = \begin{cases} 2.5 \times 10^{-3}, & \text{if } d_* = 0, \\ 4.10 \times 10^{-4}, & \text{if } d_* = d_{\max}. \end{cases}$$

A parallel `for` loop was used in `Matlab` to reduce the computational time of Monte Carlo simulations.

Here, we provide a table of details of all CDF figures in this chapter and Chapter 2. The "95% Confidence bounds of `ecdf`()" is computed via Greenwood's formula [77].

For Example 1, the discrepancy between the approximated value function and the MC simulation at a given $(q_0, p_0, \bar{s})$ is around 10^{-3} except for two examples where the discrepancy is around 0.01. For Example 2, the discrepancy between the approximated value function and the MC simulation at a given $(q_0, p_0, \bar{s})$ is around 10^{-4} except for one example where the discrepancy is around 10^{-3}. These discrepancies are due to (a) the discretization errors in solving the Hamilton-Jacobi PDE, (b) subsampling of the optimal policy, and (c) the variability of outcomes and a time discretization in MC simulations. All of these can be reduced with a higher computational cost (e.g., using finer discretizations and more MC simulations).

Example 1	Initial configuration $(q_0, p_0, \bar{s})$	Volatilities $(\sigma_1 = \sigma_2 = \sigma_3)$	Value function $v(q_0, p_0, \bar{s})$	Empirical CDF at $\bar{s}$	95% Confidence bounds of `ecdf()` at $\bar{s}$
Fig 2.1(c)	$(0.26, 0.665, 5.0)$	0.15	0.6763	0.6737	$[0.6708, 0.6766]$
Fig 2.1(c)	$(0.26, 0.665, 4.5)$	0.15	0.4117	0.4041	$[0.4011, 0.4072]$
Fig 2.5(c)	$(0.27, 0.4, 4.71)$	0.15	0.6421	0.6371	$[0.6341, 0.6401]$
Fig 2.5(c)	$(0.27, 0.4, 4.35)$	0.15	0.4684	0.4562	$[0.4532, 0.4593]$
Fig 3.2(c)	$(0.8, 0.4, 3.77)$	0.15	0.6481	0.6503	$[0.6474, 0.6533]$
Fig 3.2(c)	$(0.8, 0.4, 3.45)$	0.15	0.3238	0.3145	$[0.3116, 0.3173]$
Fig 3.7(e)	$(0.27, 0.4, 4.46)$	0.5	0.8906	0.8936	$[0.8917, 0.8956]$
Fig 3.7(e)	$(0.27, 0.4, 3.65)$	0.5	0.7742	0.7761	$[0.7735, 0.7787]$

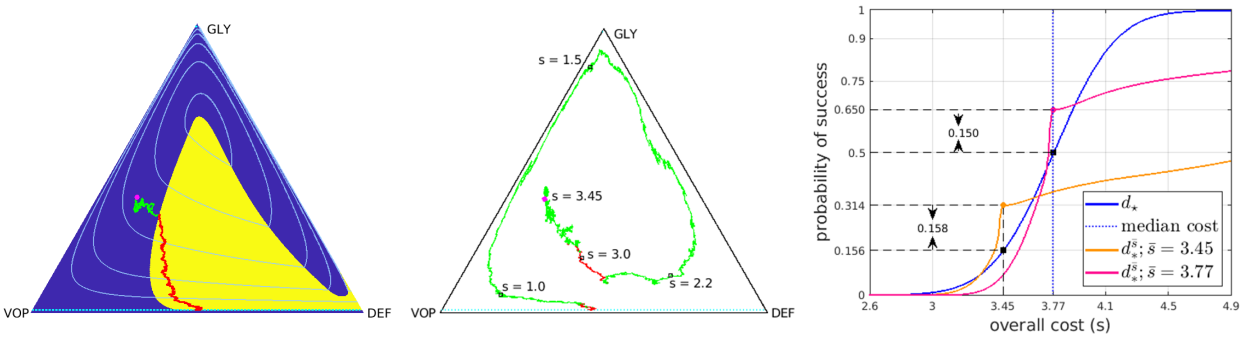
Example 2	Initial configuration $(q_0, p_0, \bar{s})$	Volatilities $(\sigma_S = \sigma_R)$	Value function $v(q_0, p_0, \bar{s})$	Empirical CDF at $\bar{s}$	95% Confidence bounds of `ecdf()` at $\bar{s}$
Fig 2.7(c)	$(0.45, 0.9, 69.45)$	0.15	0.6746	0.6739	$[0.6710, 0.6768]$
Fig 2.7(c)	$(0.45, 0.9, 60)$	0.15	0.4013	0.3984	$[0.3954, 0.4015]$
Fig 3.3(c)	$(0.55, 0.9, 72.75)$	0.15	0.6697	0.6693	$[0.6664, 0.6722]$
Fig 3.3(c)	$(0.55, 0.9, 60)$	0.15	0.3040	0.3039	$[0.3010, 0.3067]$

3.5 More numerical results

3.5.1 More policies, trajectories, and CDFs for the EGT Example

Using the same model and parameter values as in Chapter 2 §2.3.1, we now consider another initial tumor configuration at $(q_0, p_0) = (0.8, 0.4)$ located in the blue (drugs-off) region of $d_\star$.

Our threshold-aware policy (pink) still improves the $\mathbb{P}(\mathcal{J} \leq \bar{s}_{\text{med}})$ from 50% to 65%, where $\bar{s}_{\text{med}} = 3.77$ is the median cost of $\mathcal{J}$ associated with $d_\star$. When starting from a lower initial budget $\bar{s} = 3.45$, the deterministic-optimal policy provides only a 15.6% chance of stabilization, while our threshold-aware policy (orange) doubles this $\mathbb{P}(\mathcal{J} \leq 3.45)$ to 31.5%.



(a) **(b)** **(c)**

Figure 3.2. Comparison between threshold-aware policies and the deterministic-optimal policy (EGT model; Case II). Starting from an initial state $(q_0, p_0) = (0.8, 0.4)$ (magenta dot): (a) a sample path with cost 3.94 under the deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 3.45$ with a total cost of 3.41 under the (orange) threshold-aware policy; (c) CDFs of the cumulative cost $\mathcal{J}$ approximated using 10^5 random simulations. In (c), the *solid blue* curve is the CDF generated with the deterministic-optimal policy. Its median (*dashed blue* line) is 3.77 and its mean conditioning on success is also 3.77. The *solid orange* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 3.45$; and the *solid pink* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 3.77$.

We can see from Fig 3.2(a) that the deterministic-optimal policy basically prescribes drugs till stabilization once the (random) tumor state enters the yellow region. In contrast, Fig 3.2(b) shows that under $d_*^{\bar{s}}$ there is more than one period of MTD drug use for many sample paths.

3.5.2 More policies, trajectories, and CDFs for the SR model

Except for $d_{\max}$, most of the parameter listed below correspond to those from Table 1 in [30]. We have used these for all numerical experiments in this chapter and Chapter 2.

Symbol	Meaning	Value	Unit
C	Carrying capacity of the Petri dish	4.8×10^6	cells
m	Size ratio between S and R cells	30	adimensional
g_{S}	intrinsic growth rate of the sensitive	0.031	hour^{-1}
g_{R}	intrinsic growth rate of the resistant	0.026	hour^{-1}
d	drug concentration	Maximum: 3	nM
α	drug efficiency	0.06	$(\text{nM} \cdot \text{hour})^{-1}$
β	action of sensitive on resistant	6.25×10^{-7}	$(\text{cells} \cdot \text{hour})^{-1}$

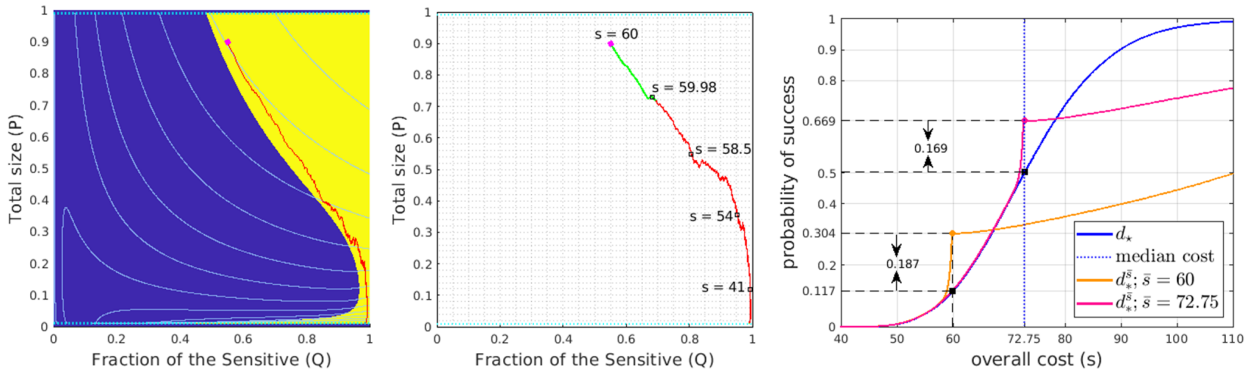


Figure 3.3. Comparison between threshold-aware policies and the deterministic-optimal policy (SR model; Case II). Starting from an initial state $(q_0, p_0) = (0.55, 0.9)$ (magenta dot): (a) a sample path with cost 56.4 under the deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 60$ with a total cost of 54.26 under the (orange) threshold-aware policy; (c) CDFs of the cumulative cost $\mathcal{J}$ with 10^5 samples. In (c), the *solid blue* curve is the CDF generated with the deterministic-optimal policy. Its median (*dashed blue* line) is 72.75 while its mean conditioning on success is 73.8. The *solid orange* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 60$; and the *solid pink* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 72.75$.

Similar to Fig 2.5, we here consider an initial tumor configuration $(q_0, p_0) = (0.55, 0.9)$ inside the yellow (drugs-on) region of $d_\star$.

In Fig 3.3, we see our threshold-aware policy (pink) improves the $\mathbb{P}(\mathcal{J} \leq 72.75)$ to 67%

from 50%. In addition, when the budget is tight, the $\mathbb{P}(\mathcal{J} \leq 60)$ under the deterministic-optimal policy is only around 12%, while our threshold-aware policy (orange) almost triples this probability to 30%. This is mainly because our threshold-aware policies leverage the inherent competitiveness of the sensitive for a brief period, prescribing drugs once the tumor size slightly diminishes. In contrast, the deterministic-optimal policy will prescribe $d_{\max}$ all the way till remission, yielding a significantly higher cumulative cost.

3.5.3 Time-evolution plots for Fig 2.7 in Chapter 2

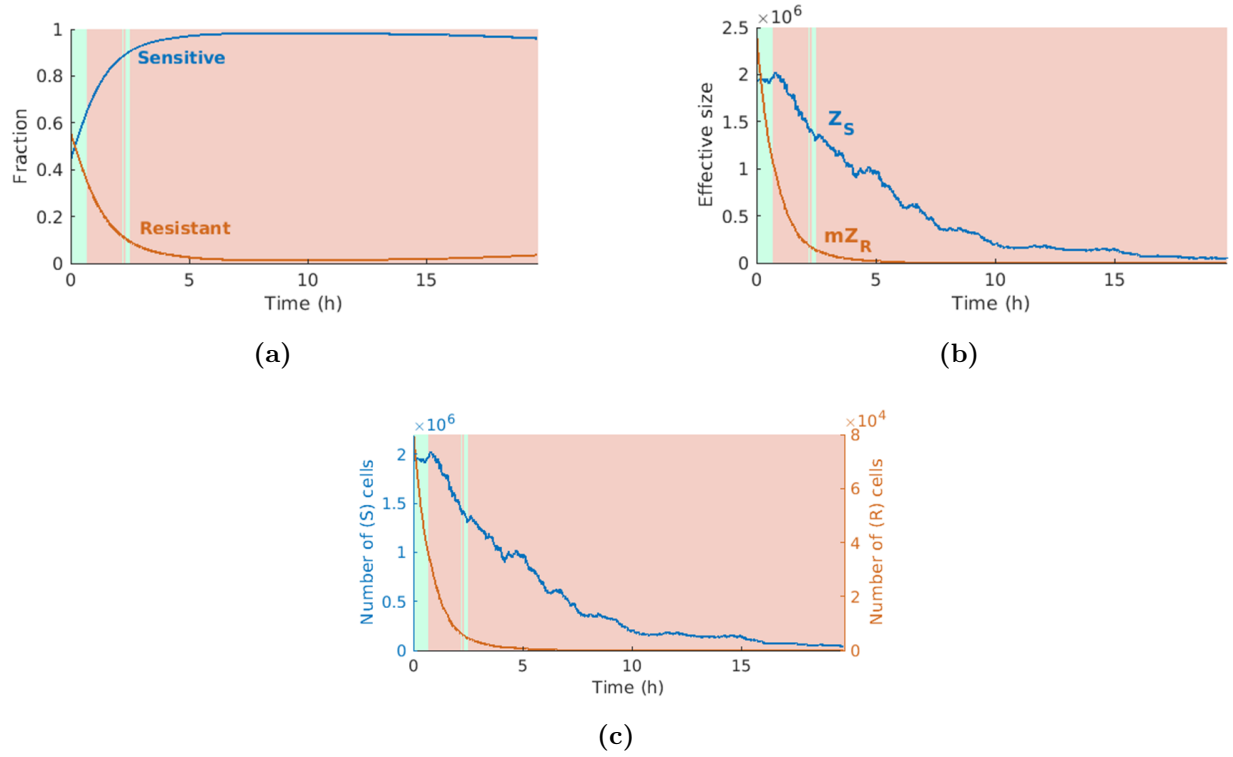


Figure 3.4. Time-evolution plots correspond to the sample path in Fig 2.7(a). Starting from an initial state $(q_0, p_0) = (0.45, 0.9)$, the subfigures show (a) time traces of *fractions* of effective tumor size $q(t)$ and $1 - q(t)$, taken by the sensitive and the resistant, respectively; (b) time traces of the actual effective tumor sizes Z_{S} and mZ_{R}; (c) time traces of their respective number of cells. Here we use light green and light pink backgrounds to indicate the time interval(s) of prescribing no drugs and of prescribing drugs at the MTD-rate, respectively. In (c), the left vertical axis, representing the number of sensitive (S) cells, and the right vertical axis, denoting the number of resistant (R) cells, are color-matched to their respective line plots.

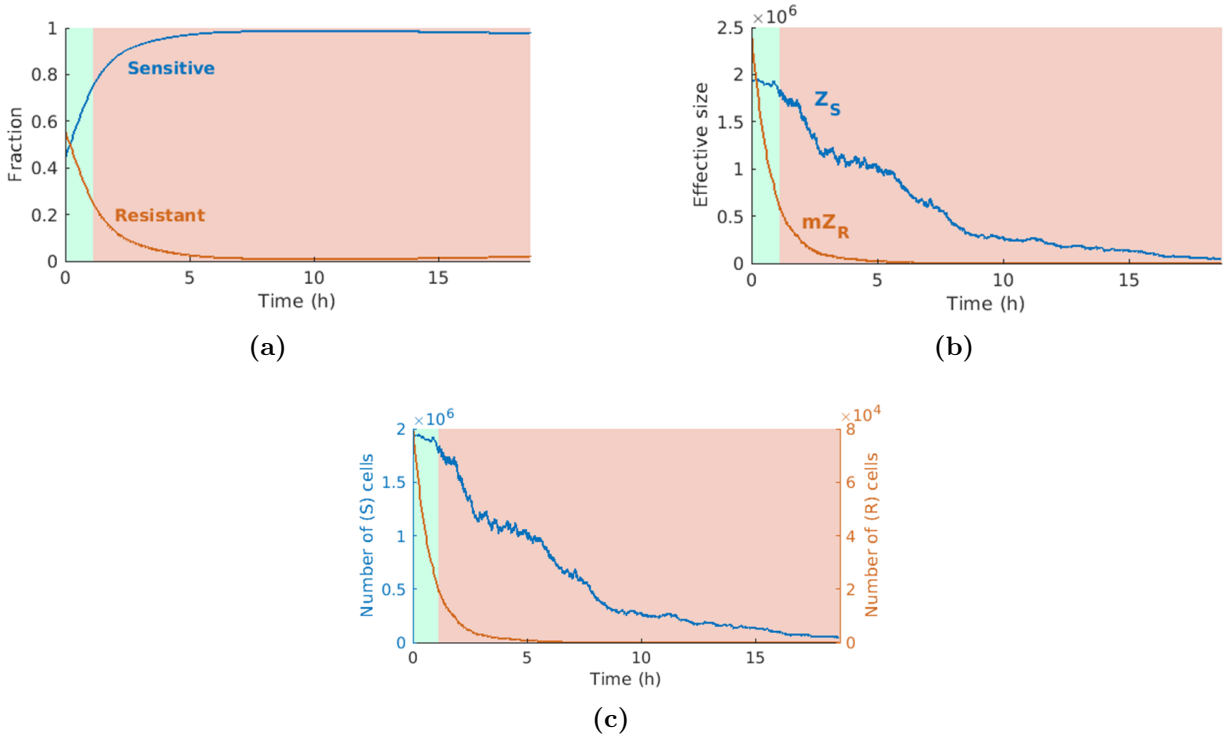


Figure 3.5. Time-evolution plots correspond to the sample path in Fig 2.7(b). Starting from an initial state $(q_0, p_0) = (0.45, 0.9)$, the subfigures show (a) time traces of *fractions* of effective tumor size $q(t)$ and $1 - q(t)$, taken by the sensitive and the resistant, respectively; (b) time traces of the actual effective tumor sizes Z_{S} and mZ_{R}; (c) time traces of their respective number of cells. Here we use light green and light pink backgrounds to indicate the time interval(s) of prescribing no drugs and of prescribing drugs at the MTD-rate, respectively. In (c), the left vertical axis, representing the number of sensitive (S) cells, and the right vertical axis, denoting the number of resistant (R) cells, are color-matched to their respective line plots.

The time-evolution plots for Fig 3.3 are similar to those shown above, and are therefore omitted here for brevity.

3.5.4 The EGT Example with higher volatilities

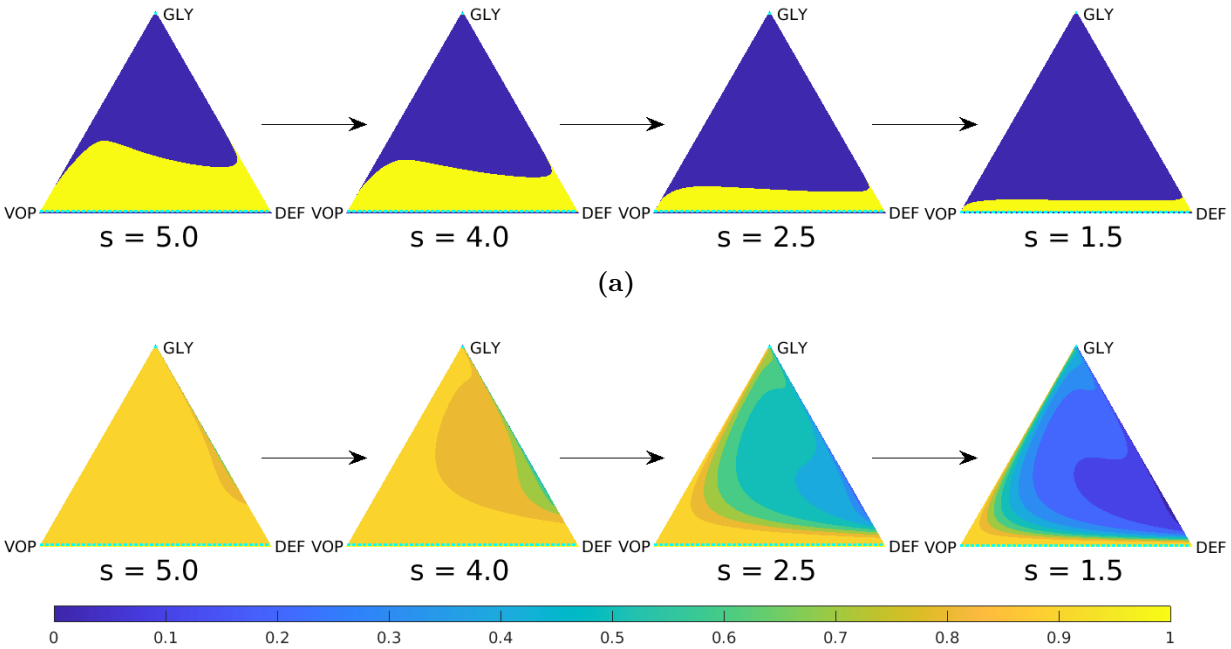


Figure 3.6. Representative slices of the threshold-aware optimal policy (top row) and the corresponding probability of success (bottom row) for the EGT-based model with $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$. Each triangle represents all possible tumor compositions (proportions of GLY/VOP/DEF cells in the population). Top row shows the policy, which prescribes the optimal instantaneous decisions on drug usage given the indicated remaining budget (s) and the current tumor state. Bottom row shows the probability of "stabilization within the budget" if the optimal policy is followed from this time point and onward. Each column corresponds to a specific budget level s, which is shown below each triangle. The arrows indicate the natural decrease of the remaining budget while implementing the policy.

The results in this section are obtained by solving (3.21) with $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$. (Other parameter values $d_{\max} = 3, b_a = 2.5, b_v = 2, c = 1, n = 4$ are the same ones as in §2.3.1.)

Fig 3.6 presents some representative s-slices of the optimal policy and their corresponding optimal probability of success. One observes that both Fig 3.6(a) and Fig 3.6(b) are significantly different from those computed with $\sigma_1 = \sigma_2 = \sigma_3 = 0.15$; see Fig 2.3 in Chapter 2 §2.3.1.

Since in this case, the random perturbations significantly affect the cancer evolution dynamics, the deterministic portion of Eq. (3.8) (the drift terms) is no longer the dominant force that leads to stabilization. As a consequence, one can see from Fig 3.6(b) that the closer to the right half of the GLY-DEF-VOP triangle, the higher the optimal probability of stabilization is. Furthermore, unlike with $\sigma_1 = \sigma_2 = \sigma_3 = 0.15$, the drugs-on (yellow) region in Fig 3.6(a) has just one connected component away from the GLY vertex of the triangle and the drug use is prescribed along the entire stabilization barrier. Consequently, vast majority of samples from simulations shows that once MTD-based therapy is turned on, it stays on until stabilization.

We again compare results from both the deterministic-optimal policy and threshold-aware optimal policies subject to this new stochastic dynamics. We use $(q_0, p_0) = (0.27, 0.4)$ (labeled as a magenta dot in Fig 3.7) as the initial tumor state, which is inside the drugs-on (yellow) region of the deterministic-optimal policy. Although the resulting sample paths would stay inside the yellow region with high probability (shown in Fig 3.7(a)), once the random perturbations bring it outside the yellow region, it has 5.3% chance of crossing the failure barrier as shown in Fig 3.7(b). The corresponding CDF (*solid blue* in Fig 3.7(e)) shows that the probability of stabilization under $\bar{s} = 3.65$ is only 0.18.

However, if our threshold-aware policy is used instead, the CDF (*solid orange* in Fig 3.7(e)) shows the maximized probability of stabilization under $\bar{s} = 3.65$ is 0.78, which significantly improves the chance of stabilization by 0.596. The median cost of $\mathcal{J}$ associated with the deterministic-optimal policy is 4.46 (*dashed blue* line in Fig 3.7(e)). Our threshold-aware policy (*solid pink* in Fig 3.7(e)) would increase the probability of stabilization under this budget from 0.5 to 0.89. Interestingly, it does not result in a markedly lower probability of success for s values above $\bar{s}$: the pink and blue CDF curves are quite close for large s values. Two representative sample paths under the threshold-aware policy with $\bar{s} = 4.46$ are

given in Fig 3.7(c)&(d). As we see, with higher σ_i values the stabilization might be attained without an overall counterclockwise direction of the random trajectory.

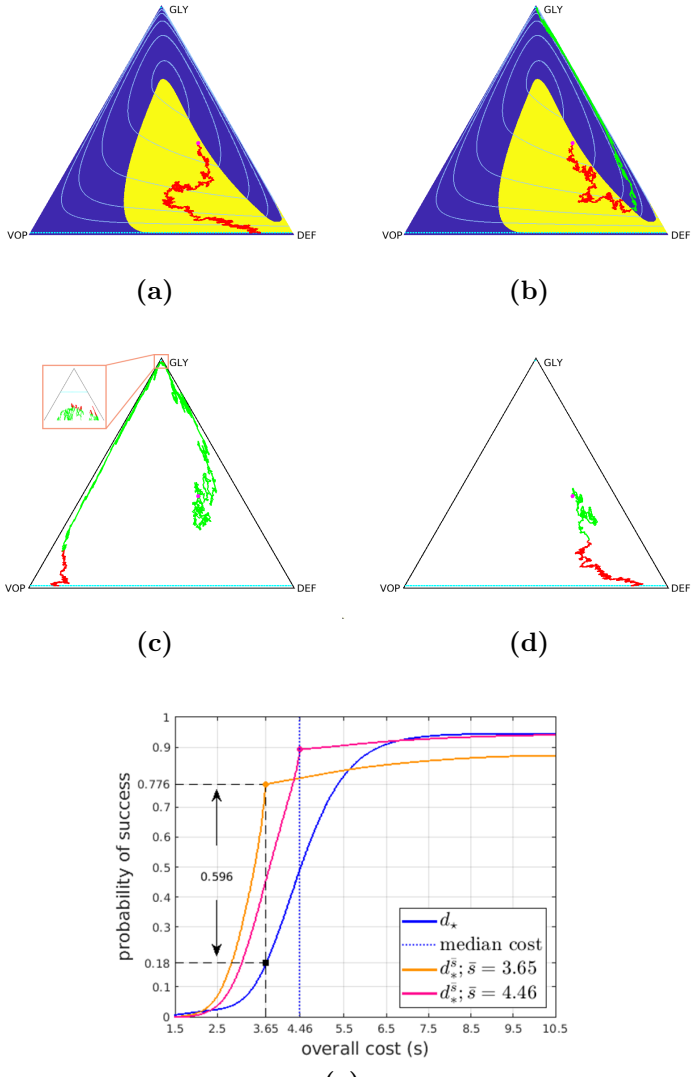


Figure 3.7. Comparison between threshold-aware policies and the deterministic-optimal policy (EGT-based model with $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$). Starting from an initial state $(q_0, p_0) = (0.27, 0.4)$ (magenta dot): (a) a sample path with cost 5.33 under the deterministic-optimal policy; (b) a sample path that leads to failure under the deterministic-optimal policy; (c) & (d) two sample paths starting at $\bar{s} = 4.46$ under the (pink) threshold-aware policy with respective total costs 3.47 and 3.21; (e) CDFs of the cumulative cost $\mathcal{J}$ approximated using 10^5 random simulations. In (e), the *solid blue* curve is the CDF generated with the deterministic-optimal policy. Its median (*dashed blue* line) is 4.46 while its mean conditioning on success is 4.48. The *solid orange* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 3.65$; and the *solid pink* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.46$.

CHAPTER 4

RISK-AWARE STOCHASTIC CONTROL OF A SAILBOAT

4.1 Introduction

Sailboat racing is one of the many areas where game-theoretic and control-theoretic tools are valuable in improving the competitive performance. The uncertainty in weather patterns gives rise to hybrid stochastic control models with many reasonable choices of performance measures to optimize. The previously developed methods have focused on risk-neutral optimization (e.g., minimizing the expected time to destination) [59, 29, 117]. In contrast, here we focus on maximizing the probability of desirable outcomes (e.g., arriving prior to a specified deadline). Our *risk-aware* approach addresses a notion of robustness very different from the traditional H^∞ control [13] and has important advantages for many applications. Indeed, it has been already successfully used in piecewise-deterministic Markov processes [32] and in bang-bang stochastic control models of adaptive drug therapies [169]. Unlike the typical *risk-averse* approaches [170], the method that we develop here for the hybrid control setting allows finding the optimal control policies for a large set of starting sailboat positions and *a range of deadlines* simultaneously. This is accomplished in the general framework of dynamic programming and requires solving a pair of quasi-variational inequalities on a 3D computational domain.

The hybrid nature of this problem is due to "tacking": to travel upwind, sailors must use a zigzag pattern, periodically swinging the bow to the other side of the wind. Each such "tack-switch" incurs a significant slow down while the wind pushes against the boat. We adopt a commonly used simplified model which assumes that the boat's velocity vector can be changed instantaneously (choosing among all directions available in its current tack) but

a switch to the opposite tack incurs a fixed time-penalty.

Optimization of sailboat routing is a topic of increasing mathematical interest. In [132] the task of minimizing the expected time to target was considered in a discrete setting, with a discrete-time Markov chain modeling occasional changes in weather conditions. The idea was extended to continuous-time Markov chains in [166], with a tack-switching curve defined in the state space to encode the optimal policy. In [59], this switching curve was found using dynamic programming for indefinite-horizon hybrid control problems, but under the assumption that the wind direction stays constant for the duration of each tack switch. In [117], it was shown how this assumption can be avoided, yielding an improvement in control policies.

We start by introducing the hybrid dynamics in Section 4.2, and describe both the risk-neutral and risk-aware optimal control problems in Section 4.3. Our numerical approach to the latter is presented in Section 4.4, followed by the summary of computational experiments in Section 4.5. We conclude by listing directions for future work in Section 4.6.

4.2 System Dynamics

Following [59, 117], we assume the strength of the wind is fixed but its direction (measured counterclockwise from the y-axis) undergoes a Brownian drift/diffusion process:

$$\mathrm{d}\phi = a\mathrm{d}t + \sigma\mathrm{d}B, \tag{4.1}$$

where $\phi(t)$ denotes the current upwind direction, a is a constant drift, σ is the diffusion coefficient, and B is a standard Brownian motion. The state of the system can be represented as (x, y, q, ϕ), where x and y encode the boat's current position, while $q \in \{1, 2\}$ is its current "tack", which determines the range of available steering directions. Our continuous control

is the steering angle, $u \in [0, \pi]$, measured relative to the wind. In the starboard tack ($q = 1$), u is measured counterclockwise from the upwind direction, while in the port tack ($q = 2$) it is measured clockwise; so, the boat's direction of motion relative to upwind is $(-1)^q u$. The boat's angle-dependent speed $f(u)$ is encoded in the speed profile (often called the "polar"), which is determined by the geometry of each specific boat. Fig 4.1(a) shows a typical polar used in all numerical tests here and in [117]. With these assumptions, the boat's dynamics is described by

$$dx = -f(u)\sin(\phi - (-1)^q u)dt, \tag{4.2}$$

$$dy = f(u)\cos(\phi - (-1)^q u)dt. \tag{4.3}$$

For the sake of computational efficiency, we adopt a dimensional reduction of (4.1)-(4.3) suitable when aiming for a circular target $\mathcal{D}$ in a domain with no obstacles [117]. Assuming a target at the origin, $r = \sqrt{x^2 + y^2}$ represents the boat's distance to the center of $\mathcal{D}$, while $\theta = \phi + \text{atan2}(-x, -y)$ encodes the upwind direction measured counterclockwise from the line connecting the boat to the center of $\mathcal{D}$; see Fig 4.1(b). This results in the system dynamics:

$$\begin{cases} \mathrm{d}r = r_{\mathrm{d}}(\theta, q, u)\ \mathrm{d}t, \\ \mathrm{d}\theta = \theta_{\mathrm{d}}(r, \theta, q, u; a)\ \mathrm{d}t + \sigma \mathrm{d}B, \end{cases} \tag{4.4}$$

where $r_{\mathrm{d}}(\theta, q, u) = -f(u)\cos\left(\theta - (-1)^q u\right)$ and $\theta_{\mathrm{d}}(r, \theta, q, u; a) = \dfrac{f(u)}{r}\sin\left(\theta - (-1)^q u\right) + a$ define the *deterministic portion* of the system dynamics.

4.3 Stochastic Optimal Control

Let $\boldsymbol{\xi}(t) = (r(t), \theta(t))$ denote the continuous component of the system state at the time t. We define $\Omega := [0, R_{\max}] \times [0, 2\pi) \times \{1, 2\}$ as the full state space (with the last component

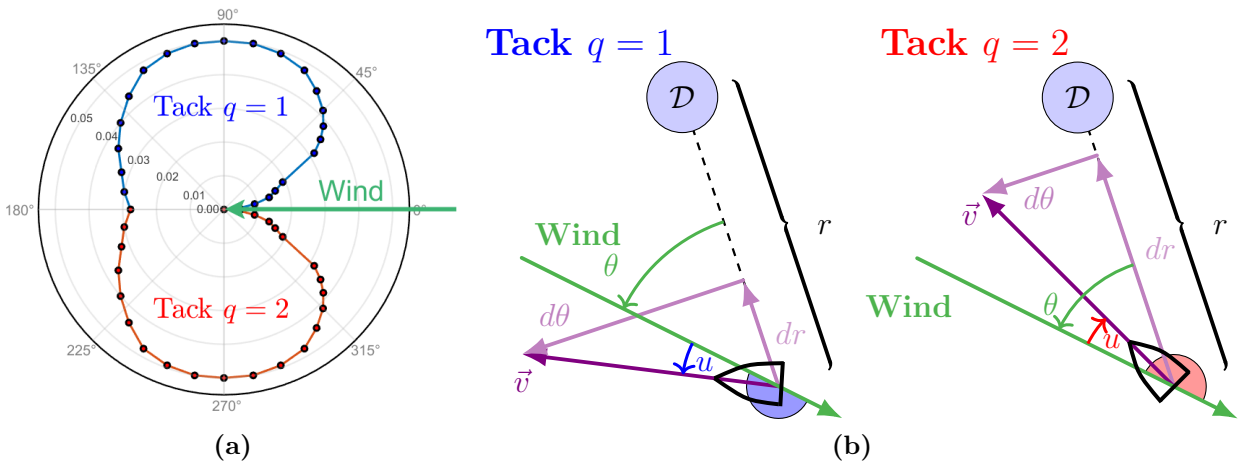


Figure 4.1. System dynamics relative to the wind and relative to the target $\mathcal{D}$. The subfigures show (a) the polar speed plot $f(u)$ used in this work (the same as Fig 1(a) in [117]) and (b) system setups in the polar coordinate centered at $\mathcal{D}$ for different tacks. In (a), $f = 0$ at $u = 0°$.

encoding the tack q). We will further use $\tilde{q} = 3 - q$ to refer to "the opposite tack" and $\Upsilon := [0, \pi] \cup \{\blacktriangle\}$ to define the policy domain (the prescribed steering angle or the special tack-switch action $\blacktriangle$).

Given a starting configuration of the boat ($\boldsymbol{\xi}(0) = \hat{\boldsymbol{\xi}}, q(0) = \hat{q}$) and a feedback control policy $\mu : \Omega \to \Upsilon$, the main quantity of interest is the random time to target $T_\mu(\hat{\boldsymbol{\xi}}, \hat{q}) = \inf\{t > 0 \mid \boldsymbol{\xi}(t) \in \mathcal{D},\ \boldsymbol{\xi}(0) = \hat{\boldsymbol{\xi}},\ q(0) = \hat{q}\}$ based on that μ. Due to the hybrid nature mentioned in section 4.2, $T_\mu(\hat{\boldsymbol{\xi}}, \hat{q})$ is the sum of the time spent in both steering and tack-switching. A typical *risk-neutral* approach is to define the value function as $v(\hat{\boldsymbol{\xi}}, \hat{q}) = \inf_\mu \mathbb{E}[T_\mu(\hat{\boldsymbol{\xi}}, \hat{q})]$, which can be then recovered by solving a system of quasi-variational HJB-type inequalities. Assuming each tack-switch takes a fixed time C and the boat stays in place ($f(u) = 0$) while switching, as shown in [117] this leads to

$$\max\left\{H(r, \theta, q, \nabla v) - \frac{\sigma^2}{2}\frac{\partial^2 v}{\partial \theta^2},\ v - \mathcal{N}v - C\right\} = 0, \tag{4.5}$$

where the switching operator $\mathcal{N}$ is defined below, $\nabla v = (\frac{\partial v}{\partial r}, \frac{\partial v}{\partial \theta})$, and the Hamiltonian is

$$H(r, \theta, q, \mathbf{p}) = \max_u \left(-r_{\mathrm{d}}(\theta, q, u)p_1 - \theta_{\mathrm{d}}(r, \theta, q, u; a)p_2\right) - 1. \tag{4.6}$$

In (4.5), we take the maximum over two alternative courses of action. The first clause corresponds to the system states, from which it is optimal to continue in the current tack and the maximization in (4.6) selects the optimal steering angle. On the other hand, in the second clause $\mathcal{N}v$ encodes the expected remaining time-to-target if we switch to the opposite tack; i.e., $\mathcal{N}v(\hat{r},\hat{\theta},q) = \mathbb{E}[v(\hat{r},\theta(C),\tilde{q}) \mid \theta(0)=\hat{\theta}, f=0]$. If we define $\psi_{r,q}(z) = v(r,z,\tilde{q})$ and

$$\mathcal{G}_\theta[\psi] = \frac{1}{\sigma\sqrt{2\pi C}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(z-\theta-aC)^2}{2C\sigma^2}} \psi(z)\,\mathrm{d}z, \tag{4.7}$$

then this switching operator can be conveniently evaluated as $\mathcal{N}v(r,\theta,q) = \mathcal{G}_\theta\left[\psi_{r,q}\right]$. Since the part of Ω on which it is better to switch tacks is a priori unknown, this is a problem with a *free boundary*. The optimal feedback policy μ_* (found by solving (4.5) with the boundary condition $v=0$ on $\mathcal{D}\times\{1,2\}$) captures both the optimal switching states and the optimal steering angles [117].

Despite its frequent use, the risk-neutral planning has a significant drawback: it is indifferent to the level of variability in the distribution of times to target. The resulting μ_* might be impractical if the risk of significantly exceeding $\mathbb{E}[T_{\mu_*}]$ is high (e.g., in right-heavy-tailed distributions). To address this, we change the perspective and search for a policy α that maximizes the probability of reaching the target before a specified deadline $\hat{s}$. We refer to such α as a *risk-aware* (or, more precisely, as an $\hat{s}$-*threshold-aware*) policy.

Letting $\Omega_{\bar{s}} = \Omega\times[0,\bar{s}]$, we define our new value function

$$w(\hat{\boldsymbol{\xi}},\hat{q},\hat{s}) = \sup_{\alpha} \mathbb{P}\left(T_\alpha(\hat{\boldsymbol{\xi}},\hat{q}) \le \hat{s}\right), \quad 0\le\hat{s}\le\bar{s}. \tag{4.8}$$

The supremum is taken over all measurable *threshold-aware* feedback policies $\alpha : \Omega_{\bar{s}} \to \Upsilon,\ (r,\theta,q,s) \to \alpha(r,\theta,q,s)$. If we treat $s(0)=\hat{s}$ as an *initial budget*, then the remaining time budget $s(t)$ is strictly decreasing along the path-to-target as t increases ($\dot{s}(t) = -1$ while continuously steering and a negative jump of C units of time when switching tack).

Consequently, the Stochastic Dynamic Programming Principle [62] yields an s-dependent quasi-variational inequality

$$w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \max \left\{ \sup_u \mathcal{E}_{u,\tau} w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}), \mathcal{E}_{C} w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) \right\} + o(\tau), \tag{4.9}$$

where

$$\mathcal{E}_{u,\tau}\ w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \mathbb{E}_{u,\tau}[w(\boldsymbol{\xi}(\tau), \hat{q}, \hat{s} - \tau) \mid \hat{\boldsymbol{\xi}}, \hat{s}], \tag{4.10}$$

$$\mathcal{E}_{C}\ w(\hat{\boldsymbol{\xi}}, \hat{q}, \hat{s}) = \mathbb{E}_{C}[w(\boldsymbol{\xi}(C), \tilde{q}, \hat{s} - C) \mid \hat{\boldsymbol{\xi}}, \hat{s}, f = 0]. \tag{4.11}$$

Similarly to the structure of (4.5), $\mathcal{E}_{u,\tau} w$ refers to the best probability of reaching the target before the deadline if we stay on the current tack for a small time τ while $\mathcal{E}_{C} w$ is the best probability if we immediately switch tacks.

From the stochastic Taylor expansion of Eq. (4.9), one can show that, if $w(r, \theta, q, s)$ is sufficiently smooth, it must satisfy

$$\max \left\{ \max_{u \in [0,\pi]} \left(\nabla w^{\top} \boldsymbol{\xi}_{\mathrm{d}} + \frac{\sigma^2}{2} \frac{\partial^2 w}{\partial \theta^2} - \frac{\partial w}{\partial s} \right), \mathcal{E}_{C} w - w \right\} = 0, \tag{4.12}$$

where $\nabla w = (\partial w / \partial r,\ \partial w / \partial \theta)$ and $\boldsymbol{\xi}_{\mathrm{d}}(r, \theta, q, u) = \Big(r_{\mathrm{d}}(\theta, q, u),\ \theta_{\mathrm{d}}(r, \theta, q, u; a) \Big)$.

As in the risk-neutral case, if we define $\psi_{r,q,s}(z) = w(r, z, \tilde{q}, s - C)$, then $\mathcal{E}_{C} w(r, \theta, q, s) = \mathcal{G}_{\theta}[\psi_{r,q,s}]$ following the definition of operator (4.7). Note that, in general, the value function does not have to be smooth or or even continuous. Nonetheless, even a discontinuous value function can be often interpreted as a unique *discontinuous viscosity solution* of a HJB PDE [8, Chapter 5], which allows recovering the optimal feedback policy $\alpha_*(r, \theta, q, s)$ as an argmax of (4.12).

4.4 Numerical Implementation

4.4.1 Semi-Lagrangian Discretization

We approximate the solution to (4.12) for both tacks simultaneously using a semi-Lagrangian discretization [54] on a uniform rectangular grid over the (r, θ, s) space. I.e., $(r_i, \theta_j, s_k) = (i\Delta r, j\Delta\theta, k\Delta s)$, where $\Delta r = R_{max}/N_r$, $\Delta r = 2\pi/N_\theta$, $\Delta s = \bar{s}/N_s$, while $i = 0, ..., N_r$, $j = 1, ..., N_\theta$ (due to periodic boundary conditions), and $k = 0, ..., N_s$. We will use $W_{i,j}^{k,q} \approx w(r_i, \theta_j, q, s_k)$ to denote the discretized approximate solution at (r_i, θ_j, q, s_k). Recall from section 4.3 that $s(t)$ is strictly decreasing along the path-to-target. We can thus *causally* march from smaller s to larger s and compute the value function from $s = 0$ to $s = \bar{s}$ in a single sweep. In particular, we choose $\tau = \Delta s$ when solving Eq. (4.10) so that we can march from the s_{k-1}-slice to the s_k-slice. The expectations in both (4.10) and (4.11) can be approximated using Gauss-Hermite quadratures (GHQ), but the details are somewhat different due to the contrast in elapsed times.

To approximate $\mathcal{E}_{u,\tau}$, we use a first-order weak approximation [58, 99] of the distribution of the Brownian increment ΔB_τ for τ time units. Starting from any gridpoint (r_i, θ_j, s_k) and using any admissible steering angle u, we consider two possible locations of $\boldsymbol{\xi}(\tau, u)$ in the s_{k-1}-slice:

$$\boldsymbol{\xi}_{i,j,u}^{\pm} \;=\; (r_i + \tau r_{\mathrm{d}}(u),\, \theta_j + \tau\theta_{\mathrm{d}}(u) \pm \sigma\sqrt{\tau}).$$

Averaging the value function at these points is equivalent to a two-node GHQ approximation of (4.10); i.e.,

$$\mathrm{M}_{u,\tau} W_{i,j}^{k,q} \;=\; \frac{1}{2}\Big(W^{k-1,q}\left(\boldsymbol{\xi}_{i,j,u}^{+}\right) + W^{k-1,q}\left(\boldsymbol{\xi}_{i,j,u}^{-}\right)\Big). \tag{4.13}$$

Since these $\boldsymbol{\xi}_{i,j,u}^{\pm}$ are usually not gridpoints, we implement (4.13) by using a *bi-cubic ENO* interpolation [152] with a 2π-periodicity in θ. We adopt a two stage process for finding the

optimal u_* that maximizes $\mathrm{M}_{u,\tau}$: first, we perform a direct comparison over a grid of angle values $\mathcal{U}$ and identify an interval containing the best $u_\star \in \mathcal{U}$; we then perform a Golden Section Search (GSS) over that interval to obtain a more accurate approximation of u_*.

The accuracy of (4.13) improves under grid refinement[1] since the diffusion time $\tau = \Delta s \to 0$. Finding a good approximation for $\mathcal{E}_{_C}$ is a bit harder since the diffusion time C is constant. To address this, one can use a higher order accurate GHQ; e.g., our implementation uses a version with 3 GH nodes

$$\eta_{j,m} \;=\; \theta_j + aC + \sigma\sqrt{2C}x_m, \quad m \in \{1,2,3\}, \tag{4.14}$$

where x_m are the roots to the 3rd Hermite polynomial. Assuming that $s_k \geq C$, we use

$$\mathrm{M}_{_C} W_{i,j}^{k,q} \;=\; \frac{1}{\sqrt{\pi}} \sum_{m=1}^{3} \gamma_m \, W\left(r_i, \eta_{j,m}, \tilde{q}, s_k - C\right), \tag{4.15}$$

where γ_m's are the weights of the third GHQ. We choose Δs to be a fraction of C, ensuring that $s_k - C = s_l$ for some $l < k$. But $\eta_{j,m}$ are usually not multiples of $\Delta\theta$ and we use a 1D periodic cubic ENO interpolation to evaluate (4.15).

The grid value is then computed as

$$W_{i,j}^{k,q} = \max\left(M_{u_*,\tau} W_{i,j}^{k,q},\, \mathrm{M}_{_C} W_{i,j}^{k,q}\right),$$

and we recover the optimal steering/switching policy $\alpha_*(r_i, \theta_j, q, s_k)$ as a by-product. Our full method is summarized in Algorithm 2 using the target radius $R_{\mathcal{D}}$, the maximum sailboat speed $f_{\max}$, and $\Xi = \{(i\Delta r, j\Delta\theta) \mid i = 0, \ldots, N_r,\ j = 0, \ldots, N_\theta\}$.

[1]Under mild technical conditions, semi-Lagrangian schemes have been proven to converge to the discontinuous viscosity solution of first-order HJB PDEs on every compact set away from discontinuity [10, 9]. For the second-order HJBs, the method closest to ours has been studied (with rigorous error estimates) in [6] but without hybrid dynamics or degenerate parabolicity. While our setting is more general, the numerical results in Section 4.5 and online repository provide strong evidence of convergence. A rigorous proof of numerical convergence to the discontinuous viscosity solution of PDE (4.12) remains an open problem to be addressed in the future.

Algorithm 2: Risk-aware value function computation

for $s_k = k\Delta s,\ k = 0, 1, \ldots N_s$ **do**
  for *every* $\boldsymbol{\xi}_{i,j} \in \Xi$ *and* $q \in \{1, 2\}$ **do**
    if $(r_i - R_{\mathcal{D}})/f_{\max} > s_k$ **then**
      $W_{i,j}^{k,q} \leftarrow 0$;
    else
      $W_{i,j}^{k,q} \leftarrow \max_u M_{u,\tau} W_{i,j}^{k,q}$;
      if $s \geq C$ **then**
        $W_{i,j}^{k,q} \leftarrow \max\left(W_{i,j}^{k,q},\ \mathrm{M}_C W_{i,j}^{k,q}\right)$;

4.4.2 Trajectory synthesis and ECDF generation

The above PDE solution process yields the optimal threshold-aware policy in feedback form, with the optimal action $\alpha_*(r_i, \theta_j, q, s_k)$ stored at each gridpoint in Ξ and for a range of deadlines ($k = 0, ..., N_s$). To recover a sample path-to-target from any specific initial configuration $(\hat{r}, \hat{\theta}, \hat{q})$ and the intended deadline $\hat{s}$, we use Euler-Maruyama scheme [99] with a fixed time step Δt on Eqs. (4.4) . At each time step, we normally use the optimal steering/switching action from the policy recorded for the nearest gridpoint. But the threshold-aware control formulation leaves two ambiguities that have to be resolved in the implementation. First, if $w(\hat{r}, \hat{\theta}, \hat{q}, \hat{s}) = 1$, the current $\hat{s}$ may be more than sufficient to reach the target with probability one and the actions taken until the remaining time budget s becomes "barely sufficient" are not important. To address this, we use a "*Deadline-Upgrade*" approach, decreasing the initial time-budget to $\hat{s} = \min\left\{s_k \mid w(\hat{r}, \hat{\theta}, \hat{q}, s_k) = 1\right\}$. Second, if during a simulation the sailor is unlucky and later finds herself with $w(\hat{r}, \hat{\theta}, \hat{q}, \hat{s}) = 0$, the PDE provides no guidance on what to do after that (since she will now definitely miss the original deadline). Rather than dismiss such simulations as complete failure, from there on we simply apply the risk-neutral policy μ_* recovered from Eq. (4.5).

Empirical cumulative distribution function (ECDF) for both α_* and μ_* are obtained through Monte Carlo simulations, with sample paths generated starting from a specific

$(\hat{r}, \hat{\theta}, \hat{q}, \hat{s})$ under different realizations of wind evolution. The ECDFs for the total time-to-target are obtained through the the Kaplan-Meier estimate [93] using the `MATLAB`'s built-in function `ecdf()`.

4.5 Numerical Experiments

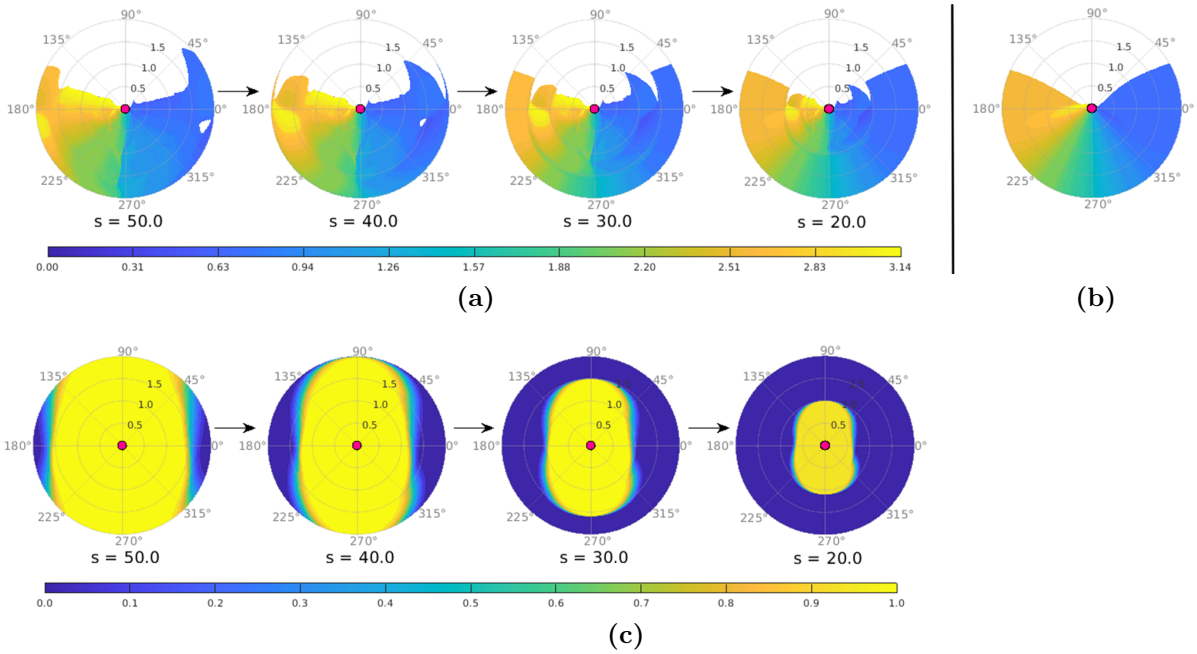
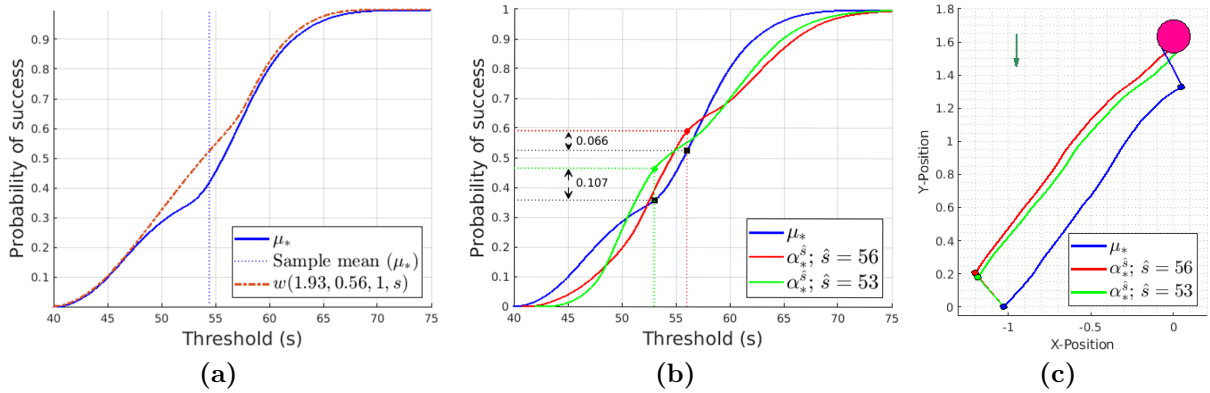


Figure 4.2. Representative s-slices of risk-aware policy, their corresponding optimal probability of reaching the target $\mathcal{D}$, and the risk-neutral policy with $a = 0$ and $\sigma = 0.05$: (a) risk-aware optimal policy α_*; (b) risk-neutral optimal policy μ_*; (c) optimal probability of reaching $\mathcal{D}$ associated with (a). All shown for the starboard tack $q = 1$ only and in relative (r, θ) coordinates. In all figures, the target $\mathcal{D}$ is shown as a magenta disk in the center. In (a) and (b), it is optimal to switch to $q = 2$ from wherever the space is left *blank*. Otherwise, it is optimal to stay with $q = 1$ and the best steering angle is shown in color (with the same colorbar used in both subfigures).

We use several examples to compare the performance of risk-aware and risk-neutral policies. In all cases, the value functions are computed on a $1601 \times 1601 \times 2$ grid for $(r, \theta, q) \in [0, 2] \times [0, 2\pi] \times \{1, 2\}$. When solving (4.12), we use $\Delta s = 0.025$ for any preset maximum deadline $\bar{s}$. The other parameter values are $R_{\mathcal{D}} = 0.1$, $C = 2$, and $f_{\max} = 0.05$.



(a) (b) (c)

Figure 4.3. Sailing against the wind: a comparison between the risk-aware and risk-neutral approaches with $(a = 0, \sigma = 0.05)$ **starting from** $(\hat{r}, \hat{\theta}, \hat{q}) = (1.93, 0.56, 1)$ Subfigures: (a) ECDF (empirical cumulative distribution function) generated with the optimal risk-neutral policy μ_* (*solid blue*) vs. the s-dependent risk-aware optimal probability of success $w(\hat{r}, \hat{\theta}, \hat{q}, s)$ (*dash-dotted orange*); (b) ECDFs of the random total time to target generated with different policies; (c) Sample sailboat trajectories in the *absolute xy-coordinates* generated with different policies under the same random wind path. The target set is plotted as a magenta disk at the top, and top-left *dark green arrow* encodes the initial wind direction $\phi(0) = 0$. Trajectory colors correspond to the policies used to generate ECDFs in (b). The colored dots indicate the tack-switching points for respective trajectories. Observed sample means for the arrival time T are 54.46, 55.57, and 55.95 for the policies μ_*, α_*^{53}, and α_*^{56} respectively. Arrival times for the specific trajectories in (c) are 56.55 (blue), 53.54 (green), and 54.07 (red).

All ECDFs are built through Monte Carlo simulations (see section 4.4.2) with 10^5 samples and $\Delta t = 0.005$.

We illustrate our s-dependent risk-aware policies for a wind with zero drift ($a = 0$) and small diffusivity $(\sigma = 0.05)^2$. In Fig 4.2(a,c), we show some representative s-slices of the risk-aware optimal policies α_* and the corresponding value function w for the starboard tack[3] (i.e., the optimal probability of reaching $\mathcal{D}$ in less than s units of time if we start with $q = 1$ and use α_*). In Fig 4.2(a), colors indicate the optimal steering angle u_* in the current tack, while the complement (left blank) shows all the (r, θ) configurations at which the immediate tack-switch ▲ is optimal. We observe that α_* is strongly s-dependent

[2]Examples with different a and σ values are provide in §4.7.1.

[3]In the interest of reproducibility, our full source code, additional examples, and movies (for both tacks) are available from https://eikonal-equation.github.io/Threshold-Aware-Sailing-Public.

and significantly differs from the risk-neutral optimal policy μ_* shown in Fig 4.2(b). The arrows in Fig 4.2(a,c) indicate the natural progression when threshold-aware policies are used in practice: once we start with a particular deadline $\hat{s}$, our initial time-budget[4] $s(0) = \hat{s}$ is progressively decreasing, making it necessary to use α_* from the lower s-slices. This decrease is gradual ($\dot{s}(t) = -1$) while we are steering, but becomes abrupt whenever we decide to tack-switch at some time $t^{\blacktriangle}$; i.e., $s(t) = s(t^{\blacktriangle}) - C, \quad \forall t \in (t^{\blacktriangle}, t^{\blacktriangle} + C]$.

Since α_* is somewhat more complicated to implement in practice, it is reasonable to ask whether it is significantly better (in meeting the desired deadlines) than the risk-neutral μ_*. For any specific starting configuration, the answer can be found by comparing the ECDF of μ_* with the risk-aware value function w plotted across the range of s values. The graph of w will be always above, though often this difference is minimal, making the use of μ_* a preferred option. However, in many cases the gap between the two graphs will be more significant for a specific range of s values. This is illustrated in Fig 4.3(a) for $(\hat{r}, \hat{\theta}, \hat{q}) = (1.93, 0.56, 1)$. If we are interested in some deadline between $\hat{s} = 52$ and $\hat{s} = 57$, the threshold-aware policies provide a noticeable advantage. For example, our α_*^{56} (red in Fig 4.3(b)) increases $\mathbb{P}(T \leq 56)$ from 52.5% to 58.9%, while our α_*^{53} (green in Fig 4.3(b)) yields a 10.6% improvement in $\mathbb{P}(T \leq 53)$ while increasing $\mathbb{E}[T]$ by less than 2.8%.

It is also revealing to examine sample trajectories resulting from each of these policies (shown in Fig 4.3(c) for a particular random realization of wind evolution). According to μ_*, the boat starts in the "wrong" tack, and thus needs to switch immediately, with another tack-switch (back to $q = 1$) almost always needed later to reach $\mathcal{D}$. This strategy produces the best $\mathbb{E}[T]$, but makes it hard to reach the target much earlier and does not hedge against the bad outcomes (e.g., $T_{\mu_*} > 58$ in more that 33% of simulations). In contrast, the threshold-aware policies make a calculated bet (that the wind direction will soon change to help us),

[4]In the following discussion, we use the superscript ($\alpha_*^{\hat{s}}$) to refer to a version of policy α_* implemented with a specific initial time-budget $\hat{s}$.

stay with the original $q = 1$ at first, and reach $\mathcal{D}$ with only one tack-switch.

Larger improvements can be similarly realized with a non-zero wind-drift, particularly when the chosen deadlines are fairly aggressive (in the left tail of T_{μ_*} PDF). In Fig 4.4(a) we show such an example with ($a = 0.05$, $\sigma = 0.05$). Unlike in Fig 4.3, here the initial direction of the wind is largely toward the target, but we are in the wrong initial tack to fully take advantage of this. The risk-neutral μ_* prescribes an immediate tack-switch followed by another one a bit later and yields a low $\mathbb{P}(T \leq 42) \approx 5.8\%$. In contrast, the threshold-aware α_*^{42} recognizes, based on the sign of a, that the wind is likely to change in the right direction soon and (in the specific wind-evolution example presented in the bottom row of Fig 4.4(a)) manages to reach $\mathcal{D}$ without any tack-switches at all. The result of this calculated bet is to almost triple $\mathbb{P}(T \leq 42)$ to 17.2% and make the tack-switches far less common. Even more dramatic improvements can be obtained when the drift is stronger. In Fig 4.4b with ($a = 0.15$, $\sigma = 0.05$), the threshold-aware policy boosts $\mathbb{P}(T \leq 43.5)$ from 8.8% to 26.6%, largely by reducing the number of tack-switches (in most cases, from 3 switches under μ_* to only 1 under $\alpha_*^{43.5}$).

We end this section with two caveats. First, it is usually impossible to optimize the entire CDF of the arrival time. As should be clear from Fig 4.3(b), a policy increasing $\mathbb{P}(T \leq \hat{s}_1)$ might be decreasing $\mathbb{P}(T \leq \hat{s}_2)$ even compared to a risk-neutral μ_*. Typically, each $\alpha_*^{\hat{s}}$ is only optimal for its particular threshold/deadline $\hat{s}$. This is why we do not use the usual nomenclature of *risk-aversion* [170] and instead describe or methods as risk (or threshold) *aware*. Second, our decision to revert to μ_* in the "unlucky" α_*-based simulations (once the time-budget is reduced to zero) is fairly arbitrary and one can certainly use other approaches instead. However, this choice does not affect $\mathbb{P}\left(T_{\alpha_*^{\hat{s}}} \leq \tilde{s}\right)$ for any $\tilde{s} \leq \hat{s}$; thus, the primary goal of threshold-aware policies is still achieved.

4.6 Discussion

We have introduced a robust (risk/deadline-aware) approach to controlling a sailboat in stochastically evolving wind conditions. The efficiency of our approach hinges on the numerical method for a pair of quasi-variational HJB-type inequalities, which yield deadline-aware policies for all initial configurations and a broad range of deadlines simultaneously. Numerical experiments demonstrate the advantages of these policies over the traditional risk-neutral approach [117], particularly when it is possible to reduce the number of likely tack-switches.

Several extensions will obviously increase the impact of this approach in the future. Solving the problem in absolute coordinates will allow for a better modeling of the domain geometry (e.g., accounting for obstacles and other target shapes). Incorporating more realistic wind models and more detailed boat dynamics will be clearly of interest to practitioners. Similarly, stochastic differential games might be used to reflect the competitive aspect of sailing races [29]. [E.g., if T_i is a (random) arrival time of the i-th competitor, one could try to maximize $\mathbb{P}\left(T \leq \min_i T_i\right)$.] In addition, it will be important to explore multi-objective versions (e.g., Pareto-optimal tradeoffs between $\mathbb{E}[T]$ and $\mathbb{P}\left(T \leq \hat{s}\right)$) and compare our approach with risk-averse methods that minimize the "Conditional Value at Risk" [170].

Finally, we hope that a similar threshold-aware approach will prove to be useful in many indefinite-horizon hybrid control applications unrelated to sailing.

4.7 Supplementary Materials

4.7.1 Optimal policies with non-zero drift

In this subsection, we provide additional s-dependent risk-aware policies α_* and the corresponding value functions w for the starboard tack[5] using the same format as Fig 4.2 with

- Fig 4.5: $a = 0.05$ and $\sigma = 0.05$;
- Fig 4.6: $a = 0.15$ and $\sigma = 0.05$;
- Fig 4.7: $a = 0.05$ and $\sigma = 0.1$.

On the top row of all three figures, colors indicate the optimal steering angle u_* in the current tack, while the complement (left blank) shows all the (r, θ) configurations at which the immediate tack-switch ▲ is optimal. We again observe that α_* is strongly s-dependent and significantly differs from the risk-neutral optimal policy μ_* shown in sub-panels (b).

When $a > 0$, the probability of success appears highly concentrated across all three scenarios, indicating a high gradient in w. Increasing σ only slightly accelerates the dispersion/smoothing rate; see the comparison between Fig 4.5(b) and Fig 4.7(b).

As the drift increases, we observe that the switchgrid and the contours of w increasingly skew towards negative θ. This indicates that it is optimal to navigate the sailboat ahead of the wind's current direction, anticipating its future shifts. By doing so, the sailboat positions itself advantageously within the state space, maintaining the ability to navigate at angles that enable maximum speed as the wind conditions evolve.

[5]Movies of all additional examples (for both tacks) are available from https://eikonal-equation.github.io/Threshold-Aware-Sailing-Public.

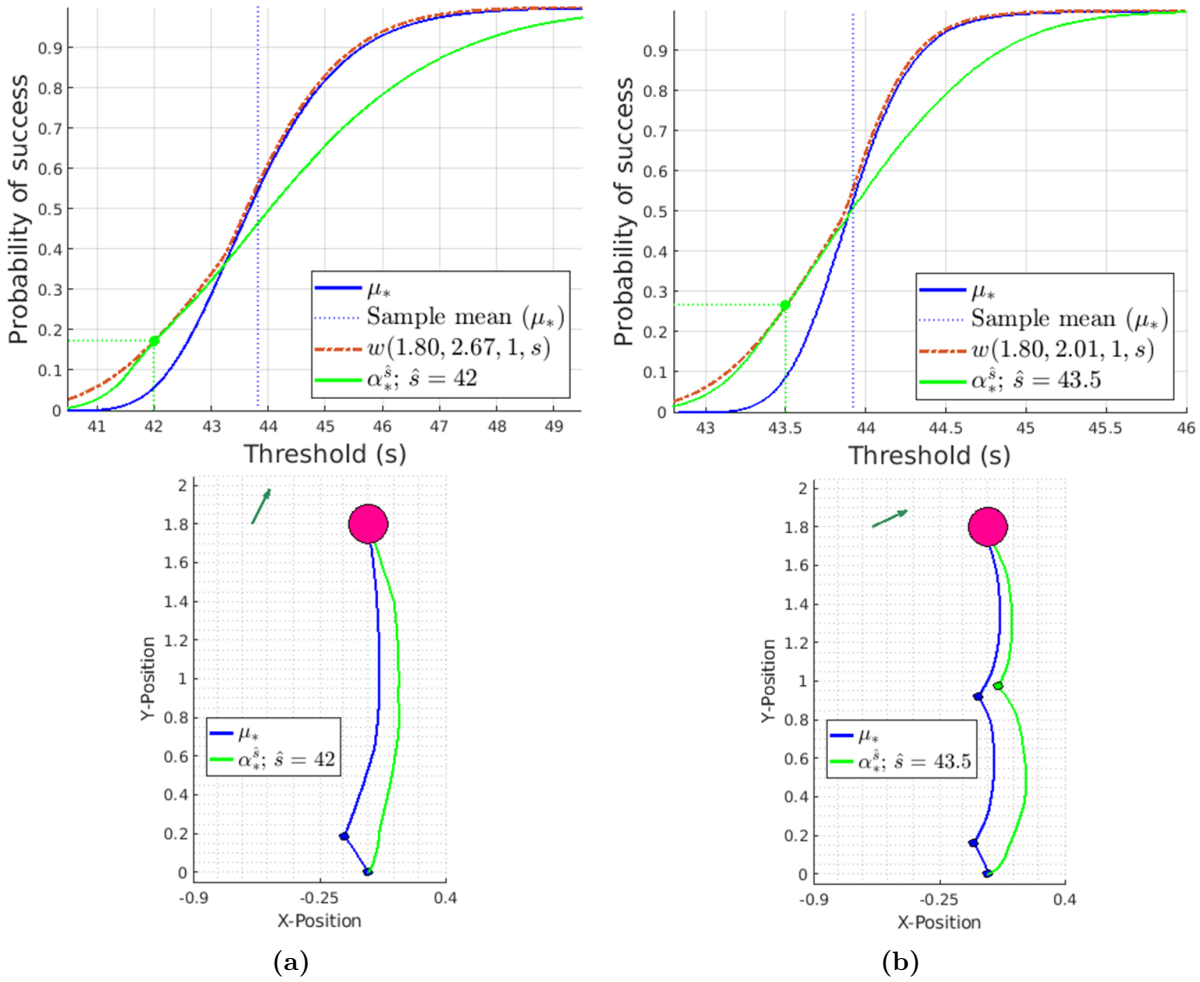


Figure 4.4. Exploiting the wind-drift: (a) ($a = 0.05$, $\sigma = 0.05$); (b) ($a = 0.15$, $\sigma = 0.05$). **Top Row:** ECDF for μ_* (*solid blue*), the s-dependent risk-aware optimal probability of success $w(\hat{r}, \hat{\theta}, \hat{q})$ (*dash-dotted orange*), and ECDF for $\alpha_*^{\hat{s}}$ (*solid green*). In (a), $(\hat{r}, \hat{\theta}, \hat{q}) = (1.80, 2.67, 1)$ and $\hat{s} = 42$. The sample means for μ_* and $\alpha_*^{\hat{s}}$ are 43.83 and 44.30. In (b), $(\hat{r}, \hat{\theta}, \hat{q}) = (1.80, 2.01, 1)$ and $\hat{s} = 43.5$. The sample means for μ_* and $\alpha_*^{\hat{s}}$ are 43.93 and 43.97. **Bottom Row**: Two representative paths generated with the same wind evolution (with colors corresponding to respective policies in the top row). The *dark green arrow* encodes the initial wind direction. *Time-to-target:* (a) blue: 42.23, green: 41.44 ; (b) blue: 43.12, green: 42.98. In (a) μ_* led to 2 tack-switches in 99.9% of simulations, while α_* required none in 99.1% of cases with 2 switches needed in all others. In (b) μ_* led to 3 tack-switches in 99.9% of simulations (with others requiring 4), while α_* required 1 switch in 99.9% of cases with 2 switches needed in the rest.

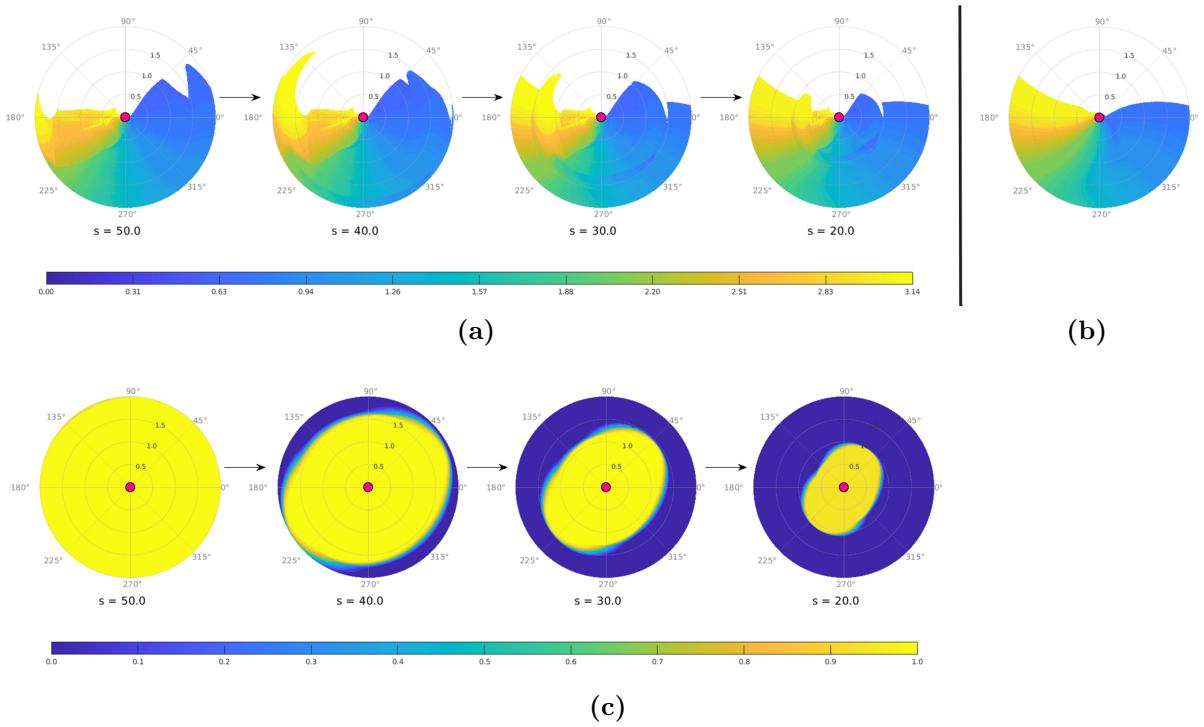


Figure 4.5. Representative s-slices of risk-aware policy, their corresponding optimal probability of reaching the target $\mathcal{D}$, and the risk-neutral policy with $a = 0.05$ and $\sigma = 0.05$: (a) risk-aware optimal policy α_*; (b) risk-neutral optimal policy μ_*; (c) optimal probability of reaching $\mathcal{D}$ associated with (a). All shown for the starboard tack $q = 1$ only and in relative (r, θ) coordinates. In all figures, the target $\mathcal{D}$ is shown as a magenta disk in the center. In (a) and (b), it is optimal to switch to $q = 2$ from wherever the space is left *blank*. Otherwise, it is optimal to stay with $q = 1$ and the best steering angle is shown in color (with the same colorbar used in both subfigures).

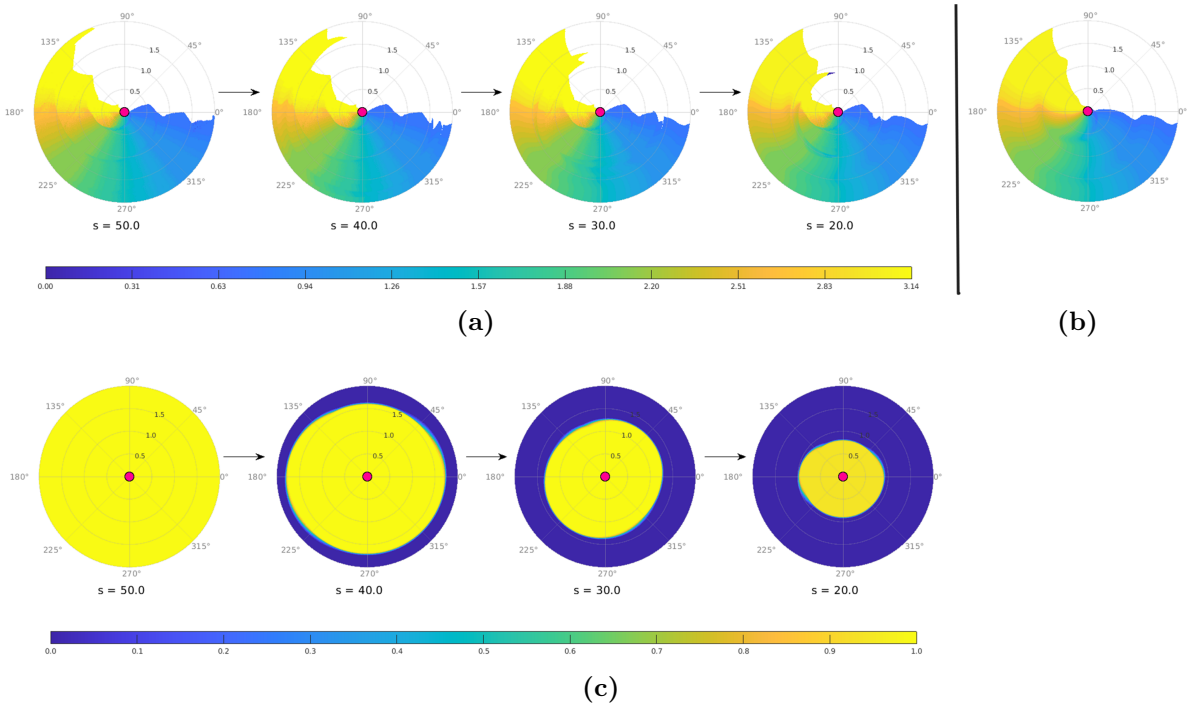


Figure 4.6. Representative s-slices of risk-aware policy, their corresponding optimal probability of reaching the target $\mathcal{D}$, and the risk-neutral policy with $a = 0.15$ and $\sigma = 0.05$: (a) risk-aware optimal policy α_*; (b) risk-neutral optimal policy μ_*; (c) optimal probability of reaching $\mathcal{D}$ associated with (a). All shown for the starboard tack $q = 1$ only and in relative (r, θ) coordinates. In all figures, the target $\mathcal{D}$ is shown as a magenta disk in the center. In (a) and (b), it is optimal to switch to $q = 2$ from wherever the space is left *blank*. Otherwise, it is optimal to stay with $q = 1$ and the best steering angle is shown in color (with the same colorbar used in both subfigures).

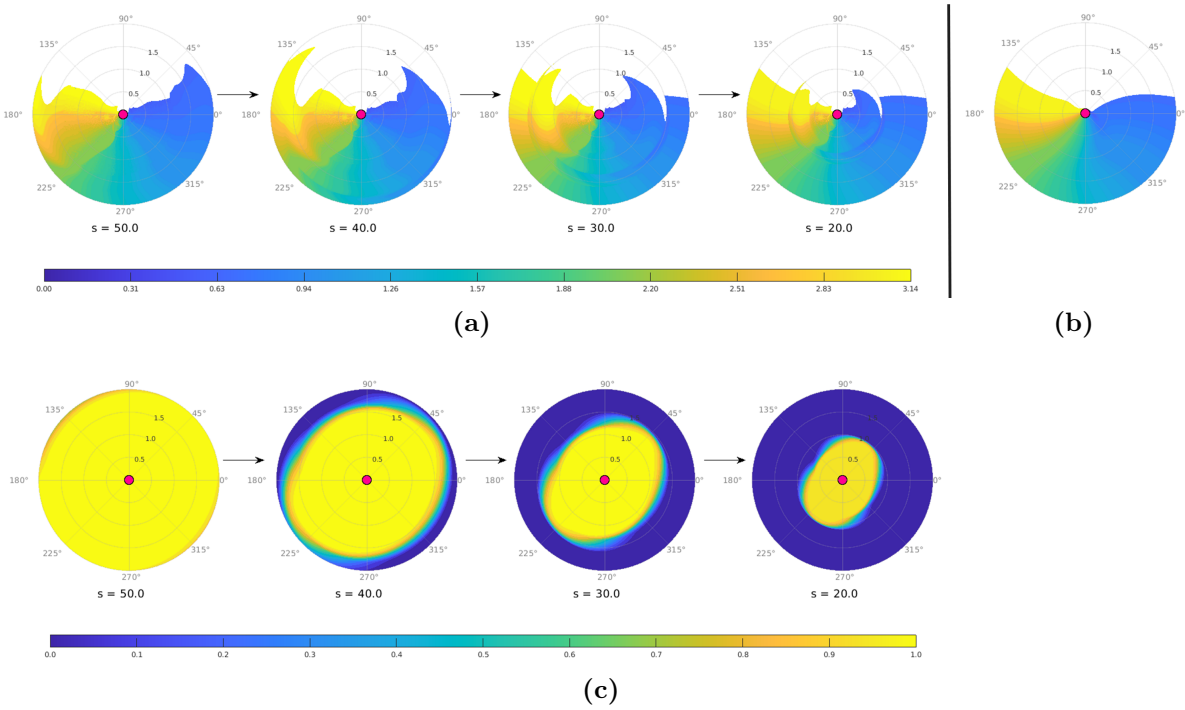


Figure 4.7. Representative s-slices of risk-aware policy, their corresponding optimal probability of reaching the target $\mathcal{D}$, and the risk-neutral policy with $a = 0.05$ and $\sigma = 0.1$: (a) risk-aware optimal policy α_*; (b) risk-neutral optimal policy μ_*; (c) optimal probability of reaching $\mathcal{D}$ associated with (a). All shown for the starboard tack $q = 1$ only and in relative (r, θ) coordinates. In all figures, the target $\mathcal{D}$ is shown as a magenta disk in the center. In (a) and (b), it is optimal to switch to $q = 2$ from wherever the space is left *blank*. Otherwise, it is optimal to stay with $q = 1$ and the best steering angle is shown in color (with the same colorbar used in both subfigures).

4.7.2 An initial "forced-to-switch"

In this subsection, we present a method to identify location(s) in the state space where the differences in the probability of success between implementing policies α_* and μ_* are most pronounced.

The idea is based on the observation that the probability of success is notably reduced (at a specific threshold $\hat{s}$) when μ_* prescribes additional tack-switches. This is particularly evident when an immediate tack-switch is recommended at the outset, as demonstrated in Figs 4.3 & 4.4. We thus propose computing an alternative value function, $\tilde{w}$, which represents the optimal probability of reaching the target within the deadline $\hat{s}$, subject to a new constraint: forcing an initial tack-switch for all state configurations (r, θ, s) where $s \geq C$, followed by optimal behavior thereafter. It is worth noting that $\hat{w}$ can be easily computed as a byproduct of solving for w. Adopting the same notations as in the main text, our full method is summarized in Algorithm 3 (with only a one-line modification to Algorithm 2) with $\widetilde{W}_{i,j}^{k,q} \approx \tilde{w}(r_i, \theta_j, q, s_k)$.

Algorithm 3: Risk-aware value function + "Forced-to-switch" initially

for $s_k = k\Delta s,\ k = 0, 1, \ldots N_s$ **do**
  for *every* $\boldsymbol{\xi}_{i,j} \in \Xi$ *and* $q \in \{1, 2\}$ **do**
    if $(r_i - R_{\mathcal{D}})/f_{\max} > s_k$ **then**
      $W_{i,j}^{k,q} \leftarrow 0$;
    else
      $W_{i,j}^{k,q} \leftarrow \max_u M_{u,\tau} W_{i,j}^{k,q}$;
      if $s \geq C$ **then**
        $\widetilde{W}_{i,j}^{k,q} \leftarrow \mathrm{M}_C W_{i,j}^{k,q}$;
        $W_{i,j}^{k,q} \leftarrow \max\left(W_{i,j}^{k,q}, \widetilde{W}_{i,j}^{k,q}\right)$;

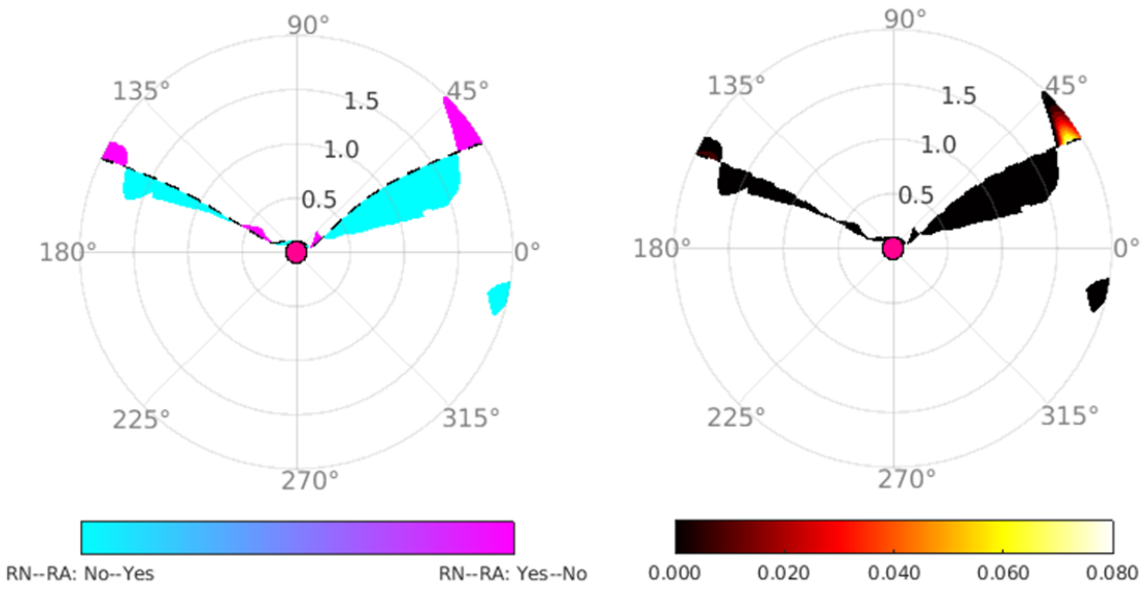
We calculate the difference in the switchgrid between policy α_* and μ_*, denoted as D_*, for all s. (Note that this results in D_* being s-dependent as well.) By "difference", we refer to instances where the risk-aware (RA) policy prescribes a "tack-switch", whereas the

risk-neutral (RN) policy does not, and vice versa. Following this, we compute the differences in the value functions $w - \tilde{w}$ over all states for $0 \leq s \leq \bar{\mathcal{S}}$. We then identify locations where the difference $w - \tilde{w}$ is significant within the regions defined by D_*.

In Fig 4.8(a), we present the switchgrid difference D_* at $\hat{s} = 53$, under the wind characterization $a = 0$ and $\sigma = 0.05$. Fig 4.8(b) illustrates the corresponding reduction in the probability of success, $w - \tilde{w}$, observed on D_*. It is clear that a significant discrepancy is near $r \approx 1.9$ and $\theta \approx 40^\circ$ in the magenta region (where RN prescribe a tack-switch while RA does not). This observation aligns with the example depicted in Fig 4.3 from the main text, confirming our intuition is correct.

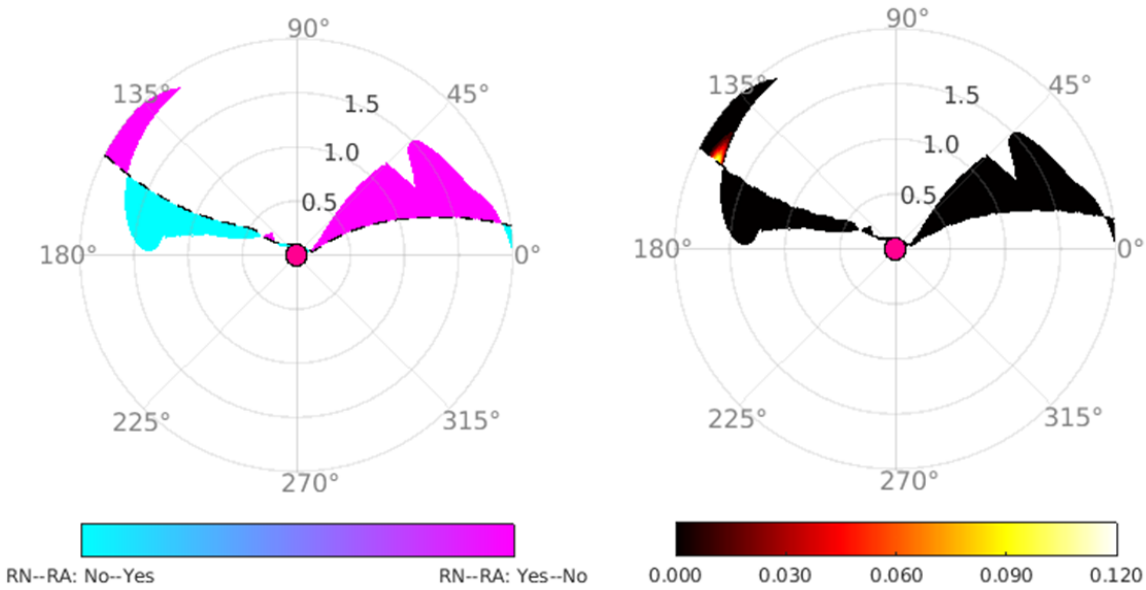
In Fig 4.9, we show the results of wind characterization with a non-zero drift ($a = 0.05$) while keeping $\sigma = 0.05$, at $\hat{s} = 42$ [6]. Fig 4.9(b) highlights the greatest reduction in the chance of success occurring near $r \approx 1.8$ and $\theta \approx 150^\circ$, which is again within the magenta region. This is consistent with the scenario described in the main text. Under conditions of a smaller budget and wind from behind, a strategically calculated bet based on a allows us to avoid unnecessary tack-switches. However, the presence of the magenta region does not always indicate a large discrepancy, particularly when the probability of success is already near 1, despite an earlier additional tack-switch.

[6] Movies of how D_* and $w - \tilde{w}$ change with s are available from https://eikonal-equation.github.io/Threshold-Aware-Sailing-Public.



(a) (b)

Figure 4.8. Forced to switch initially with wind characterization $a = 0$ and $\sigma = 0.05$ at $\hat{s} = 53$: (a) the switchgrid difference D_*; (b) success reduction in probability $w - \tilde{w}$. All shown for tack $q = 1$ only. In (a), the magenta region means the risk-aware (RA) policy does not prescribe a tack-switch while the risk-neutral (RN) policy does. The cyan region means the opposite. The boundary of the RN switchgrid is plotted with a black-dashed line.



(a) (b)

Figure 4.9. Forced to switch initially with wind characterization $a = 0.05$ and $\sigma = 0.05$ at $\hat{s} = 42$: (a) the switchgrid difference D_*; (b) success reduction in probability $w - \tilde{w}$. All shown for tack $q = 1$ only. In (a), the magenta region means the risk-aware (RA) policy does not prescribe a tack-switch while the risk-neutral (RN) policy does. The cyan region means the opposite. The boundary of the RN switchgrid is plotted with a black-dashed line.

CHAPTER 5

OVERCOMING TOXICITY: WHY BOOM-AND-BUST CYCLES ARE GOOD FOR TOXIN-SENSITIVE BACTERIA

5.1 Introduction

Antagonistic interactions are found throughout the microbial tree of life [130, 7, 139, 75, 159], in almost any environment [110, 148, 128, 142, 5] and host-associated microbiomes [144, 181, 66, 83]. Microbes have evolved a large variety of mechanisms to interact antagonistically with each other [130, 22], from contact dependent antagonism (e.g., via type IV, V and VI secretion systems), to short-distance interaction mediated by diffusible antimicrobial metabolites (e.g., bacteriocins), to long-range interaction via secretion of volatile antimicrobials [147]. These antagonistic interactions are thought to be a strong determinant of microbial community structure [130, 14], to provide benefits to hosts, such as protection against pathogen invasion [66], and to be a promising avenue for antimicrobial control in both natural ecosystems and animal hosts [56, 66, 155]. Recent experimental investigations have explored the ecological [115, 72, 38] and evolutionary dynamics of microbial antagonism [28] in the laboratory, where toxin-producing strains are found to dominate over toxin-sensitive ones. In many microbiomes, however, one finds both antagonistic and non-antagonistic microbes [74], raising the question of whether the dominance of toxin-producing strains observed in the laboratory might be caused by idealized growth conditions in those settings. Here, we explore how the interplay of environmental fluctuations, costs associated with toxin production, and regulation of toxin production, play different roles in determining the outcome of the competition. Although spatial heterogeneity of populations is one of the popular explanations of why toxin-sensitive strains might be doing well in practice [48, 97, 120, 75], we show that environmental fluctuations (e.g., dilutions) can favor toxin-sensitive strains

even in spatially homogeneous populations.

Ecological theory has long recognized the significant impact of environmental fluctuations on community composition [37, 88, 108, 34]. Changes in temperature, nutrient levels, and other abiotic factors critically shape the structure and dynamics of these communities. Such disturbances not only disrupt resident populations, possibly allowing new colonizers to invade, but also affect both immediate and long-term ecological outcomes [3, 112]. Microbial communities in a diverse array of habitats are subject to "boom-and-bust" dynamics, where periods of rapid population growth are often followed by sharp declines. Such dynamics have been observed across various environments, including phytoplankton and particle-attached microbial communities in marine ecosystems [15, 165, 42], soil [154, 153, 160, 12], host-associated microbiomes [138, 140, 164, 157], and the built environment [63, 69]. These boom-and-bust cycles are not only influenced by abiotic factors such as nutrient availability and environmental disturbances, but also by biotic interactions including competition, predation, and parasitism. Particularly, the interactions with phages and predators can drastically alter microbial community structure, trigger population crashes, and thereby influence the overall dynamics of microbial ecosystems [158, 150, 161, 26].

Despite the interest in the effect of microbial antagonism and environmental fluctuations on microbial community composition [70, 112, 121], the interplay between environmental fluctuations and antagonistic microbial interactions is surprisingly underexplored. Although data relating the relative abundance of toxin-producing strains to environmental fluctuations is very scarce, there is evidence that environments with higher turnover rates harbor a reduced number of toxin-producing strains [11, 105]. Mathematical modeling and laboratory experiments in stationary environments suggest that the efficacy of toxin-mediated killing is dependent on high population densities [72] and consequently frequent population busts could favor sensitive strains that avoid the metabolic costs associated with toxin production.

Certain microbial species have evolved regulatory mechanisms for toxin production that are triggered by quorum sensing signals [49] or environmental indicators suggestive of a stationary phase [124]. This regulation ensures that resources are not expended on toxin production at times when it would be least effective, possibly allowing microbes to optimize the cost-benefit ratio of toxin production. Activation of toxin production genes in response to quorum sensing signals produced by other strains, a phenomenon known as eavesdropping or cross-talk, has also been reported [118]. Regulation of toxin production in response to self and non-self abundances is thus theoretically possible and may be exploited to design synthetic genetic systems of pathogen eradication via targeted secretion of antimicrobials [143].

Our goal is to investigate how the regulation of toxin production, combined with antagonistic dynamics and environmental fluctuations, affects competition between toxin-producing and non-producing strains in environments characterized by frequent boom-and-bust cycles. This approach aims to elucidate the survival strategies of microbial populations and provide a deeper understanding of the ecological and evolutionary consequences of microbial interactions under variable environmental conditions.

5.1.1 Why are disruptions disruptive?

A simple mathematical model of a competition between toxin-producers and toxin-sensitives (to be described in detail in the next sections), shows that the former will eventually defeat the latter even if at first both strains are equally represented. This is because, near the carrying capacity, both strains have similarly low division rates, and the toxin will significantly affect the sensitive cells; see Fig 5.1. However, since the toxin-producers typically have a lower intrinsic division rate, there is an intermediate time period when the toxin-sensitives

are in the majority thanks to their higher growth rate at low densities. So, it is plausible that a significant disruption (e.g., a dilution) happening during that time will help the sensitives at least temporarily. To check whether this is the case, we turn to a biological experiment, in which the dilutions are performed regularly and affect both strains equally.

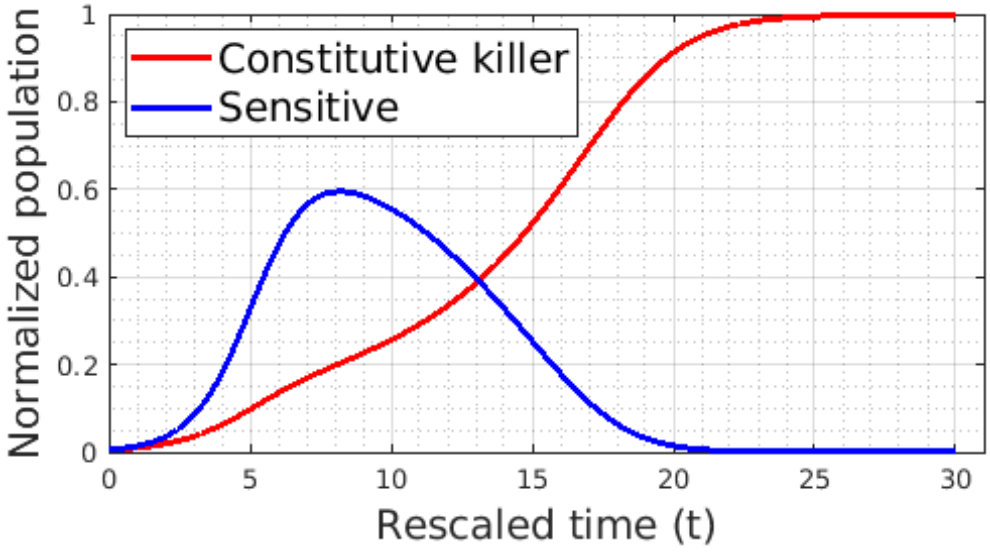


Figure 5.1. Mathematical model-based population trajectories for a strain of constitutive toxin-producers ($a = 1$ in (5.1)) and a strain of toxin-sensitive bacteria. Starting with 50% sensitive bacteria and a total population at 1% of the carrying capacity, sensitives initially grow much faster due to their higher intrinsic growth rate. However, as the overall population approaches the carrying capacity, the sensitives' intrinsic advantage shrinks, and the produced toxin leads to an eventual domination by the constitutive killers. But could the sensitives escape this fate if dilutions happen at an early stage, when the toxin-producers are still in the minority?

5.1.2 Experimental antagonism with periodic dilutions

To gain intuition for the impact of environmental disturbances on the dynamics of microbial antagonism, we conducted competition experiments between a sensitive (S) and a killer (K) strain of *Saccharomyces cerevisiae*, with the latter engineered to secrete the killer toxin K1 at experimentally adjustable rates [72].

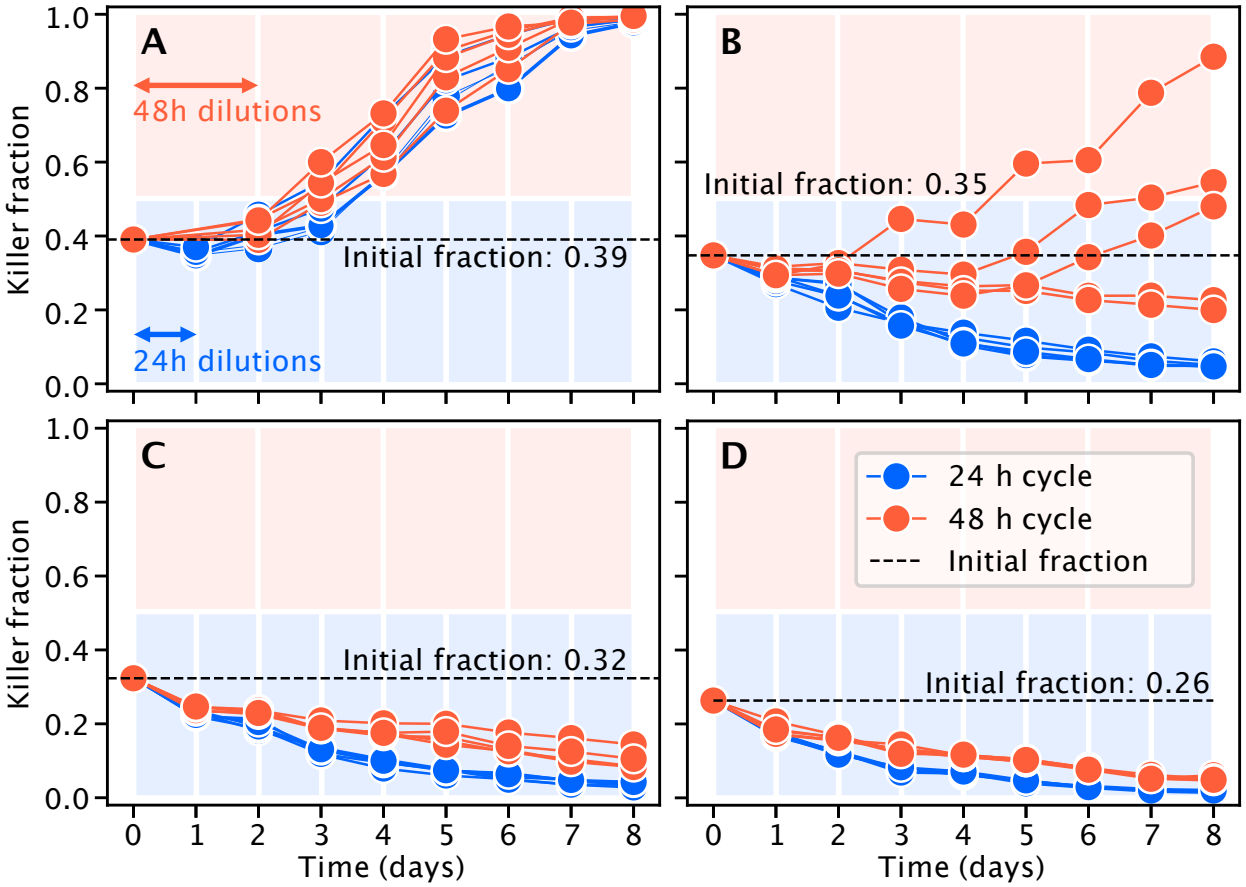


Figure 5.2. Experimental competitions between a toxin-producing (killer) and a sensitive strain of *S. cerevisiae* in environments diluted periodically with periods of $T = 1$ day (blue) and $T = 2$ days (orange). Different subfigures show different initial fractions of the killer strain. Both the initial killer fraction and the period of the dilution cycles determine the outcome of the competition and the rate of extinction of the losing strain.

In isolation and at low densities, the growth rates of strains S and K were 0.28 ± 0.01 h^{-1} and 0.26 ± 0.03 h^{-1} (mean±SD), respectively. At low cell densities, when the toxin is too dilute to significantly impact the growth of sensitive cells, one would thus expect the relative abundance of S to increase over time compared to K. At higher (killer) cell densities, however, one would expect the toxin to reduce the growth rate of the S strain due to increased cell death. This led us to hypothesize that frequent disturbance events, such as dilutions, could favor the sensitive strain allowing it to increase its relative abundance over time, despite the presence of the toxin-producing killer strain. Indeed, we found experimentally

that both the initial fraction and the interval between successive dilutions influenced the competition outcome between S and K, favoring S when dilutions occurred more frequently (Fig 5.2). Notably, the killer strain K tended towards extinction when dilutions occurred every 24 hours, starting from an initial killer fraction equal to 35% K versus S (Fig 5.2(b)). However, at 48-hour dilution intervals with 35% initial fraction of K versus S, three out of five replicates tended towards domination of the killer, while others tended towards domination of the sensitive strain. At other initial fractions, the outcome of competition (dominance vs extinction) was consistent between the 24 h and 48 h dilution cycles, with frequent dilutions either increasing the rate at which the killer went extinct, or slowing down the rate at which it became dominant. Overall, these experiments suggest that frequent dilutions can favor sensitive cells when their growth rates in isolation are larger than those of the killers. To more comprehensively characterize how growth rates, toxin production rates and their regulation, and environmental fluctuations jointly affect the dynamics of microbial antagonism, we now turn to mathematical modeling inspired by these experimental results.

5.1.3 Population dynamics

We start by describing a basic competition model between the toxin-producing "killer" strain (K) and the toxin-sensitive strain (S). The goal is to keep the modeled mechanisms general, although our main ideas can be similarly applied to more complex models tied to specific microorganisms or microbial communities.

We assume that the growth of both strains is logistic, with respective intrinsic growth rates $(r_{\mathrm{K}}, r_{\mathrm{S}})$ and a shared carrying capacity C. We assume that the killer's *ability* to produce toxin confers to it a growth rate deficit (i.e., $r_{\mathrm{K}} < r_{\mathrm{S}}$) independent of the actual toxin production. We also assume an additional growth rate deficit that scales linearly with

the normalized toxin production rate $a \in [0, 1]$, yielding the realized growth rate $r_{\mathrm{K}}(1 - \varepsilon a)$, where $\varepsilon > 0$ is the cost associated with producing toxin at maximal rate. The rate of toxin-induced death of the sensitive strain is assumed to be proportional to the product of the strain densities (n_{K} and n_{S}, respectively, assuming mass-action kinetics) with the killing rate μ. After non-dimensionalizing, $n_{\mathrm{K}}(t) \to n_{\mathrm{K}}(t)/C$, $n_{\mathrm{S}}(t) \to n_{\mathrm{S}}(t)/C$, and $t \to r_{\mathrm{S}}t$, the resulting dynamics are

$$\begin{cases} \dfrac{dn_{\mathrm{K}}}{dt}(t) = r_{\mathrm{KS}}(1 - \varepsilon a)\,(1 - n_{\mathrm{K}} - n_{\mathrm{S}})\,n_{\mathrm{K}}, \\ \dfrac{dn_{\mathrm{S}}}{dt}(t) = (1 - n_{\mathrm{K}} - n_{\mathrm{S}})\,n_{\mathrm{S}} - a\gamma n_{\mathrm{K}} n_{\mathrm{S}}, \end{cases} \tag{5.1}$$

where $r_{\mathrm{KS}} = r_{\mathrm{K}}/r_{\mathrm{S}}$ is the intrinsic growth rates ratio and $\gamma = \mu/(Cr_{\mathrm{S}})$ is the rescaled killing rate. For the sake of consistency, all rates in the rest of this paper are dimensionless (i.e., scaled by r_{S}). Distinguishing between the relative values of r_{KS} and ε in published datasets is challenging, as both parameters influence the realized growth rate of the killer strain. A review of various studies [172, 134, 33] that compared the growth rates of killer and sensitive strains suggests that the ratio of realized growth rates $r_{\mathrm{KS}}(1 - \varepsilon)$ typically falls between 0.68 and 0.98 for strains that produce the toxin at a constant, maximal rate (referred to as *constitutive killers*). Since proteinaceous toxins are often expressed from plasmids, one can estimate characteristic r_{KS} values by comparing the growth rates of *Escherichia coli* strains harboring such plasmids, with the toxin-producing genes deleted, to strains lacking these plasmids. From [171], we derive that $r_{\mathrm{KS}} \approx 0.85$ is a plausible value, based on comparisons of growth rates between cells with and without ColE1-type plasmids. In our experiments, the cost of constitutive toxin production is minor [72], but the titratable inducible system used to modulate the toxin production rate carries a metabolic cost that reduces r_K compared to r_S, resulting in the ratio of realized growth rates $r_K(1 - \varepsilon)/r_s = 0.92 \pm 0.03$ (mean±SE). In general, in addition to describing the cost associated with the ability to produce the toxin, r_{KS} may also reflect other differences in metabolism or genotype between killer and sensitive strains that may be unrelated to toxin production. Unless otherwise noted, we will adopt

$r_{\mathrm{KS}} = 0.85$ and $\varepsilon = 0.2$ in our computations, aligning with the lower bound $r_{\mathrm{KS}}(1-\varepsilon) = 0.68$ of the range reported in the literature.

It is also more convenient to restate the dynamics in terms of the normalized total population $N(t) = n_{\mathrm{K}}(t) + n_{\mathrm{S}}(t)$ and the fraction of killers $f(t) = n_{\mathrm{K}}(t)/N(t)$. This change of coordinates yields the ODE model (5.2) on a unit square, summarized in Box 7. Under this transformation, the entire horizontal line $N = 0$ maps to the origin $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0, 0)$ in the original (5.1) coordinates. Similarly, the entire horizontal line $N = 1$ corresponds to $n_{\mathrm{K}} + n_{\mathrm{S}} = 1$, indicating that the system is at carrying capacity. Fig 5.3(a) shows the phase portrait for constitutive killers, with all trajectories approaching $f = 1$ and $N = 1$ (or, alternatively, $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (1, 0)$ – the competitive exclusion of sensitives by killers), which is the only attracting fixed point of (5.2) for any fixed $a > 0$. It is worth noting that when starting from a small initial population N, the trajectories bend left (i.e., decreasing the fraction of killers) for a significant amount of time, which likely explains the strong initial reduction in the fraction of killers during the first growth cycle in many of our experiments. This reduction is due to killers' initial disadvantage ($r_{\mathrm{KS}}(1-\varepsilon) < 1$), which does not prevent their eventual domination once the population size gets closer to the carrying capacity.

Box 7: Population growth model

$$\begin{cases} \dfrac{\mathrm{d}f}{\mathrm{d}t} = \overbrace{f(1-f)\Big((1-N)\big[r_{\mathrm{KS}}(1-\varepsilon a)-1\big] + a\gamma f N\Big)}^{F(f,N,a)}, \\ \dfrac{\mathrm{d}N}{\mathrm{d}t} = \overbrace{N(1-N)\Big(1+\big[r_{\mathrm{KS}}(1-\varepsilon a)-1\big]f\Big) - a\gamma N^2 f(1-f)}^{G(f,N,a)}. \end{cases} \tag{5.2}$$

Definitions and Parameters:

- $f(t), N(t) \in [0,1]$: fraction of killers, normalized total population;
- $x = f(0)$, $y = N(0)$: initial killer fraction, initial total population;
- $r_{\mathrm{KS}} := r_{\mathrm{K}}/r_{\mathrm{S}}$: ratio between intrinsic growth rates;
- $a(t) \in [0,1]$: toxin-production rate;
- $\varepsilon > 0$: cost of producing the toxin;
- $\gamma = \mu/(Cr_{\mathrm{S}})$: rescaled killing rate.

5.2 Results

5.2.1 Do regular dilutions protect the sensitive?

Box 8: Periodic dilution events

Periodic dilution events: A deterministic process with a fundamental period of T.

- $(T, 2T, 3T, \ldots)$, an infinite sequence of dilution times;
- $\rho \in (0, 1)$, a surviving proportion of the population.

After the n-th dilution occurring at nT:

- $f\big((nT)^+\big) = f\big((nT)^-\big)$, relative abundances are preserved;
- $N\big((nT)^+\big) = \rho N\big((nT)^-\big)$, total population decreases;
- population grows until $t = (n+1)T$ according to (5.2).

See Fig 5.3(bc) for examples of multi-dilution trajectories.

We begin by assuming that dilutions occur regularly every T time units and that at each dilution the relative strain abundances are preserved, but only a fixed fraction ρ of the total population survives[1]. As in our experiments, we observe that with a relatively short waiting time ($T = 1$, i.e. dilution inter-arrival time equal to the inverse growth rate of sensitive cells), the killers may either progressively increase their relative abundance (see Fig 5.3(b)) or decrease it (see Fig 5.3(c)), depending on the initial condition. This suggests that while dilution events can disrupt the killers' dominance, a further investigation is needed to determine under which circumstances these interventions help the sensitives if T is small.

Numerical methods make it easy to analyze the performance of constitutive killers for all possible initial states over one cycle; i.e., we use linear partial differential equations (PDEs) to compute the pre-dilution $f(T^-) = f(T^+)$ and $N(T^-) = N(T^+)/\rho$ corresponding to all initial

[1]We focus here on such "strictly proportional" dilutions for the sake of simplicity and computational efficiency. The results in SI Appendix §5.4.6 show that for most initial conditions the conclusions remain largely the same even with a probabilistic dilution model, where each cell has probability ρ of surviving each dilution.

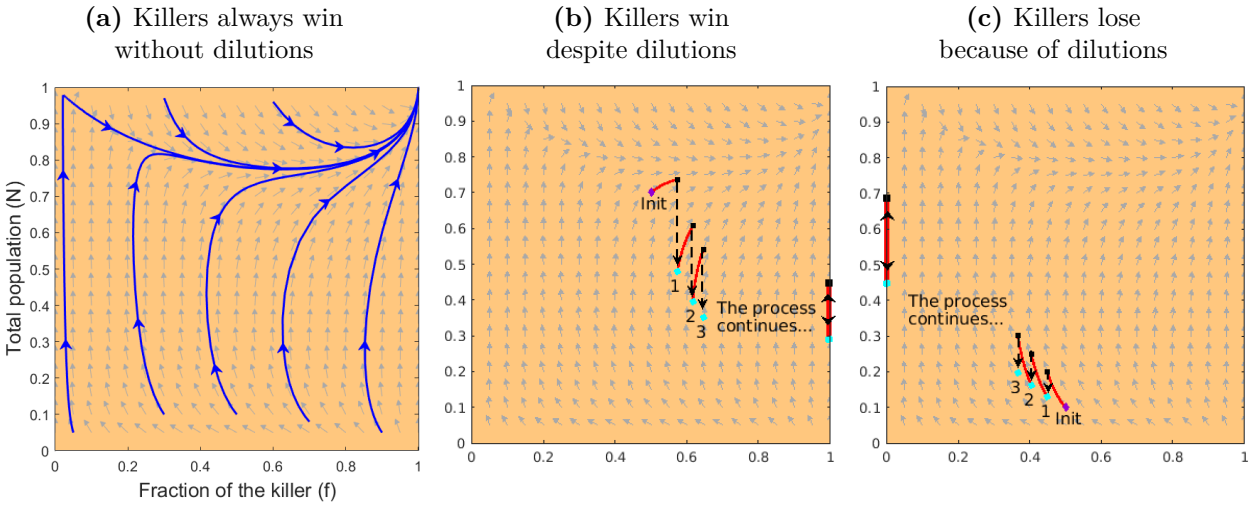


Figure 5.3. Trajectories of competitions between constitutive killers and sensitives in undisturbed and periodically-diluted populations. (a) Killers always win without dilutions. (b,c) With periodic dilutions, their fate depends on the initial condition. With a high enough initial population size, e.g., $(f(0) = x, N(0) = y) = (0.5, 0.7)$ (b), a sequence of dilutions carries killers to an eventual victory (i.e., $f \to 1$). Starting at a lower population size, e.g., $(x, y) = (0.5, 0.1)$ (c), the dilutions lead to their demise (i.e., $f \to 0$). In both cases, the period of dilutions is $T = 1$, and the pre- and post-dilution states are shown with black squares and cyan dots respectively. Once either strain dominates, the population oscillates between the terminal cyan dot and black square at $f = 0$ or $f = 1$. Temporal trajectories associated with subfigure (b) are shown in Fig 5.4(b).

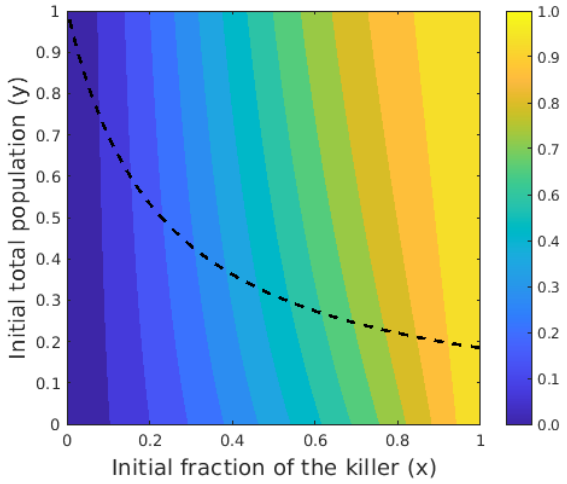
$(f(0) = x, N(0) = y)$; see Fig 5.4(a) and SI Appendix §5.4.4. Once this mapping is computed, we iterate it to determine the asymptotic outcomes as the number of dilutions approaches infinity. Fig 5.4(a) shows a general trend: larger initial killer fractions x and population sizes y allow constitutive killers to moderately increase their relative abundance by the end of the first cycle. The region below the black-dashed curve in Fig 5.4(a) indicates initial conditions for which killers decrease their fraction in the first cycle; i.e. $f(T^-) < x = f(0)$. Thus, one may expect that populations that start below the black-dashed curve are exactly the ones that become dominated by the sensitive strain with successive dilutions. However, our calculation of the limiting pre-dilution fraction of killer strains, $\bar{f}(x, y) = \lim_{n\to\infty} f((nT)^-)$, shows that this is not the case. Fig 5.4(c) shows that, in the limit of infinite dilutions, the state space is divided into two regions corresponding to the competitive exclusion of the

killer by the sensitive strain (blue) and vice versa (red). The actual shades of blue and red in Figs 5.4(c,d) represent the time it takes from the initial (x, y) to come within the machine accuracy of the limit $\bar{f}$ (which in real systems would be correlated with the time until the competitive exclusion). We highlight three features generic in these computations, which mirror the observations from the experiments:

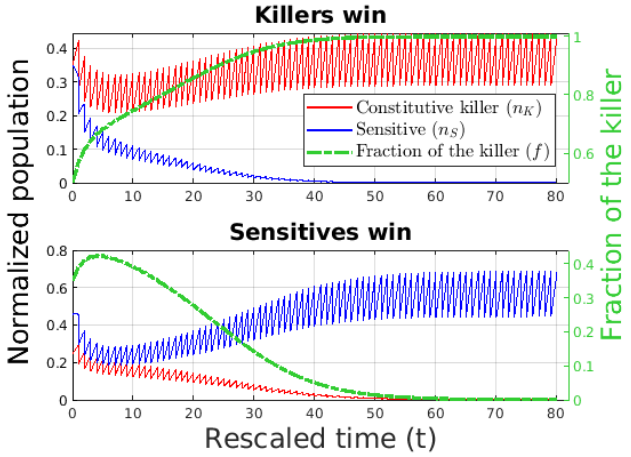
1. In Fig 5.4(c), the boundary of the blue-hued region is significantly different from the black-dashed curve (reproduced from subfigure (a)), showing that changes in relative abundance in the first cycle are not predictive of the asymptotic limit. Accordingly, in our experiments, the killer's fraction almost always decreased in the first cycle due to the low initial population size.
2. With dilutions occurring at timescales comparable to the inverse growth rate of the sensitive strain r_{S}^{-1} ($T = 1$ in nondimensional time), the sensitives can eventually prevail if they start in the majority (e.g., see the magenta marker in Fig 5.4(c)), or if the initial population size is sufficiently small.
3. When the dilution period grows, this might change the asymptotic limit and the time necessary to approach it. For example, the same magenta-marked initial condition leads to the victory of killers when $T = 2$ in Fig 5.4(d). While the cyan-marked initial condition was already leading to killers' victory even with $T = 1$, with $T = 2$ this exclusion happens much faster. This is consistent with the comparison of 24-hour and 48-hour dilution trajectories in the experiments (Fig 5.2).

In addition, with $T = 1$, neither strain can reach the carrying capacity within one cycle even after the other strain is excluded in the limit. The range of such single-strain oscillations can be found analytically (§5.4.2 in SI Appendix), is observed in Fig 5.4(b).

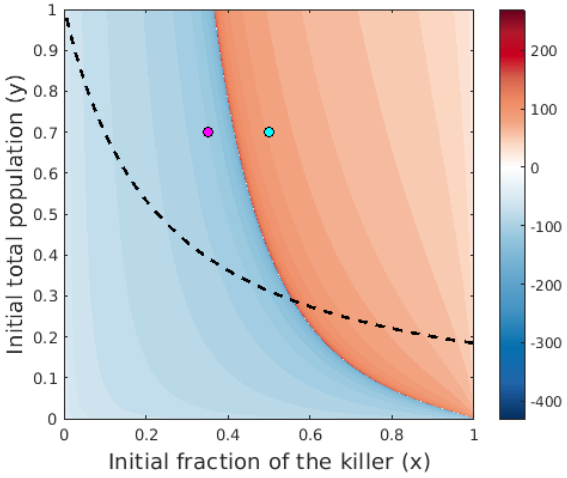
(a) Pre-dilution fraction of killers $f(T^-)$ for $T = 1$



(b) Two scenarios of competitive exclusion (different initial conditions, same $T = 1$)



(c) Time until competitive exclusion (with $T = 1$)



(d) Time until competitive exclusion (with $T = 2$)

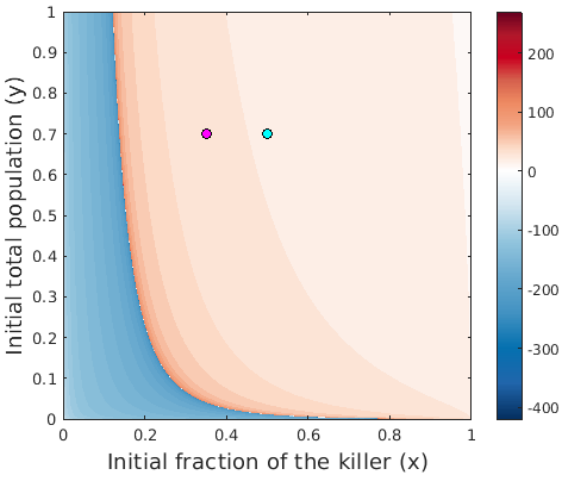


Figure 5.4. Constitutive killers: pre-dilution fraction and limiting behavior. (a) Pre-dilution fraction of the killer at the end of the first cycle, $f(T^-)$, for $T = 1$ and any initial condition ($f(0) = x, N(0) = y$). Initial conditions below the black, dashed curve lead to $f(T^-) < f(0)$. (b) Temporal trajectories of killer (blue) and sensitive (red) population sizes (left axes), and killer fraction (right axes), with two different initial conditions (corresponding to the magenta and cyan dots in subfigure (c)) and dilution period $T = 1$. Black dashed line as in subfigure (c). (c) Time until competitive exclusion (red/blue shades) and limiting killer fraction (red and blue indicate $f = 1$ and $f = 0$, respectively) vary with the initial condition. Within the red and blue regions, the absolute killer and sensitive population sizes reach $n_K \approx 0.45$ and $n_K \approx 0.69$, respectively (see also subfigure (b)). (d) Doubling the dilution period to $T = 2$ extends the range of initial conditions leading to domination by the killer and reduces/increases the timescale over which the killer/sensitive reach domination, respectively. Note that both initial conditions marked by dots now lead to killer domination. Parameter values: $\varepsilon = 0.2$, $r_{KS} = 0.85$, $\gamma = 1$, and $\rho = 0.65$.

5.2.2 Do toxin-producers benefit from population-sensing?

The results for constitutive killers are revealing and align well with our experiments. However, antagonistic strains often use quorum sensing or environmental signals to regulate toxin production, and thus may not engage in antagonistic behavior at all times. Here, rather than modeling specific types of quorum sensing mechanisms, we use a phenomenological approach and explore different notions of optimality for toxin production policies of "omniscient killers." The results will serve as an upper bound on how well more realistic killers could do given the limits to their sensing abilities.

Supposing that the killers can sense the current size and composition of the population, they might use this to regulate their rate of toxin-production, also taking into account the remaining time until the next dilution. More precisely, we will consider a theoretical possibility of their evolving the optimal toxin production rate *in feedback form*: $a_* = a_*(f, N, t)$ to optimize the resulting pre-dilution fraction $f(T^-)$. We will refer to such killers as "myopically-optimal" or simply "myopic" since they optimize the results over a single cycle only, without any regard to the sequence of future dilutions. We use the methods of control theory [61] to find this optimal policy. Using $u(x, y, t)$ to denote the best $f(T^-)$ still achievable when $(f(t) = x, N(t) = y)$, we can derive a time-dependent Hamilton-Jacobi-Bellman (HJB) PDE (5.3) satisfied by u and solve it numerically, recovering the optimal control policy $a_*(\cdot)$ as a byproduct (SI Appendix §5.4.4). The properties of that PDE guarantee that the optimal policy will generally be *bang-bang*; i.e., for generic (x, y, t), it will be optimal to either not produce the toxin at all ($a_* = 0$) or produce it at the maximum rate ($a_* = 1$). As Fig 5.5(c) shows, these myopic killers try to maximize the early exponential growth by opting not to produce the toxin at first if the initial population and/or their initial fraction are low. Fig 5.5(a) demonstrates that they do better than the constitutive killers over the first cycle, but at least for these parameter values, this sensing-based advantage is minor and mostly

pronounced when the initial populations are low. Once the optimal policy is identified, we can also find the corresponding pre-dilution population size $\phi(x, y, t) = N(T^-)$ from a similar linear PDE (5.4) and then iterate (u, ϕ) to obtain the limiting pre-dilution fraction of killers $\hat{u}^\infty(x, y)$ under an infinite sequence of dilutions (SI Appendix §5.4.4). Compared to constitutive killers, the region starting from which the killers eventually dominate expands only slightly, for large x and small y (Fig 5.5(b)). Consequently, the sensitive strain is still protected by periodic dilutions, which allow them to achieve competitive exclusion starting from a large fraction of initial configurations.

The periodic setting investigated up till now provides useful insights but may be overly simplistic for capturing behaviors across various ecosystems. For example, boom-and-bust dynamics are often driven by abiotic fluctuations that occur randomly in time, rather than periodically. In the next section we extend our analysis to randomly distributed dilution events.

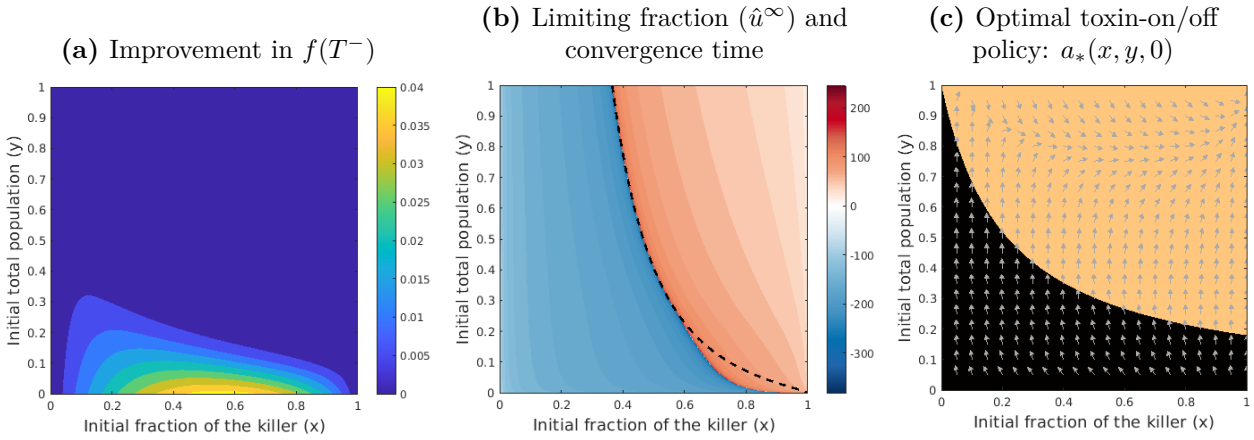


Figure 5.5. Myopic killers: pre-dilution and limiting behaviors with regular dilutions and $T = 1$. (a) Myopic population sensing provides a minor improvement to the killer frequency $f(T^-)$ at the end of the first cycle, mostly for initial conditions with low population size. Shown here is the maximized $f(T^-)$ for myopic killers, minus the corresponding $f(T^-)$ for constitutive killers. (b) In the infinite-dilution limit, myopic population sensing expands the set of initial conditions leading to the killer dominance, compared to constitutive killers (black dashed curve reports the blue/red boundary of Fig 5.4(c)). (c) The initial optimal toxin production strategy $a_*(x, y, 0)$ at $t = 0$ is *bang-bang*, equal to 1 in the orange region, and to 0 in the black region. Grey arrows denote the vector field corresponding to (5.2) with $a = a_*(x, y, 0)$.

Box 9: Myopic optimal policy for population-sensing killers under regular dilutions

The *value function* $u(x, y, t) = \sup_{a(\cdot)} f(T^-)$ is the best pre-dilution killer fraction achievable from $f(t) = x$, $N(t) = y$.

It satisfies a HJB PDE

$$-\frac{\partial u}{\partial t}(x, y, t) = \max_{a\in[0,1]} \left\{ \nabla u(x, y, t) \cdot \begin{bmatrix} F(x, y, a) \\ G(x, y, a) \end{bmatrix} \right\} \tag{5.3}$$

Terminal condition: $u(x, y, T) = x$ if $y > 0$ and $u(x, y, T) = 0$ otherwise.

- F and G are defined in (5.2);
- optimal toxin-production policy $a_*(x, y, t)$ is an argmax in (5.3).

The corresponding pre-dilution population size $\phi(x, y, t) = N(T^-)$ starting from $f(t) = x$, $N(t) = y$, and using policy $a_*(\cdot)$ satisfies

$$-\frac{\partial \phi}{\partial t}(x, y, t) = \nabla \phi(x, y, t) \cdot \begin{bmatrix} F\big(x, y, a_*(x, y, t)\big) \\ G\big(x, y, a_*(x, y, t)\big) \end{bmatrix} \tag{5.4}$$

with the terminal condition $\phi(x, y, T) = y$.

5.2.3 Who benefits from randomness in dilution times?

We now turn to model dilutions as random events governed by a Poisson process; so, the duration of inter-dilution time intervals $\mathcal{T}$ are independent exponentially distributed random variables with a fixed rate $\lambda > 0$. As before, we will consider proportional dilutions, preserving f but instantaneously switching from N to ρN. Mathematically, this continuous evolution of $f(t)$ and $N(t)$ punctuated by randomly timed jumps in N can be described as a Piecewise-Deterministic Markov Process (PDMP) [43] and we take advantage of a well-developed theory for optimal control of PDMPs throughout the rest of this paper. We first note that any toxin-production policies will now be independent of time since the last dilution; we will use $\alpha = \alpha(f, N)$ to denote such feedback policies, to distinguish them from $a(f, N, t)$ used in the periodic case above.

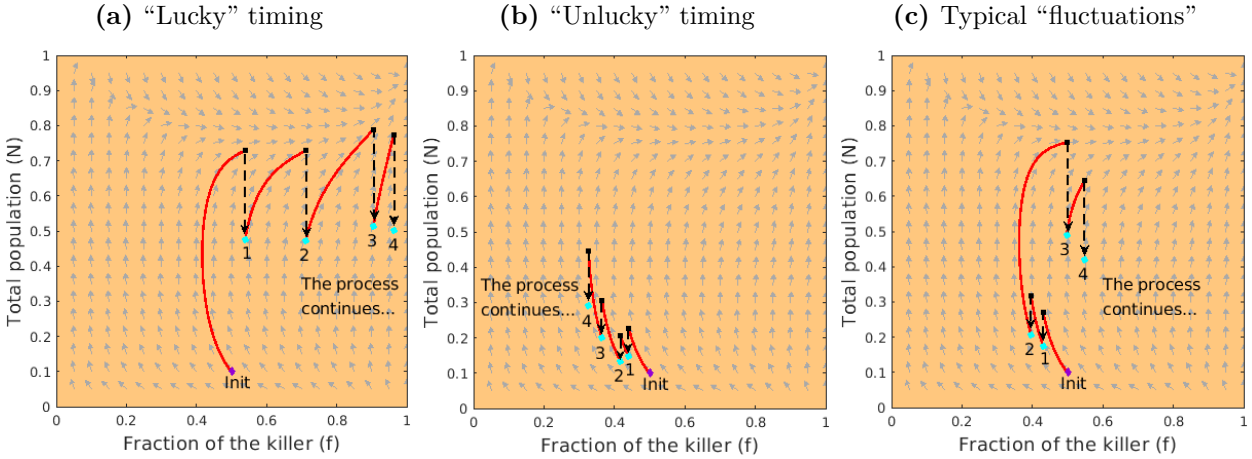


Figure 5.6. Constitutive killer with random dilution times. Killers can either progressively increase or decrease their fraction in a "lucky" or "unlucky" scenario depicted in subfigures (a, b), respectively. However, in most cases, their fraction will fluctuate instead (subfigure (b)). In all cases, the initial configuration $(x, y) = (0.5, 0.1)$ is plotted with a purple diamond and followed by 4 dilution events. The pre-dilution / post-dilution states are again shown with black squares and cyan dots, respectively.

Second, we note that our criteria for evaluating initial conditions and the quality of policies become more subtle. Starting from the same initial $(f(0) = x, N(0) = y)$ and using any reasonable fixed policy α, a sequence of randomly timed dilutions might lead to a competitive exclusion of either strain. Fig 5.6 illustrates this for the simplest case of constitutive killers; we will now use $\alpha_0 = 1$ to denote their policy. Since a trajectory reaches neither $f = 0$ nor $f = 1$ in finite time, we will select small threshold values γ_{d} and $(1 - \gamma_{\mathrm{v}})$, declaring killers' victory[2] as soon as $f(t) > \gamma_{\mathrm{v}}$ or sensitives' victory as soon as $f(t) < \gamma_{\mathrm{d}}$. We then define a new probabilistic metric for policy performance: $\hat{w}^{\alpha}(x, y)$ is now the probability of the killers attaining their victory (before the toxin-sensitives do) after an arbitrary number of dilutions starting from $(f(0) = x, N(0) = y)$ and using policy α. The rigorous definition and the PDE that $\hat{w}^{\alpha}$ must satisfy are covered in Box 10. We will use this metric to compare the performance of all toxin-production policies in stochastic environments.

[2]It is similarly possible to declare the victory/defeat criteria in terms of strain populations rather than fractions. For most initial conditions, this does not affect the qualitative policy and victory probabilities; see SI Appendix 5.4.5.

Since the inter-dilution intervals are random, this also affects the notion of optimal myopic policy. We will call the killers "stochastic-myopic" if they follow a policy α_1 chosen to maximize the *expected* killers' fraction just before the next dilution; i.e., $\mathbb{E}[f(\mathcal{T}^-)]$. This bang-bang policy can be found by solving the so-called "randomly-terminated" problem [4]. See the HJB PDE (5.11) in Box 11 and also SI Appendix §5.4.3 for the derivation.

Box 10: Probabilistic performance metric for policies

Definitions and Parameters:

- $\Delta_{\text{vic}} = \{(x,y) \in [0,1]^2 \mid x > \gamma_{\text{v}}\}$, victory zone ($\gamma_{\text{v}}$–victory threshold);
- $\Delta_{\text{dft}} = \{(x,y) \in [0,1]^2 \mid x < \gamma_{\text{d}}\}$, defeat zone ($\gamma_{\text{d}}$–defeat threshold);
- $\Delta = \Delta_{\text{vic}} \cup \Delta_{\text{dft}}$, terminal set.

(Random) victory time for the killer:

$$T_v(x,y,\alpha(\cdot)) = \inf\Big\{t > 0 \mid f(t) \in \Delta_{\text{vic}};\, f(0) = x,\; N(0) = y,\; \text{with many dilutions}\Big\}. \tag{5.5}$$

(Random) defeat time for the killer:

$$T_d(x,y,\alpha(\cdot)) = \inf\Big\{t > 0 \mid f(t) \in \Delta_{\text{dft}};\, f(0) = x,\; N(0) = y,\; \text{with many dilutions}\Big\}. \tag{5.6}$$

(Random) termination time:

$$T := T(x,y,\alpha(\cdot)) = \min\Big\{T_v\big(x,y,a(\cdot)\big),\; T_d\big(x,y,a(\cdot)\big)\Big\}. \tag{5.7}$$

Terminal cost: $g = \begin{cases} 1, & \text{if } (x,y) \in \Delta_{\text{vic}}, \\ 0, & \text{if } (x,y) \in \Delta_{\text{dft}}. \end{cases}$

Performance metric (probability of killers' victory with α):

$$\hat{w}^{\alpha}(x,y) = \mathbb{P}\Big(T_v\big(x,y,\alpha\big) < T_d\big(x,y,\alpha\big)\Big) \tag{5.8}$$

can be found by numerically solving a first-order linear equation:

$$\lambda\Big[\hat{w}^{\alpha}(x,\rho y) - \hat{w}^{\alpha}(x,y)\Big] + \left(\nabla\hat{w}^{\alpha}(x,y) \cdot \begin{bmatrix} F\left(x,y,\alpha(x,y)\right) \\ G\left(x,y,\alpha(x,y)\right) \end{bmatrix}\right) = 0, \tag{5.9}$$

with the boundary condition $\hat{w}^{\alpha} = g$ on Δ.

See **Remark 5.7** in SI Appendix §5.4.4 for the numerics.

We first focus on the case $\lambda = 1$, to ensure that $\mathbb{E}[\mathcal{T}] = 1/\lambda = 1$ matches the period of regular dilutions $T = 1$ considered in the previous section. Fig 5.7 compares the performance of constitutive and stochastic-myopic killers. Unlike in the deterministic/periodic case

(Fig 5.5(a)), here the advantage of population-sensing (myopic) killers is significant: they have noticeably better chances of winning than constitutives starting from most initial conditions. Another simple comparison is to focus on the previous boundaries between the blue (deterministic defeat) and red (deterministic winning) in Figs 5.4(c) and 5.5(b). Plotting these boundaries as black dashed lines in Figs 5.7(a) and 5.7(b) respectively, we provide a different quantitative measure of stochastic myopic killers' advantage: their average chances for success starting near this deterministic "no microbe's land" are $\approx 72\%$, while for the constitutives the same number is only $\approx 46\%$. Interestingly, the stochastic myopic optimal policy $\alpha_1(x, y)$ shown in Fig 5.7(c) is quite close to the zeroth time-slice of the deterministic myopic optimal policy $a_*(x, y, 0)$ from Fig 5.5(c). The difference in performance comes primarily from the fact that $\alpha_1(x, y)$ is stationary and that occasional long intervals ($\mathcal{T} > 1/\lambda$) really help the myopic killers. Nevertheless, the toxin-sensitives still have a significant probability of winning on a large set of initial conditions, particularly when the toxin-producers are not starting in the majority.

To check whether the randomness in dilution times has a similar impact in other stochastic environments, we introduce a slightly different metric of competitive advantage and use it across a range of (ρ, λ) values. Assuming that the initial population size $N(0) = y$ is fixed while $f(0) \in [0, 1]$ is selected uniformly at random, we examine the probability of killers' winning. For regular/deterministic dilutions with period $T = 1/\lambda$, this probability is simply the width of the killer's "deterministically-winning" (red) region at y, denoted by $\mathcal{L}(y)$. With random dilution times, this probability is $\mathcal{P}(y) = \int_0^1 \hat{w}^{\alpha}\Big(x, y\Big)\, \mathrm{d}x$. To quantify the impact of randomness, we define $\bar{Q}(y) = \mathcal{L}(y) - \mathcal{P}(y)$. Focusing on $y = 0.5$, Fig 5.8 presents a heat map of $\bar{Q}(0.5)$ for $(\rho, \lambda) \in [0.52, 0.6] \times [0.75, 1.0]$. Subfigure (a) shows that, for constitutive killers, the randomness is beneficial in the upper left half of the parameter space (where the dilutions are more severe and frequent) but actually slightly detrimental in the bottom right half (where the dilutions are more moderate and rare). In contrast, for

the population-sensing (myopic, optimized for a specific T or λ) killers, the randomness in dilution times appears to be beneficial across all (ρ, λ).

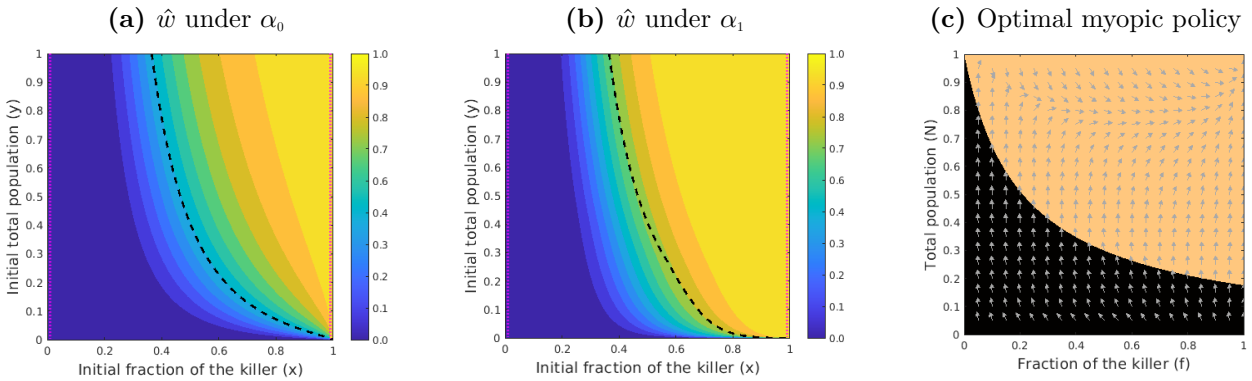


Figure 5.7. Performance of (a) constitutive and (b) "stochastic myopic" killers under random dilutions ($\lambda = 1$). The probability of attaining competitive exclusion is noticeably higher for the population-sensing ("stochastic myopic") toxin-producers starting from most initial conditions. Dashed black lines show the boundary of the set from which they could deterministically win under periodic dilutions with $T = 1$. In the current random dilutions setting, starting near that boundary gives the stochastic-myopic killers a $\approx 72\%$ chance of winning, while the same number for constitutives is only $\approx 46\%$. This is mainly because the "stochastic-myopic" killers opt not to produce the toxin when their fraction or the overall population size is low (subfigure (c)). In both (a) & (b), the victory and defeat barriers (γ_v and γ_d, respectively) are indicated by vertical magenta dotted lines. All parameter values are the same as in Fig 5.4(c).

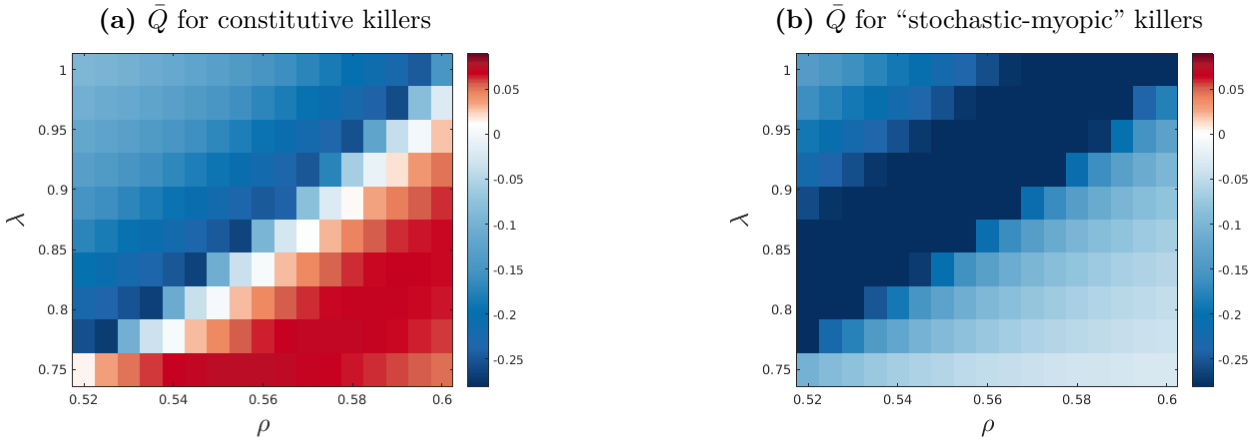


Figure 5.8. The impact of dilution-time randomness on the performance of constitutive and myopic killers across a range of (ρ, λ) values. The performance metric $\bar{Q}$ (defined in the text) is shown in red wherever the toxin-producers have better chances of winning with regular/periodic dilutions and in blue wherever their chances are better with randomly-timed dilutions (assuming the same average frequency: $\lambda = 1/T$). For constitutives, the randomness is beneficial when dilutions are more severe and frequent, but it is sightly detrimental when dilutions happen more rarely and are less drastic. For myopic killers (with policies optimized for each λ and $T = 1/\lambda$), the randomness seems beneficial across all tested parameters.

5.2.4 Can toxin-producers do better if they are non-myopic?

The optimality of toxin-production policy α_1 is myopic because it is selected with only one (upcoming) dilution in mind, ignoring the ultimate goal of killers winning after arbitrarily many dilutions. It is natural to ask whether they would gain a substantial advantage by selecting a policy which maximizes the probability of attaining their victory before the sensitives, $\hat{w}^\alpha$. For fixed values of dilution frequency λ and survival factor ρ, such "ultimately smart" policy $\alpha_\infty(x, y)$ can be found numerically by solving an HJB-type equation with non-local coupling; see (5.14) (Box 11).

Fig 5.9 shows that this α_∞ prescribes producing toxin slightly more conservatively than the myopic α_1, but in the end its performance is only marginally better for our chosen parameter values. However, both of these population-sensing-enabled policies have a very significant advantage over the constitutive $\alpha_0 = 1$; see Fig 5.9(c).

These observations appear to be robust, holding true for a variety of stochastic environments. In Fig 5.10, we focus on a single initial condition $(f(0), N(0)) = (0.5, 0.1)$ and compare the performance of these toxin-production policies for a range of (ρ, λ) values. Predictably, all three of them yield higher chances of winning against toxin-sensitives when the dilutions are rare and weak (small λ, large ρ) – in these regimes, the population gets closer to the carrying capacity in between dilutions, the growth of both strains slows down, and the toxin's effect becomes more noticeable. As expected, the constitutives are far less effective on most of this map. The biggest surprise is how well the myopic killers do – their chances of winning are in the worst case only 2.5% below those of "ultimately smart" killers. This is impressive since the myopic α_1 is formulated without any reference to ρ, but works well across a fairly broad range of dilution strengths. Even against the ultimately optimized α_∞, the toxin-sensitives still have a chance of winning above 50% on at least half of this (ρ, λ)

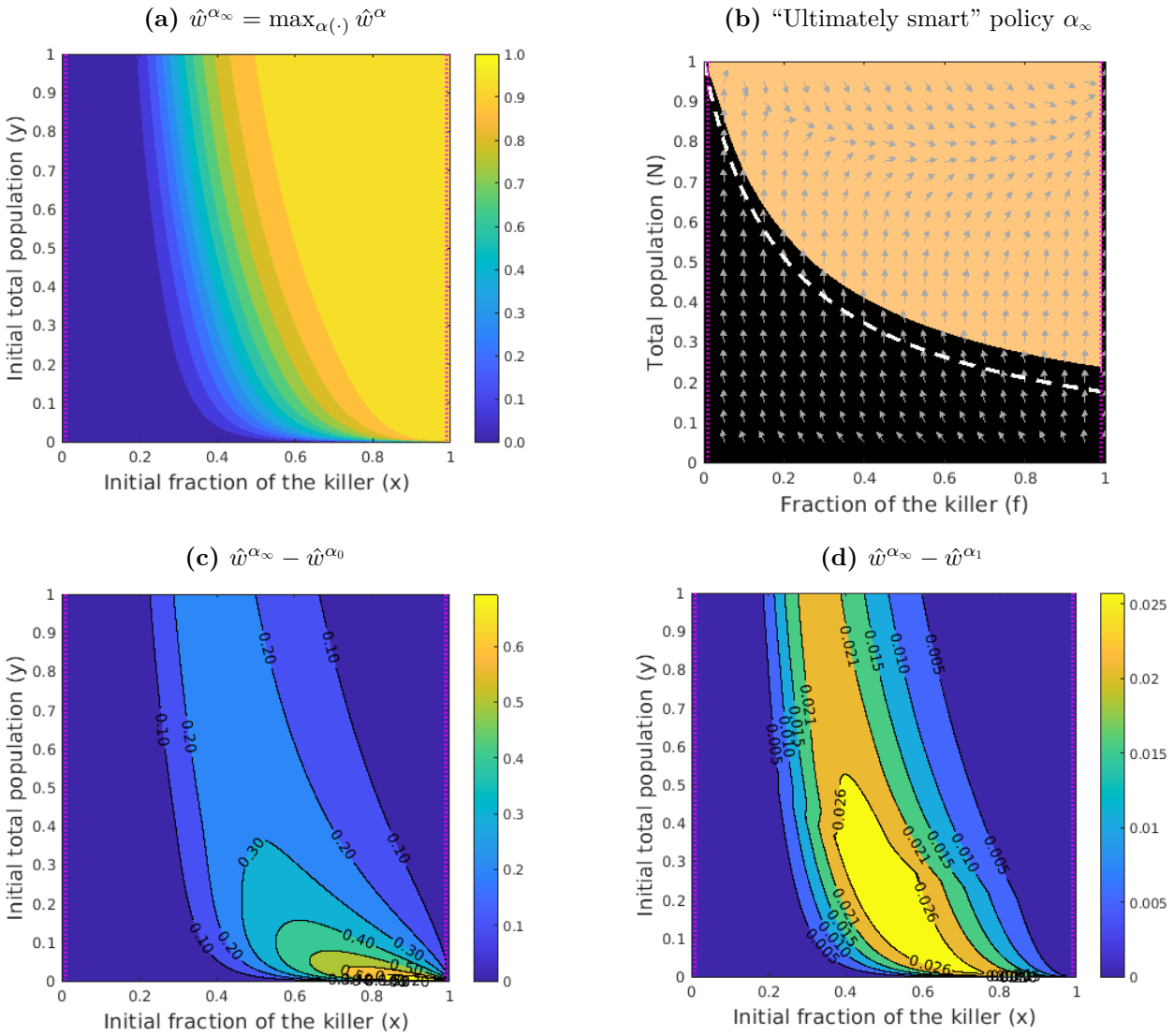


Figure 5.9. "Ultimately smart" killers: performance, policy, and comparison with constitutive and myopic killers. The optimal toxin-on region for the "ultimately smart" killers (orange in subfigure (b)) slightly shrinks compared to that of the "stochastic-myopic" killers (the boundary of which is shown by a white-dashed line). As a result, the maximized probability of winning ($\hat{w}^{\alpha_\infty}$ in subfigure (a)) is only marginally better than $\hat{w}^{\alpha_1}$, with the maximum difference of just 0.025 (see the absolute difference map in subfigure (d)). However, compared to constitutive killers , the advantage is significant: on a large part of the domain, the improvement in chances of winning is above 20% (subfigure (c)). For really small N and relatively large f, this advantage is even above 60% – this is the set of initial conditions where constitutives grow much slower and are thus more affected by occasional short inter-dilutions intervals. In all subfigures, the victory and defeat barriers (γ_{v} and γ_{d}, respectively) are plotted with a magenta dotted line In both (c) & (d), the contour lines are labeled with their respective probability values.

map.

Therefore, whether dilutions are randomly timed or periodic and despite the benefits brought by toxin-production regulation to killers, the sensitive strain can prevail by taking advantage of disruptions caused by dilutions.

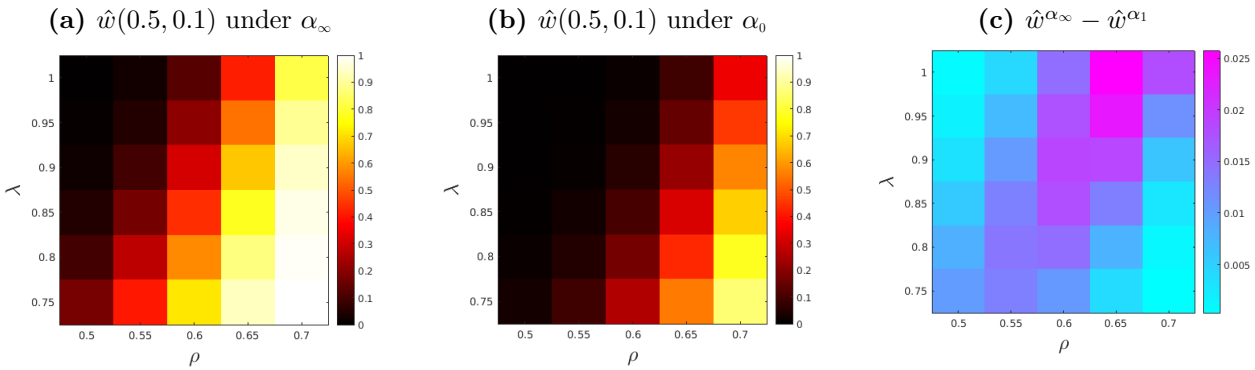


Figure 5.10. Comparison of probabilistic performance for different types of killers starting from $(f(0), N(0)) = (0.5, 0.1)$ for a range of dilution strengths and frequencies. Policy α_1 is recomputed for each λ, while policy α_∞ is recomputed for each (ρ, λ) combination. In general, a stronger survival rate (larger ρ) combined with a slower arrival rate (smaller λ) increases the chances of toxin-producers to win for all three policies. It is clear that the ultimately smart (subfigure (a)) and stochastic-myopic killers significantly outperform the constitutives (subfigure (b)). The differences in $\hat{w}(0.5, 0.1)$ between α_∞ and α_1 are still small, with the discrepancy increasing toward the upper right corner (subfigure (c)).

Box 11: Different types of toxin-production policies

1. Constitutive killers: always producing the toxin (i.e., $\alpha_0 = 1$).

2. Stochastic-myopic killers: The value function

$$v(x,y) = \sup_{\alpha(\cdot)} \mathbb{E}\big[f\big(\mathcal{T}^-;\alpha(\cdot)\big) \mid f(0)=x,\ N(0)=y\big]. \tag{5.10}$$

The myopic-optimal policy α_1 can be found by solving

$$\lambda\big[x - v(x,y)\big] + \max_{a\in\{0,1\}} \left\{ \nabla v(x,y) \cdot \begin{bmatrix} F(x,y,a) \\ G(x,y,a) \end{bmatrix} \right\} = 0, \tag{5.11}$$

with the boundary condition

$$v(x,y) = \begin{cases} 1, & \text{if } x = 1 \text{ and } y \neq 0 \\ 0, & \text{if } y = 0 \text{ or } x = 0. \end{cases} \tag{5.12}$$

See SI Appendix §5.4.3 for the derivation.

3. Ultimately smart killers: The value function

$$w(x,y) = \sup_{\alpha(\cdot)} \mathbb{P}\Big(T_v\big(x,y,\alpha(\cdot)\big) < T_d\big(x,y,\alpha(\cdot)\big)\Big). \tag{5.13}$$

The "ultimately smart" policy α_∞ can be found by solving a first-order non-local HJB equation satisfied by w:

$$0 = \lambda\big[w(x,\rho y) - w(x,y)\big] + \max_{a\in\{0,1\}} \left\{ \nabla w(x,y) \cdot \begin{bmatrix} F(x,y,a) \\ G(x,y,a) \end{bmatrix} \right\}, \tag{5.14}$$

with the boundary condition

$$w(x,y) = \begin{cases} 1, & \text{if } x > \gamma_{\mathrm{v}} \text{ and } y \neq 0, \\ 0, & \text{if } x < \gamma_{\mathrm{d}} \text{ or } y = 0. \end{cases} \tag{5.15}$$

See SI Appendix §5.4.3 for the derivation and §5.4.4 for the numerics.

5.3 Discussion

In this chapter, we have shown both theoretically and experimentally that dilution events can benefit toxin-sensitive strains, when their rate of growth in the exponential phase is larger than that of toxin-producing ones. Using high-throughput experiments with strains of *S. cerevisiae*, one engineered to constitutively produce the killer toxin K1, and another

one sensitive to it, we found that the outcome and dynamics of competition between the two varied with the frequency of periodic dilution events. Because toxin production is often regulated in response to quorum sensing, we used tools of (stochastic) optimal control theory to explore how regulating toxin production in response to population-sensing can benefit toxin-producing strains, introducing several types of toxin-regulating strategies that are designed to receive information from various sources. Along this approach, we developed two effective algorithms that (i) calculate the deterministic limit of the relative abundances of these strains as the number of periodic dilutions approaches infinity; and (ii) address the non-local Hamilton-Jacobi type equation that includes the discontinuities introduced by dilution events in a stochastic environment. Our numerical experiments consistently support the conclusion that, regardless of toxin production regulation, dilution events disrupt the dominance of the killer, thereby protecting the sensitive strains from extinction.

Rather than focusing on specific mechanistic models of toxin production regulated by quorum sensing, which would vary across species and would require a large number of parameters and modeling assumptions, we adopted a phenomenological approach to identify theoretical performance bounds for "omniscient killers" capable of measuring population density and fractions. The optimal policies derived here should thus be regarded as upper bounds on the ability of toxin-producing strains to out-compete sensitive ones in fluctuating environments. Future work will explore how more realistic, mechanistic models of toxin production regulated by quorum sensing compare to the optimal policies investigated here. It is also of interest to ask what the optimal policy would be, in different environments, if the killer strain could only sense its own population density or the concentration of a quorum sensing molecule, which would increase the dimensionality of the problem and necessitate modeling metabolite concentration dynamics. These extensions will also require handling a partially observed system [17], for which substantial mathematical and numerical challenges are inevitable [92, 176].

In a seminal paper on allelopathy in spatially distributed populations [48], Durrett and Levin showed that the competition of toxin-producing and toxin-sensitive strains displays bi-stability in well-mixed, undisturbed competitions, with either the killer or sensitive strain dominating in the long-term limit depending on the initial condition. In their model, such bi-stability depends critically on the magnitude of a death rate term in their governing equations. In the SI Appendix, we show that using typical values for bacterial (maximal) growth and death rates (the former being typically much larger than the latter), the range of initial conditions for which the sensitive strain competitively exclude the killer is extremely small, and is thus unlikely to explain why sensitive strains are found in many natural populations. It is of interest to ask how the perturbations investigated here would affect the spatial competition between killer and sensitive cells explored theoretically in [48], where the two strain types were shown to coexist by forming dynamic, single-strain clusters. Similarly, it would be useful to explore the impact of dilutions on the nucleation criteria that control the killers' invasion success of a spatially-distributed, resident sensitive population [72].

Given the many simplifications we have made in our analysis, several extensions will obviously enrich this approach in the future. One significant improvement would involve considering random outcomes of dilution events rather than assuming that the relative abundances are preserved and a fixed fraction ρ of the population consistently survives. For example, modeling the outcomes with a Binomial distribution with a surviving rate ρ could provide a more nuanced understanding. While preliminary Monte Carlo simulations in SI Appendix §5.4.6 suggest that the results remain qualitatively the same, randomness of the dilution factor in the experiment is likely responsible for the spread in the outcome of competition of our experiments at the initial fraction of 45%, where we observed some replicates being dominated by the toxin-producing strain and others by the toxin-sensitive one (Fig 5.2B). Incorporating randomness of dilution factors in our mathematical models will likely lead to more sophisticated equations and the challenge of devising an efficient algorithm to solve

them.

Another limitation of our approach is our reliance on a deterministic model of population growth, which may be subject to the "atto-fox" problem [64] by which either strain may recover from extremely small population sizes. This issue can be alleviated by introducing a demographic noise term in the governing equations, which may lead either strain or even the entire population to extinction, depending on their absolute size. Moreover, a more realistic model would consider random parameter values, particularly for the growth and killing rates $(r_{\mathrm{K}}, r_{\mathrm{S}})$ and γ, as well as the dilution strength and frequency ρ and λ. In real-world scenarios, these parameters are often subject to variability due to environmental fluctuations, biological heterogeneity, and other stochastic factors. Such fluctuations in biological systems would more accurately capture the behavior and outcomes of population dynamics. These extensions lead to a hybrid model combining discrete and continuous random perturbations in the framework of general jump-diffusion processes [127, 151] and leading to additional computational challenges.

Incorporating evolutionary adaptation such as mutational dynamics, would greatly enhance our understanding of the competition between toxin-producing and toxin-sensitive strains in natural environments over longer timescales. Experiments indicate that killer strains can lose or alter their toxin-producing ability, and sensitive strains may develop resistance to the toxin [133, 28, 72]. The ultimate success of the killer strain in our model is negatively impacted by the costs associated with toxin production, suggesting that evolutionary adaptation may aim to minimize these costs [133]. Additionally, evolutionary adaptation may enable toxin-producing strains to regulate toxin production in response to their environment, population abundance, or the presence of competitors [124], possibly approaching the performance of the optimal policies described here. Toxin resistance can arise through various mechanisms, such as alterations in toxin receptors or translocation pathways, which

may have antagonistic pleiotropic effects where resistance incurs a cost in terms of growth rate [57]. This growth rate penalty will, in turn, influence the competitive dynamics with the killer strain. Finally, in environments experiencing disturbances, evolutionary adaptation may promote increased retention (ρ) [87] or even alter the rate of disturbances (λ). For example, production of surface-attachment molecules, pili or fimbriae by microbes such as *Pseudomonas aeruginosa, Vibrio cholerae, Clostridium difficile* and *Streptococcus salivarius* can help them adhere to surfaces in their environment and prevent them from being washed away in fluid environments such as the gastrointestinal tract, the oral cavity, or natural water bodies [131, 144, 162]. *C. difficile* and *V. cholerae*, in addition to adhering to surfaces such as the intestinal mucosa, can also cause diarrhea and thus potentially control the environment dilution rate and intensity, at least transiently.

In conclusion, we propose that the fitness cost incurred by toxin-producing strains to engage in antagonistic behavior may be detrimental in boom-and-bust environments in which populations undergo regular or stochastic dilutions, possibly explaining why both antagonistic and non-antagonistic microbes are found in nature, and why environments with higher turnover rates may favor the latter [11, 105].

5.4 Supporting Information (SI) Appendix

5.4.1 Comparison with the Durrett-Levin model

In the main text, all of our results are based on a basic competition model between the toxin-producing "killer" strain (K) and the toxin-sensitive strain (S) ((5.1) of the main text). This model assumes the growth of both strains is logistic with respective intrinsic growth rates $(r_{\mathrm{K}}, r_{\mathrm{S}})$ and a shared carrying capacity C. The toxin-induced death rate of the sensitive cells is proportional to the product of the strain densities, $n_{\mathrm{K}} n_{\mathrm{S}}$. Focusing on *constitutive killers* who produce the toxin at the maximal rate $a = 1$, the original (5.1) reads

$$\begin{cases} \dfrac{dn_{\mathrm{K}}}{dt}(t) = r_{\mathrm{KS}}(1-\varepsilon)\left(1 - n_{\mathrm{K}} - n_{\mathrm{S}}\right) n_{\mathrm{K}}, \\ \dfrac{dn_{\mathrm{S}}}{dt}(t) = \left(1 - n_{\mathrm{K}} - n_{\mathrm{S}}\right) n_{\mathrm{S}} - \gamma n_{\mathrm{K}} n_{\mathrm{S}}, \end{cases} \tag{5.16}$$

where $r_{\mathrm{KS}} = r_{\mathrm{K}}/r_{\mathrm{S}} < 1$ is the intrinsic growth rates ratio, γ is the killing rate rescaled by r_{S}, and ε is the cost associated with producing toxin at the maximal rate ($a = 1$). We note three fixed points of (5.16):

1. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0, 0)$ – a nodal source
2. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (1, 0)$ – a nodal sink
3. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0, 1)$ – a degenerate node

It follows that starting from any point except for the fixed ones, a competitive exclusion of the sensitive by the killer will be observed.

Durrett and Levin [48] in 1997 proposed a similar competition model between a colicin-producing "killer" strain and a colicin-sensitive strain. Their model is a direct extension of

ours, incorporating natural death rates for both strains. By including these natural death terms $(\delta_{\mathrm{K}}, \delta_{\mathrm{S}})$ (rescaled by r_{S}^{-1}) in our (5.16), our model can be extended as follows:

$$\begin{cases} \dfrac{dn_{\mathrm{K}}}{dt}(t) = r_{\mathrm{KS}}(1-\varepsilon)\,(1-n_{\mathrm{K}}-n_{\mathrm{S}})\,n_{\mathrm{K}} - \delta_{\mathrm{K}} n_{\mathrm{K}}, \\ \dfrac{dn_{\mathrm{S}}}{dt}(t) = (1-n_{\mathrm{K}}-n_{\mathrm{S}})\,n_{\mathrm{S}} - \gamma n_{\mathrm{K}} n_{\mathrm{S}} - \delta_{\mathrm{S}} n_{\mathrm{S}}. \end{cases} \tag{5.17}$$

By setting both equations to zero, we find four fixed points of this system:

1. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0, 0)$
2. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = \left(1 - \dfrac{\delta_{\mathrm{K}}}{r_{\mathrm{KS}}(1-\varepsilon)},\ 0\right)$
3. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0,\ 1-\delta_{\mathrm{S}})$
4. $(n_{\mathrm{K}}, n_{\mathrm{S}}) = \left(\dfrac{\delta_{\mathrm{K}}}{\gamma r_{\mathrm{KS}}(1-\varepsilon)} - \dfrac{\delta_{\mathrm{S}}}{\gamma},\ \dfrac{\delta_{\mathrm{S}}+\gamma}{\gamma} - \dfrac{\delta_{\mathrm{K}}(1+\gamma)}{\gamma r_{\mathrm{KS}}(1-\varepsilon)}\right)$

Stability analysis shows that the origin is a *nodal source*, the two boundary equilibria are both *nodal sinks*, and the interior fixed point is *hyperbolic saddle* if

$$\delta_{\mathrm{K}} < r_{\mathrm{KS}}(1-\varepsilon), \quad \delta_{\mathrm{S}} < 1, \quad \delta_{\mathrm{S}} < \frac{\delta_{\mathrm{K}}}{r_{\mathrm{KS}}(1-\varepsilon)} < \frac{\delta_{\mathrm{S}}+\gamma}{1+\gamma}. \tag{5.18}$$

The system thus exhibits a "bi-stability": starting above the stable manifold (e.g., magenta dotted-dashed line in Fig 5.11(a)) of the hyperbolic saddle leads to competitive exclusion of the killer by the sensitive, while starting below it results in competitive exclusion of the sensitive by the killer. See Fig 5.11(a) for a detailed phase portrait.

It is worth noting that when the natural death rates approach zero (i.e., as δ_{S}, $\delta_{\mathrm{K}} \to 0$), (5.17) reduces to (5.16). In the meantime,

$$\begin{aligned} 1 - \frac{\delta_{\mathrm{K}}}{r_{\mathrm{KS}}(1-\varepsilon)} &\to 1, \\ 1 - \delta_{\mathrm{S}} &\to 1, \\ \frac{\delta_{\mathrm{K}}}{\gamma r_{\mathrm{KS}}(1-\varepsilon)} - \frac{\delta_{\mathrm{S}}}{\gamma} &\to 0, \\ \frac{\gamma+\delta_{\mathrm{S}}}{\gamma} - \frac{\delta_{\mathrm{K}}(\gamma+1)}{\gamma r_{\mathrm{KS}}(1-\varepsilon)} &\to 1. \end{aligned}$$

Consequently, the last two equilibria collapse to $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0, 1)$ while the second fixed point moves to $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (1, 0)$. This means the system loses its "bi-stability" structure as the "sensitive-winning" region shrinks when the natural death rates approach zero (see the transitions in Fig 5.11). Schink et al. [146] estimated the death rate of an *E. coli* strain K-12 at 0.018 h^{-1} in laboratory experiments, whereas the max growth rate was 0.7 h^{-1}. In our time scale, it results in $\delta_{\mathrm{K}} = \delta_{\mathrm{S}} \approx 0.0257$. We see from Fig 5.11(c) that in this scenario, the hyperbolic saddle is extremely close to the "all sensitives" equilibrium and both of them are close to $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (0, 1)$. This extremely small basin of attraction for the "all sensitives" equilibrium makes it harder to explain why sensitive strains are often found in the natural environment. This issue is one of the motivations for the current paper, and for our conjecture that, with dilutions, the sensitives can win even in the $\delta_{\mathrm{K}} = \delta_{\mathrm{S}} = 0$ limit, where the bi-stability disappears; Fig 5.11(d).

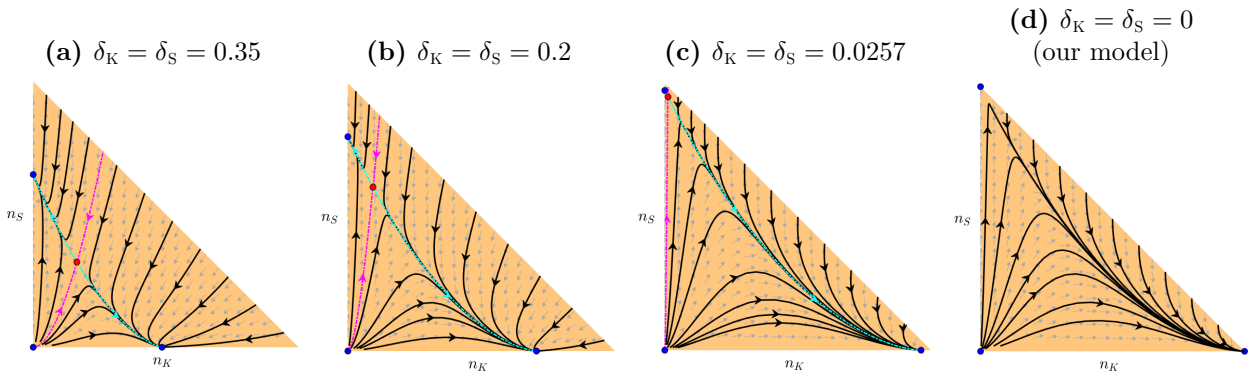


Figure 5.11. Phase portraits of the Durrett-Levin model (Eq. 5.17) with decreasing death rates. The "bi-stability" is noticeable when the death rates $(\delta_{\mathrm{K}}, \delta_{\mathrm{S}})$ are comparable in magnitude to the growth rates (subfigures (a,b)). Using laboratory-estimated death rates values [146], the hyperbolic saddle moves toward the "all sensitives" stable node, with both converging toward $(n_{\mathrm{K}}, n_{\mathrm{S}}) = (1, 0)$ (subfigure (c)). When $(\delta_{\mathrm{K}}, \delta_{\mathrm{S}}) = (0, 0)$, the Durrett-Levin model reduces to our model (5.16), and "bi-stability" disappears (subfigure (d)). In (a-c), the hyperbolic saddle is plotted with a red dot while other equilibria are plotted with a blue dot. The stable manifold is plotted with a magenta dotted-dashed line while the unstable manifold is plotted with a cyan dotted-dashed line. In (d), all equilibria are plotted with a blue dot. Parameter values: $r_{\mathrm{KS}} = 0.85$, $\varepsilon = 0.2$, and $\gamma = 1$.

5.4.2 Effect of dilutions on a single-strain logistic growth model

This section presents theoretical and numerical results relevant for a *single* strain subjected to either regular or randomly timed dilutions. These findings provide background information for the main text and are relevant after one of the strains becomes strongly dominant.

Effect of regular dilutions

In this subsection, we prove the theoretical pre-dilution population limit for a population growing according to the logistic model and undergoing regular dilutions.

Theorem 5.4.1. *Consider the following rescaled logistic growth model*

$$\dot{q} = rq(1-q), \qquad q(0) = x \in (0,1]. \tag{5.19}$$

If the system undergoes regular dilutions with a fundamental period of T, and after each dilution, a deterministic fraction ρ of the population survives, i.e.,

$$q\Big((mT)^+\Big) = \rho q\Big((mT)^-\Big), \quad \textit{for } m = 1, 2, 3, \ldots, \tag{5.20}$$

Then

$$\lim_{m\to\infty} q\Big((mT)^-\Big) = \begin{cases} \dfrac{1}{\rho + \dfrac{1-\rho}{1-\exp\Big(-(rT+\ln\rho)\Big)}}, & \textit{if } rT + \ln\rho > 0, \\ 0, & \textit{otherwise.} \end{cases} \tag{5.21}$$

Proof. We begin the proof of Theorem 5.4.1 by proving the following lemma.

Lemma 5.4.2. *Given the rescaled logistic model* (5.19)*, the pre-dilution population size up to the m-th cycle is*

$$\begin{cases} q\Big((mT)^-\Big) = \dfrac{1}{\rho + (1-\rho)\sum\limits_{k=0}^{m-1} \exp\Big(-(rT+\ln\rho)k\Big) + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)(m-1)-rT\Big)}, \\ q(0) = x \in (0,1]. \end{cases} \tag{5.22}$$

Proof. We prove the above lemma by mathematical induction. Starting with the first dilution $m = 1$, from (5.22) we have

$$q\Big(T^-\Big) = \frac{1}{\rho + (1-\rho) + \dfrac{1-x}{x}\exp(-rT)} = \frac{1}{1 + \dfrac{1-x}{x}\exp(-rT)}, \quad q(0) = x \in (0,1]. \tag{5.23}$$

We show it is true by solving (5.19) analytically. By separation of variables, one can show that the general solution to (5.19) is

$$q(t) = \frac{1}{1 + C\mathrm{e}^{-rt}}. \tag{5.24}$$

With $q(0) = x$, we have $C = \frac{1-x}{x}$. Consequently,

$$q\Big(T^-\Big) = \frac{1}{1 + \dfrac{1-x}{x}\exp(-rT)}, \quad q(0) = x.$$

Now assume Eq. (5.22) holds true for some integer $j > 1$. I.e.,

$$q\Big((jT)^-\Big) = \frac{1}{\rho + (1-\rho)\sum\limits_{k=0}^{j-1} \exp\Big(-(rT+\ln\rho)k\Big) + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)},$$

with $q(0) = x$. We now show it holds for $j+1$. Notice that $q\Big([(j+1)T]^-\Big)$ with $q(0) = x$ is equivalent to $q\Big(T^-\Big)$ with $q(0) = q\Big((jT)^+\Big) = \rho q\Big((jT)^-\Big)$ under the assumption of proportional dilutions. Given the general solution to (5.19) in (5.24), we compute the arbitrary

constant C by imposing the new initial condition $q(0) = \rho q\Big((jT)^-\Big)$. It follows that

$$\begin{aligned}
\frac{1}{1+C} = q(0) &= \frac{\rho}{\rho + (1-\rho)\sum\limits_{k=0}^{j-1} \exp\Big(-(rT+\ln\rho)k\Big) + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)} \\
&= \frac{1}{1+(1-\rho)\sum\limits_{k=0}^{j-1} \dfrac{\exp\Big(-(rT+\ln\rho)k\Big)}{\rho} + \dfrac{1-x}{x}\dfrac{\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)}{\rho}} \\
&= \frac{1}{1+(1-\rho)\sum\limits_{k=0}^{j-1} \dfrac{\exp\Big(-(rT+\ln\rho)k\Big)}{\exp(\ln\rho)} + \dfrac{1-x}{x}\dfrac{\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)}{\exp(\ln\rho)}} \\
C &= (1-\rho)\sum_{k=0}^{j-1} \frac{\exp\Big(-(rT+\ln\rho)k\Big)}{\exp(\ln\rho)} + \frac{1-x}{x}\frac{\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)}{\exp(\ln\rho)}.
\end{aligned}$$

Substituting it back into (5.24), we have

$$\begin{aligned}
q(T^-) &= \frac{1}{1+\left[(1-\rho)\sum\limits_{k=0}^{j-1} \dfrac{\exp\Big(-(rT+\ln\rho)k\Big)}{\exp(\ln\rho)} + \dfrac{1-x}{x}\dfrac{\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)}{\exp(\ln\rho)}\right]e^{-rT}} \\
&= \frac{1}{1+(1-\rho)\sum\limits_{k=0}^{j-1} \dfrac{\exp\Big(-(rT+\ln\rho)k\Big)}{\exp(rT+\ln\rho)} + \dfrac{1-x}{x}\dfrac{\exp\Big(-(rT+\ln\rho)(j-1)-rT\Big)}{\exp(rT+\ln\rho)}} \\
&= \frac{1}{1+(1-\rho)\sum\limits_{k=0}^{j-1} \exp\Big(-(rT+\ln\rho)(k+1)\Big) + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)j-rT\Big)} \\
&= \frac{1}{[\rho+(1-\rho)]+(1-\rho)\sum\limits_{k=0}^{j-1} \exp\Big(-(rT+\ln\rho)(k+1)\Big) + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)j-rT\Big)} \\
&= \frac{1}{\rho+(1-\rho)\left[1+\sum\limits_{k=0}^{j-1} \exp\Big(-(rT+\ln\rho)(k+1)\Big)\right] + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)j-rT\Big)} \\
&= \frac{1}{\rho+(1-\rho)\sum\limits_{k=0}^{j} \exp\Big(-(rT+\ln\rho)k\Big) + \dfrac{1-x}{x}\exp\Big(-(rT+\ln\rho)j-rT\Big)}, \\
&= q\Big([(j+1)T]^-\Big), \quad q(0) = x.
\end{aligned}$$

Consequently, (5.4.2) holds for all $j \in \mathbb{N}$. Q.E.D.

Now, we take the limit of (5.4.2) as m approaches infinity. Notice that the second term in the denominator is a geometric series. It follows that

$$\lim_{m\to\infty}(1-\rho)\sum_{k=0}^{m-1}\exp\Big(-(rT+\ln\rho)k\Big)=\begin{cases}\dfrac{1-\rho}{1-\exp\Big(-(rT+\ln\rho)\Big)}, & \text{if } \rho>\mathrm{e}^{-rT},\\ +\infty, & \text{otherwise.}\end{cases} \tag{5.25}$$

With the third term in the denominator of (5.22) approaching 0 as $m\to\infty$, we thus conclude that

$$\lim_{m\to\infty} q\Big((mT)^-\Big)=\begin{cases}\dfrac{1}{\rho+\dfrac{1-\rho}{1-\exp\Big(-(rT+\ln\rho)\Big)}}, & \text{if } \rho>\mathrm{e}^{-rT},\\ 0, & \text{otherwise.}\end{cases}$$

Q.E.D.

Therefore, for a population consisting of sensitives only, i.e.,

$$\dot{n_{\mathrm{S}}}=n_{\mathrm{S}}(1-n_{\mathrm{S}}),$$

the limiting pre-dilution population size is

$$\lim_{m\to\infty} n_{\mathrm{S}}\Big((mT)^-\Big)=\begin{cases}\dfrac{1}{\rho+\dfrac{1-\rho}{1-\exp\Big(-(T+\ln\rho)\Big)}}, & \text{if } \rho>\mathrm{e}^{-T},\\ 0, & \text{otherwise.}\end{cases} \tag{5.26}$$

For a population consisting of constitutive killers only, i.e.,

$$\dot{n_{\mathrm{K}}}=r_{\mathrm{KS}}(1-\varepsilon)n_{\mathrm{K}}(1-n_{\mathrm{K}}),$$

the limiting pre-dilution population size is

$$\lim_{m\to\infty} n_{\mathrm{K}}\Big((mT)^-\Big)=\begin{cases}\dfrac{1}{\rho+\dfrac{1-\rho}{1-\exp\Big(-(r_{\mathrm{KS}}(1-\varepsilon)T+\ln\rho)\Big)}}, & \text{if } \rho>\mathrm{e}^{-r_{\mathrm{KS}}(1-\varepsilon)T},\\ 0, & \text{otherwise.}\end{cases} \tag{5.27}$$

For a population consisting of bacteria capable of producing toxin but "choosing" not to produce it, i.e.,

$$\dot{n_{\mathrm{K}}} = r_{\mathrm{KS}} n_{\mathrm{K}} (1 - n_{\mathrm{K}}),$$

the limiting pre-dilution population size is

$$\lim_{m\to\infty} n_{\mathrm{K}}\Big((mT)^{-}\Big) = \begin{cases} \dfrac{1}{\rho + \dfrac{1-\rho}{1 - \exp\Big(-(r_{\mathrm{KS}}T + \ln\rho)\Big)}}, & \text{if } \rho > \mathrm{e}^{-r_{\mathrm{KS}}T}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.28}$$

This formula might be relevant for the limiting population size in the "sensitives vs population-sensing killers" competition if the killers win in the limit and approach that limit through a toxin-off part of the (f, N) space. If that victory is approached through the toxin-on part of the (f, N) space, the relevant population limit is provided by Eq. (5.27). In the following figures, whenever we refer to population-sensing killers, we always use Eq. (5.28).

Fig 5.12 show the numerical values derived in (5.26), (5.27), and (5.28) in the (ρ, T) phase plane. It is evident that the larger the proportion of the population surviving after dilution (higher ρ) and the longer the period, the higher is the population size as the number of dilutions approaches infinity. For non-zero limits, sensitives can maintain a higher limiting population than any killers, due to their faster reproduction rate. Additionally, population-sensing killers outperform constitutive killers as they do not produce the toxin when sensitives are not nearby.

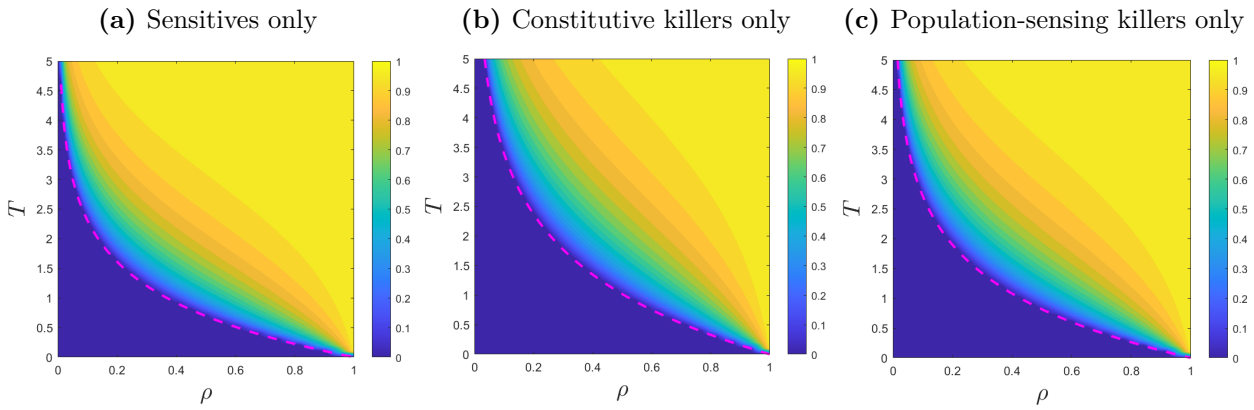


Figure 5.12. Pre-dilution single population limits under regular dilutions in the (ρ, T) phase plane. (a) sensitives only ($r = 1$); (b) constitutive killers only ($r = r_{\mathrm{KS}}(1 - \varepsilon) = 0.68$); (c) population-sensing killers only ($r = r_{\mathrm{KS}} = 0.85$). In all of them, the magenta-dashed line corresponds to $\rho = \exp(-rT)$ with their respective intrinsic growth rate r. limiting pre-dilution population is zero below this line.

Effect of randomly-timed dilutions

Next, we consider randomly timed dilutions, following a Poisson process with rate λ. In this case, any inter-dilution time $\mathcal{T}$ is exponentially distributed with the expected value $\mathbb{E}[\mathcal{T}] = 1/\lambda$. To compare with the previous results, we focus on the *expected pre-dilution population size* as the number of dilutions approaches infinity, with the population growing according to (5.19). We still assume that a fixed fraction ρ of the population survives after each dilution. As an analytical limit is unlikely to be obtained, we conducted Monte Carlo simulations with 200 dilutions starting from the initial population $q(0) = 0.5$. Fig 5.13 shows that the general trend remains unchanged: weaker dilution strength (higher ρ) and slower arrival rate (lower λ) result in higher expected pre-dilution population size in the limit. However, with randomly timed dilutions, the region where the averaged population is nearly 1 in the limit is significantly reduced in all three cases. Surprisingly, the previous "deterministic boundary" $\rho = \exp(-r/\lambda)$ (the magenta dashed line in Fig 5.13) still appears to accurately predict where the population goes extinct in the limit in the $(\rho, 1/\lambda)$ phase

plane.

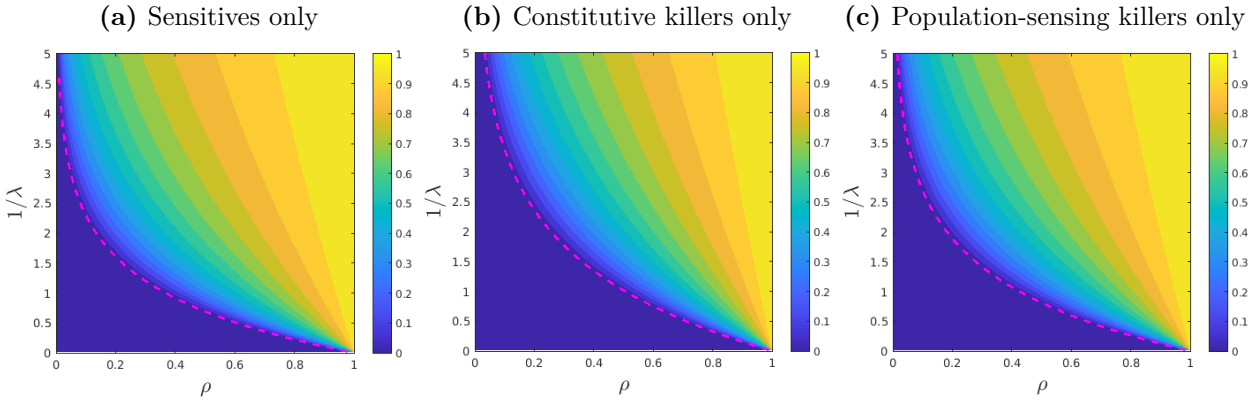


Figure 5.13. Randomly timed dilutions: mean empirical population just before the 201 dilution shown in the $(\rho, 1/\lambda)$ phase plane. (a) sensitives only ($r = 1$); (b) constitutive killers only ($r = r_{\mathrm{KS}}(1-\varepsilon) = 0.68$); (c) population-sensing killers only ($r = r_{\mathrm{KS}} = 0.85$). In all of them, the magenta-dashed line corresponds to $\rho = \exp(-r/\lambda)$ with their respective intrinsic growth rate r. The population below this line will most likely go extinct. All panels are produced by Monte Carlo simulations, conducted with 10^5 samples and a fixed initial population size 0.5.

The rest of this section follows closely the original derivation by my collaborator, Andrea Giometto. We further investigate properties of the distribution of the logistic growth model under our Poisson-distributed dilutions setting, specifically its behavior for population size $q = 1$ and $q = \rho$. The master equation [67] for a logistic growth model with instantaneous, Poisson-distributed dilutions at a rate λ, where the population $q(t)$ is reduced to a fraction ρ, is given by:

$$\frac{\partial p}{\partial t}(q,t) = -\frac{\partial}{\partial q}\left[q(1-q)p(q,t)\right] - \lambda p(q,t) + \frac{\lambda}{\rho} p\left(\frac{q}{\rho}, t\right), \tag{5.29}$$

where $-\frac{\partial}{\partial q}\left[q(1-q)p(q,t)\right]$ captures the deterministic logistic growth, $-\lambda p(q,t)$ represents the probability density leaving q to ρq due to dilutions occurring at rate λ, and $+\frac{\lambda}{\rho} p\left(q/\rho, t\right)$ reflects the probability density entering q from diluting q/ρ. The factor $1/\rho$ accounts for normalization. The stationary distribution of (5.29) is the solution of

$$-\frac{d}{dq}\left[q(1-q)p_s(q)\right] - \lambda p_s(q) + \frac{\lambda}{\rho} p_s\left(\frac{q}{\rho}\right) = 0. \tag{5.30}$$

Note that for $q > \rho$, the term $p_s(q/\rho)$ is identically zero if the population is initialized in $(0, 1]$. Thus, in this situation we have

$$-\frac{d}{dq}\left[q(1-q)p_s(q)\right] - \lambda p_s(q) = 0 \qquad \text{for } q > \rho. \tag{5.31}$$

Integrating in $(\rho, 1)$, we find

$$p_s(q) = p_s(\rho)\left(\frac{\rho}{q}\right)^{1+\lambda}\left(\frac{1-q}{1-\rho}\right)^{\lambda-1} \qquad \text{for } q > \rho. \tag{5.32}$$

It follows that

$$\lim_{q\to 1^-} p_s(q) = \begin{cases} 0, & \text{for } \lambda > 1, \\ \rho^2 p_s(\rho), & \text{for } \lambda = 1, \\ +\infty, & \text{for } \lambda < 1. \end{cases} \tag{5.33}$$

Moreover, taking $q \to \rho^+$ in the derivative of (5.32), we obtain

$$\lim_{q\to\rho^+} \frac{dp_s}{dq}(q) = -p_s(\rho)\frac{1+\lambda-2\rho}{\rho(1-\rho)} \tag{5.34}$$

which is negative if $2\rho < 1 + \lambda$.

The first derivative of p_s is zero at $q^\star = \frac{1+\lambda}{2}$, which is in $(\rho, 1)$ if $2\rho - 1 < \lambda < 1$. The second derivative there is always positive, leading to a local minimum.

It follows from (5.32) that in (ρ^2, ρ), the stationary distribution satisfies:

$$-\frac{d}{dq}\left[q(1-q)p_s(q)\right] - \lambda p_s(q) + \frac{\lambda}{\rho}p_s(\rho)\left(\frac{\rho^2}{q}\right)^{1+\lambda}\left(\frac{1-q/\rho}{1-\rho}\right)^{\lambda-1} = 0. \tag{5.35}$$

The solution to the homogeneous ODE corresponding to (5.35) is given by (5.32). Searching for a solution of the form $p_s(q) = p_h(q)p_{ih}(q)$ (for $\rho^2 < q < \rho$) we find that the inhomogeneous factor $p_{ih}(q)$ satisfies

$$q(1-q)\frac{dp_{ih}}{dq}(q) = \lambda\rho^\lambda\left(\frac{1-q/\rho}{1-q}\right)^{\lambda-1}, \tag{5.36}$$

whose solution is

$$p_{ih}(q) = 1 + \int_{\rho}^{q} \lambda\rho \frac{1}{s(1-s)} \left(\frac{\rho - s}{1-s}\right)^{\lambda-1} ds, \tag{5.37}$$

where $p_{ih}(\rho) = 1$ to ensure that $p_h(\rho)p_{ih}(\rho) = p_s(\rho)$. We thus obtain $p_s(q)$ in $\rho^2 < q < \rho$:

$$p_s(q) = p_s(\rho) \left(\frac{\rho}{q}\right)^{1+\lambda} \left(\frac{1-q}{1-\rho}\right)^{\lambda-1} \left[1 + \int_{\rho}^{q} \lambda\rho \frac{1}{s(1-s)} \left(\frac{\rho - s}{1-s}\right)^{\lambda-1} ds\right]. \tag{5.38}$$

Taking $q \to \rho^-$ in its derivative, we find

$$\lim_{q\to\rho^-} \frac{dp_s}{dq}(q) = \begin{cases} -p_s(\rho)\dfrac{1+\lambda-2\rho}{\rho(1-\rho)} = \lim\limits_{q\to\rho^+} \dfrac{dp_s}{dx}(q), & \text{for } \lambda > 1, \\ +\infty, & \text{for } \lambda < 1, \end{cases} \tag{5.39}$$

thus $q = \rho$ is a local maximum provided $\lambda < 1$ and $2\rho < 1 + \lambda$.

Remark 5.1: For $\lambda = 1$, expressions simplify significantly:

$$p_s(q) = p_s(\rho) \left(\frac{\rho}{q}\right)^2 \qquad \text{for } q > \rho \tag{5.40}$$

$$p_s(q) = \begin{cases} p_s(\rho) \left(\dfrac{\rho}{q}\right)^2, & \text{for } q > \rho, \\ p_s(\rho) \left(\dfrac{\rho}{q}\right)^2 \left[1 + \rho \ln\left(\dfrac{q}{\rho}\dfrac{1-\rho}{1-q}\right)\right], & \text{for } \rho^2 < q < \rho. \end{cases} \tag{5.41}$$

Fig 5.14 shows three empirical stationary distributions p_s computed by Monte Carlo simulations with $r = 1$ (the sensitives) and estimated as the empirical population size distribution at the end of the growth period following 200 dilutions. The value $p_s(\rho)$, which is needed to plot the analytical solutions, was taken from the empirical distribution. The histogram aligns well with our theoretical results. For $\rho^2 < q < \rho$, as λ increases, the PDF for p_s changes from increasing to decreasing. The PDF becomes unbounded as $q \to 1^-$ when $\lambda < 1$; in contrast, it is strictly decreasing on $(\rho, 1)$ when $\lambda \geq 1$.

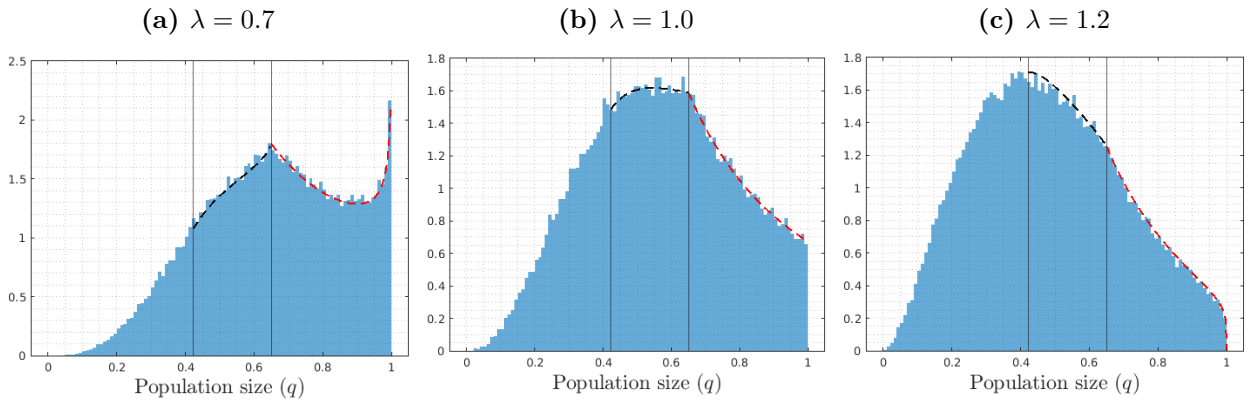


Figure 5.14. Representative empirical probability distributions of population abundance at the end of the growth period after 200 dilutions (normalized histograms) for $\lambda = 0.7$ (left), $\lambda = 1$ (middle), and $\lambda = 1.2$ (right). The two vertical lines mark ρ^2 and ρ. The red dashed curve is (5.32), and the black dashed curve is (5.38). All panels are produced with 10^5 samples, $r = 1$, and a fixed initial population 0.5. The value for $p_s(\rho)$ was taken from the empirical distribution.

5.4.3 Derivation of Hamilton-Jacobi-Bellman equations

Next, we derive the HJB equations for the "stochastic-myopic" killers ((5.11) in Box 11) and the "ultimately smart" killers ((5.14) in Box 11) via tools of dynamic programming. For the former, see also [4, 136] for more details. The derivation of the time-dependent HJB PDE for the finite-horizon problem ((5.3) in the main text) is omitted here, as it can be easily found in classical literature, such as [61].

Recall that we model the random dilution events as a Poisson process with a fixed rate $\lambda > 0$. Thus, any inter-arrival time $\mathcal{T}$ is exponentially distributed with rate λ. We assume that after each dilution, the fractions (relative abundances) are preserved while only ρ fraction of the total population survives.

For the "stochastic-myopic" killer

Recall the value function for the "stochastic-myopic" killer in Box 11 in the main text:

$$v(x,y) = \sup_{a(\cdot)} \mathbb{E}\Big[f\Big(\mathcal{T}^-\Big) \mid f(0) = x,\ N(0) = y,\ \text{following policy } a(\cdot)\Big].$$

Assume the optimal policy a_* exists, and let

$$\mathbb{E}^0 = \mathbb{E}[\,\cdot \mid f(0) = x,\ N(0) = y,\ \text{following } a_*].$$

Since $\mathcal{T} \sim \text{Exp}(\lambda)$, the expectation is defined as

$$v(x,y) = \int_0^\infty \lambda \mathrm{e}^{-\lambda t} f(t; a_*(t))\ \mathrm{d}t.$$

Rewriting it as

$$v(x,y) = \int_0^\infty \mathrm{e}^{-\lambda t}\Big[\lambda f(t)\Big]\ \mathrm{d}t,$$

we can now re-interpreted it as an *infinite horizon* problem with running cost $\Big[\lambda f(t; a_*(t))\Big]$ and a discounting factor λ.

For a sufficiently small $h > 0$, by Bellman's Optimality Principle, we have

$$\begin{aligned}
v(x,y) &= \int_0^h \mathrm{e}^{-\lambda t}\Big[\lambda f(t; a_*(t))\Big]\ \mathrm{d}t + \mathrm{e}^{-\lambda h} v\Big(f(h; a_*(h)),\ N(h; a_*(h))\Big) + o(h) \\
&= h\Big[\lambda x\Big] + (1 - \lambda h)\Big[v(x,y) + h v_x(x,y) \cdot F(x,y,a_*(0)) + h v_y(x,y) \cdot G(x,y,a_*(0))\Big] + o(h) \\
0 &= h\lambda x - h\lambda v(x,y) + h v_x(x,y) \cdot F(x,y,a_*(0)) + h v_y(x,y) \cdot G(x,y,a_*(0)) + o(h)
\end{aligned}$$

Now dividing both sides by h and sending h to 0, we obtain

$$0 = \lambda\Big(x - v(x,y)\Big) + v_x(x,y) \cdot F(x,y,a_*(0)) + v_y(x,y) \cdot G(x,y,a_*(0)).$$

Notice that the above equation involves $a_*(0)$ only. It is then natural to switch to a state-dependent optimal control in feedback form. The HJB equation that v satisfies is then

obtained by maximizing over $a = a(0) \in [0,1]$. By demanding the above equation holds for all $(x,y) \in [0,1]^2$, the PDE can be written as:

$$0 = \lambda\Big(x - v(x,y)\Big) + v_x(x,y) \cdot F(x,y,a) + v_y(x,y) \cdot G(x,y,a), \tag{5.42}$$

with the boundary condition

$$v(x,y) = \begin{cases} 1, & \text{if } x = 1 \text{ and } y \neq 0, \\ 0, & \text{if } y = 0 \text{ or } x = 0. \end{cases} \tag{5.43}$$

Substituting the actual definitions of F and G, we obtain (5.42) in the specific form:

$$\begin{aligned} 0 =& \lambda\left[x - v(x,y)\right] + \max_{a\in[0,1]} \left\{ \left(\nabla v(x,y) \cdot \begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\text{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\text{KS}}(1-y)] \end{bmatrix} \right) a \right\} \\ &+ \nabla v(x,y) \cdot \begin{bmatrix} x(1-x)(1-y)(r_{\text{KS}} - 1) \\ y(1-y)[1 + (r_{\text{KS}} - 1)x] \end{bmatrix}. \end{aligned} \tag{5.44}$$

The linear dependence on a yields the *bang-bang* property:

$$\alpha_1(x,y) = a_*(x,y) = \begin{cases} 1, & \text{if } \nabla v(x,y) \cdot \begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\text{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\text{KS}}(1-y)] \end{bmatrix} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.45}$$

For the "ultimately smart" killer

Recall its value function (5.13) from the main text:

$$w(x,y) = \mathbb{P}\Big(T_v\Big(x,y,a_*(\cdot)\Big) < T_d\Big(x,y,a_*(\cdot)\Big)\Big). \tag{5.46}$$

Let $\{\tau_i\}_{i=1}^{\infty}$ be an infinite sequence of random dilution times and $t_m = m\Delta t$, $m = 0,1,2,3,\ldots$ be a uniform time discretization. We again assume the optimal policy a_* exists, and let

$$\mathbb{E}^0 = \mathbb{E}[\,\cdot \mid f(0) = x,\ N(0) = y,\ \text{following } a_*].$$

Thus, by *law of total expectation*, we have

$$\begin{aligned} w(x,y) &= \mathbb{E}^0\Big[w(f(t_1),N(t_1)) \mid \tau_1 = t_1\Big]\mathbb{P}(\tau_1 = t_1) + \mathbb{E}^0\Big[w(f(t_1),N(t_1)) \mid \tau_1 > t_1\Big]\mathbb{P}(\tau_1 > t_1) \\ &= \Big(1 - e^{-\lambda\Delta t}\Big)w\Big(f(\Delta t), \rho N(\Delta t)\Big) + e^{-\lambda\Delta t}\, w\Big(f(\Delta t), N(\Delta t)\Big) + o(\Delta t) \end{aligned}$$

A first-order approximation around $t_0 = 0$ gives

$$w\Big(f(\Delta t), N(\Delta t)\Big) = w(x,y) + w_x(x,y)\cdot F(x,y,a_*(0))\Delta t + w_y(x,y)\cdot G(x,y,a_*(0))\Delta t + O(\Delta t^2)$$

$$w\Big(f(\Delta t), \rho N(\Delta t)\Big) = w(x,\rho y) + O(\Delta t)$$

And hence

$$\begin{aligned} \cancel{w(x,y)} &= \Big(1 - e^{-\lambda\Delta t}\Big)w\Big(f(\Delta t), \rho N(\Delta t)\Big) + e^{-\lambda\Delta t}\, w\Big(f(\Delta t), N(\Delta t)\Big) + o(\Delta t) \\ &= (\lambda\Delta t)\Big[w(x,\rho y) + O(\Delta t)\Big] \\ &\quad + (1-\lambda\Delta t)\left\{w(x,y) + w_x(x,y)\cdot F(x,y,a_*(0))\Delta t + w_y(x,y)\cdot G(x,y,a_*(0))\Delta t\right\} + O(\Delta t^2) \\ &= \lambda\Delta t w(x,\rho y) + \cancel{w(x,y)} - \lambda\Delta t w(x,y) \\ &\quad + w_x(x,y)\cdot F(x,y,a_*(0))\Delta t + w_y(x,y)\cdot G(x,y,a_*(0))\Delta t + O(\Delta t^2) \end{aligned}$$

Dividing it by Δt and take $\Delta t \downarrow 0$, we have

$$0 = \lambda\left[w(x,\rho y) - w(x,y)\right] + w_x(x,y)\cdot F(x,y,a_*(0)) + w_y(x,y)\cdot G(x,y,a_*(0)).$$

Notice that the above equation involves $a_*(0)$ only. It is then natural to switch to a state-dependent optimal control in feedback form. The HJB equation for (5.13) is then obtained by maximizing over $a = a(0) \in [0,1]$. By demanding the above equation holds for all $(x,y) \in [0,1]^2 \setminus \Delta$, the PDE can be written as:

$$0 = \lambda\left[w(x,\rho y) - w(x,y)\right] + \max_{a\in[0,1]}\Big\{w_x(x,y)\cdot F(x,y,a) + w_y(x,y)\cdot G(x,y,a)\Big\}, \tag{5.47}$$

with the boundary condition

$$w(x,y) = \begin{cases} 1, & \text{if } (x,y) \in \Delta_{\text{vic}} \text{ and } y \neq 0, \\ 0, & \text{if } (x,y) \in \Delta_{\text{dft}} \text{ or } y = 0. \end{cases} \tag{5.48}$$

Substituting the actual definitions of F and G, we obtain (5.47) in the specific form:

$$0 = \lambda\left[w(x,\rho y) - w(x,y)\right] + \max_{a\in[0,1]}\left\{\left(\nabla w(x,y)\cdot\begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\mathrm{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\mathrm{KS}}(1-y)]\end{bmatrix}\right)a\right\} \tag{5.49}$$

$$+\nabla w(x,y)\cdot\begin{bmatrix} x(1-x)(1-y)(r_{\mathrm{KS}}-1) \\ y(1-y)[1+(r_{\mathrm{KS}}-1)x]\end{bmatrix}.$$

The linear dependence on a yields the *bang-bang* property:

$$\alpha_\infty(x,y) = a_*(x,y) = \begin{cases} 1, & \text{if } \nabla w(x,y)\cdot\begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\mathrm{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\mathrm{KS}}(1-y)]\end{bmatrix} > 0, \\ 0, & \text{otherwise.}\end{cases} \tag{5.50}$$

As mentioned in the main text, our system dynamics with randomly-timed dilutions can be interpreted as a Piecewise-Deterministic Markov Process (PDMP). In general, the value function associated with a PDMP might not be smooth or even continuous. However, it can still often be interpreted as a unique (discontinuous) *viscosity solution* of the HJB equation [44].

Remark 5.2: The linear non-local equation (5.9) in Box 10 can be derived in the same way but using a fixed policy $\hat{a}$ instead of a_*.

Remark 5.3: Let

$$\vec{\Upsilon}(x,y) := \begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\mathrm{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\mathrm{KS}}(1-y)]\end{bmatrix}.$$

For both the "stochastic-myopic" killer and the "ultimately smart" killer, the respective HJB equation is linear in a, which yields a *generally* bang-bang optimal policy. However, we note that singular controls may arise when either $\nabla v(x,y)\cdot\vec{\Upsilon}(x,y)$ or $\nabla w(x,y)\cdot\vec{\Upsilon}(x,y)$ is equal to 0. But by computing the vector field associated with both α_1 and α_∞, we find no vector tangential to the boundary of $a_*(x,y) = 1$, thereby excluding the possibility of

singular arcs in our models. The same situation applies to the optimal policy $a_*(x, y, t)$ for the regular/period dilutions as well.

5.4.4 Numerical methods and implementation details

In this section, we provide the numerical schemes and implementation details of solving: (i) the finite horizon HJB PDE for $u(x, y, t)$ with regular/periodic dilutions; (ii) the limiting fractions $\hat{u}^\infty$ and total population $\hat{\phi}^\infty$ with regular dilutions; and (iii) the non-local HJB equation for $w(x, y)$ for the "ultimately smart" killers. For (i) and (iii), the optimal feedback policy is found by numerically solving the corresponding HJB equation.

For the finite-horizon HJB

Recall from the main text that we define the fraction-maximizing *value function* as

$$u(x, y, t) = \sup_{a(\cdot)} f(T^-), \quad \text{with } f(0) = x,\ N(0) = y, \tag{5.51}$$

where u satisfies a time-dependent HJB PDE

$$-\frac{\partial u}{\partial t}(x, y, t) = \max_{a\in[0,1]} \left\{ \nabla u(x, y, t) \cdot \begin{bmatrix} F(x, y, a) \\ G(x, y, a) \end{bmatrix} \right\} \tag{5.52}$$

with the terminal condition

$$u(x, y, T^-) = \begin{cases} x, \text{ if } y > 0, \\ 0, \text{ if } y = 0. \end{cases}$$

Substituting the actual definitions of F and G, we obtain (5.52) in the specific form:

$$-\frac{\partial u}{\partial t}(x,y,t) = \max_{a\in[0,1]} \left\{ \left(\nabla u(x,y,t) \cdot \begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\text{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\text{KS}}(1-y)] \end{bmatrix} \right) a \right\} \tag{5.53}$$
$$+ \nabla u(x,y,t) \cdot \begin{bmatrix} x(1-x)(1-y)(r_{\text{KS}}-1) \\ y(1-y)[1+(r_{\text{KS}}-1)x] \end{bmatrix}.$$

The linear dependence on a yields the *bang-bang* property:

$$\alpha(x,y,t) = a_*(x,y,t) = \begin{cases} 1, & \text{if } \nabla u(x,y,t) \cdot \begin{bmatrix} x(1-x)[\gamma xy - \varepsilon r_{\text{KS}}(1-y)] \\ -xy[\gamma y(1-x) + \varepsilon r_{\text{KS}}(1-y)] \end{bmatrix} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{5.54}$$

The time-dependent total population, denoted by $\phi(x,y,t)$, satisfies a linear PDE

$$-\frac{\partial \phi}{\partial t}(x,y,t) = \nabla \phi(x,y,t) \cdot \begin{bmatrix} F\Big(x,y,\alpha(x,y,t)\Big) \\ G\Big(x,y,\alpha(x,y,t)\Big) \end{bmatrix} \tag{5.55}$$

with the terminal condition $\phi(x,y,T^-) = y$.

We approximate the solution to (5.52) by a first-order semi-Lagrangian discretization [54] on a uniform rectangular grid over the (x,y,t) space. I.e., $(x_i,y_j,t_k) = (i\Delta x, j\Delta y, k\Delta t)$, where $\Delta x = 1/M_x$, $\Delta y = 1/M_y$, $\Delta t = T/M_t$, while $i = 0,...,M_x$, $j = 0,...,M_y$, and $k = 0,...,M_t$. We further simplify the notation for the spatial part as $\Xi = \{(i\Delta x, j\Delta y) \mid i = 0,\ldots,M_x,\ j = 0,\ldots,M_y\}$. We will use $U^k_{i,j} \approx u(x_i,y_j,t_k)$ to denote the discretized approximation at (x_i,y_j,t_k), and similarly, $\Phi^k_{i,j} \approx \phi(x_i,y_j,t_k)$, $A^k_{i,j} \approx \alpha(x_i,y_j,t_k)$.

For a sufficiently small $\Delta t > 0$, a first-order approximation of $\Big(f(\Delta t;a),\ N(\Delta t;a)\Big)$ starting from $(f(0),N(0)) = (x_i,y_j)$ with a control value $a \in \{0,1\}$ is $\Big(\tilde{f}_{i,a},\ \tilde{N}_{j,a}\Big) = \Big(x_i + \Delta t F(x_i,y_j,a),\ y_j + \Delta t G(x_i,y_j,a)\Big)$. Let $\tilde{U}^{k+1}_{i,j,a} \approx u(\tilde{f}_{i,a},\tilde{N}_{j,a},t_{k+1})$. Since (5.51) is in a Mayer form [61], the discretized dynamic programming equation is

$$U^k_{i,j} = \max_{a\in\{0,1\}} \tilde{U}^{k+1}_{i,j,a} + o(\Delta t), \tag{5.56}$$

where $\tilde{U}^{k+1}_{i,j,a}$ is evaluated by a bi-linear interpolation using the U values from the 4 neighboring gridpoints surrounding $(\tilde{f}_{i,a}, \tilde{N}_{j,a})$. Note that $A^k_{i,j}$ is found as the argmax of (5.56) at each gridpoint. This straightforward time-marching scheme is summarized in Algorithm 4. In all of our numerical experiments, we have used $M_x = M_y = 1600$ on each side of the unit fN-square, and $\Delta t = 6.25 \times 10^{-3}$. *To obtain the numerical solution for constitutive killers, one can simply apply Algorithm 4 with $a = 1$ without the maximization.*

Remark 5.4: Despite the time-dependent nature of the problem, one is typically interested in the 0-th time slice of the value function (i.e., $u(x, y, 0)$), which predicts the value at $t = T$ starting from $t = 0$ for all initial states. We note that by setting a sufficiently large T in Algorithm 4, any k-th time slice serves as the 0-th slice for a reduced horizon of $T - t_k$. This allows us to obtain the prediction for a range of horizons simultaneously in a single sweep.

Algorithm 4: Finite-horizon value function computation

Initialize U, Φ, A at $t = T$ $(k = M_t)$ using the terminal condition;

for $t_k = k\Delta t,\ k = M_t - 1, \ldots 0$ **do**

  for *every* $(x_i, y_j) \in \Xi$ **do**

    for $a \in \{0, 1\}$ **do**

      $\tilde{f}_{i,a} = x_i + \Delta t * F(x_i, y_j, a)$;

      $\tilde{N}_{j,a} = y_j + \Delta t * G(x_i, y_j, a)$;

      $U^k_{i,j,a} \leftarrow u(\tilde{f}_{i,a}, \tilde{N}_{j,a}, t_{k+1})$ by interpolation;

    $U^k_{i,j} \leftarrow \max_{a \in \{0,1\}} \left\{ U^k_{i,j,a} \right\}$;

    $A^k_{i,j} \leftarrow \arg\max_{a \in \{0,1\}} \left\{ U^k_{i,j,a} \right\}$;

  for *every* $(x_i, y_j) \in \Xi$ **do**

    $\tilde{f}_i = x_i + \Delta t * F(x_i, y_j, A^k_{i,j})$;

    $\tilde{N}_j = y_j + \Delta t * G(x_i, y_j, A^k_{i,j})$;

    $\Phi^k_{i,j} \leftarrow \phi(\tilde{f}_i, \tilde{N}_j, t_{k+1})$ by interpolation;

Approximating population limits under competitions

When sensitives and killers compete under regular dilutions, we are interested in whether they can coexist or if one population will dominate the other. As mentioned in the main text, this can be numerically found by a repetitive mapping using $u(x, y, 0)$ and $\phi(x, y, 0)$ computed by Algorithm 4.

Let $\hat{u}(x, y) = u(x, y, 0)$ and $\hat{\phi}(x, y) = \phi(x, y, 0)$. We will use $\hat{u}^n$ and $\hat{\phi}^n$ to denote the relative abundance of the killer, and the normalized total population, respectively, by the end of n-th cycle. Thus, by definition, $\hat{u}^1 = \hat{u}$ and $\hat{\phi}^1 = \hat{\phi}$. Based on our assumption, after the n-th dilution, $f\Big((nT)^+\Big) = f\Big((nT)^-\Big)$ and $N\Big((nT)^+\Big) = \rho N\Big((nT)^-\Big)$. It follows that

$$\hat{u}^{n+1}(x, y) = \hat{u}\Big(\hat{u}^n(x, y),\ \rho\hat{\phi}^n(x, y)\Big),$$
$$\hat{\phi}^{n+1}(x, y) = \hat{\phi}\Big(\hat{u}^n(x, y),\ \rho\hat{\phi}^n(x, y)\Big),$$

with $f(0) = x$ and $N(0) = y$ and following policy α. We repeat the process until $\|\hat{u}^{n+1} - \hat{u}^n\|$ is small, and output $\hat{u}^\infty \approx \hat{u}^{n+1}$ and $\hat{\phi}^\infty \approx \hat{\phi}^{n+1}$ as the limits. Given our focus on this limit, we describe α as a "myopic" policy. This designation highlights that α is only optimal for a single period. The truly optimal policy for an infinite number of periods would typically adapt from one period to the next. Our full method of approximating $\hat{u}^\infty$ with this "myopic" policy is summarized in Algorithm 5 with the usual notations of solution on the discretized grid: $\hat{U}_{i,j} \approx \hat{u}(x_i, y_j)$, $\hat{\Phi}_{i,j} \approx \hat{\phi}(x_i, y_j)$, $\hat{U}^n_{i,j} \approx \hat{u}^n(x_i, y_j)$, and $\hat{\Phi}^n_{i,j} \approx \hat{\phi}^n(x_i, y_j)$.

Algorithm 5: Limiting performance of the myopic policy

Initialize $\hat{U}$, $\hat{\Phi}$, $\hat{U}^1$, $\hat{\Phi}^1$ using the outputs from Algorithm 4;

$n = 1$;

`err` = 1e6 ;

while *err* > *tol* **do**

  for *every* $(x_i, y_j) \in \Xi$ **do**

    $f_{\text{temp}} = \hat{U}^n_{i,j}$;

    $N_{\text{temp}} = \rho * \hat{\Phi}^n_{i,j}$;

    $\hat{U}^{n+1}_{i,j} \leftarrow \hat{u}\Big(f_{\text{temp}}, N_{\text{temp}}\Big)$ by interpolation;

    $\hat{\Phi}^{n+1}_{i,j} \leftarrow \hat{\phi}\Big(f_{\text{temp}}, N_{\text{temp}}\Big)$ by interpolation;

  `err` $= \Big\|\hat{U}^{n+1} - \hat{U}^n\Big\|$;

  $n \leftarrow n + 1$;

$\hat{U}^\infty = \hat{U}^{n+1}$;

$\hat{\Phi}^\infty = \hat{\Phi}^{n+1}$;

Remark 5.5: Given our interest in the limit as the number of dilutions approaches infinity, we can significantly accelerate Algorithm 5. Rather than updating just one cycle with $\hat{u}$ and $\hat{\phi}$ per iteration, we can exponentially increase the number of cycles updated at each iteration. Specifically, at the n-th iteration, we can update 2^n cycles by using the results from the previous iteration. Let $\tilde{u}^n$ and $\tilde{\phi}^n$ represent the values at the n-th iteration of this accelerated algorithm, starting with $\tilde{u}^1 = \hat{u}$ and $\tilde{\phi}^1 = \hat{\phi}$, we now have

$$\tilde{u}^{n+1}(x, y) = \tilde{u}^n\Big(\tilde{u}^n(x, y),\ \rho\tilde{\phi}^n(x, y)\Big),$$
$$\tilde{\phi}^{n+1}(x, y) = \tilde{\phi}^n\Big(\tilde{u}^n(x, y),\ \rho\tilde{\phi}^n(x, y)\Big),$$

with $\tilde{u}^n = \hat{u}^{2^n}$ and $\tilde{\phi}^n = \hat{\phi}^{2^n}$. This accelerated algorithm is summarized in Algorithm 6.

Algorithm 6: Accelerated computation of $\hat{u}^{\infty}$

Initialize $\tilde{U}^1$, $\tilde{\Phi}^1$ using the outputs from Algorithm 4;

$m = 1$;

`err` $=$ 1e6 ;

while *err* $>$ *tol* **do**

  for *every* $(x_i, y_j) \in \Xi$ **do**

    $f_{\text{temp}} = \tilde{U}^m_{i,j}$;

    $N_{\text{temp}} = \rho * \tilde{\Phi}^m_{i,j}$;

    $\tilde{U}^{m+1}_{i,j} \leftarrow \tilde{u}^m\Big(f_{\text{temp}}, N_{\text{temp}}\Big)$ by interpolation;

    $\tilde{\Phi}^{m+1}_{i,j} \leftarrow \tilde{\phi}^m\Big(f_{\text{temp}}, N_{\text{temp}}\Big)$ by interpolation;

  `err` $= \Big\|\tilde{U}^{m+1} - \tilde{U}^m\Big\|$;

  $m \leftarrow m + 1$;

$\hat{U}^{\infty} = \tilde{U}^{m+1}$;

$\hat{\Phi}^{\infty} = \tilde{\Phi}^{m+1}$;

Remark 5.6: Either Algorithm 5 or Algorithm 6 can be applied to each time-slice in Algorithm 4 to compute the $\hat{u}^{\infty}$ and $\hat{\phi}^{\infty}$ for a range of horizons simultaneously in a single sweep. This can be achieved by defining $\hat{u} = u(x, y, t_k)$ and $\hat{\phi} = \phi(x, y, t_k)$ at each k-th slice, and thus obtaining the limits for each reduced horizon $T - t_k$.

For the "ultimately smart" killers

The HJB PDE (5.42) associated with "stochastic-myopic" killers can be numerically computed via standard Value [19, 18, 20] (or Value-Policy [78]) Iterations with a semi-Lagrangian discretization [54, 4, 136]. Here, we propose a similar Value-Policy Iterations (VPI) scheme to compute (5.47) for the "ultimately smart" killers.

Assuming the same spatial discretization Ξ as in §5.4.4, we denote the approximate solution to the value function as $W_{i,j} \approx w(x_i, y_j)$. For a sufficiently small amount of time $\Delta t > 0$, We have shown in §5.4.4 that the foot of the characteristics starting from a gridpoint (x_i, y_j) with a control value $a \in \{0, 1\}$ lands at a new state $(\tilde{f}_{i,a},\ \tilde{N}_{j,a})$

Therefore, from the Dynamic Programming Principle (DPP), we have

$$w(x_i, y_j) = \max_{a\in\{0,1\}} \Big\{ \underbrace{\big(1 - e^{-\lambda\Delta t}\big)}_{\text{prob of arrival}} w\big(\tilde{f}_{i,a}, \rho\tilde{N}_{j,a}\big) \ + \ \underbrace{e^{-\lambda\Delta t}}_{\text{prob of not arrival}} w\big(\tilde{f}_{i,a}, \tilde{N}_{j,a}\big) \Big\} + o(\Delta t) \tag{5.57}$$

yielding the discretized version

$$W_{i,j} = \max_{a\in\{0,1\}} \Big\{ (1 - e^{-\lambda\Delta t})\check{W}_{i,j,a} \ + \ e^{-\lambda\Delta t}\tilde{W}_{i,j,a} \Big\}, \tag{5.58}$$

where $\check{W}_{i,j,a} \approx w(\tilde{f}_{i,a}, \rho\tilde{N}_{j,a})$ and $w(\tilde{f}_{i,a}, \tilde{N}_{j,a})$ are computed through a *bi-linear* interpolation of the W values from the four neighboring gridpoints surrounding $(\tilde{f}_i, \rho\tilde{N}_j)$ and $(\tilde{f}_i, \tilde{N}_j)$, respectively. The optimal feedback policy $\Pi_{i,j} \approx \pi(x_i, y_j)$ is recovered as an argmax in (5.58).

We start with value iterations where we solve the nonlinear (5.47) by a Gauss-Seidel relaxation. Let $W^n_{i,j} \approx w^n(x_i, y_j)$ and $\Pi^n_{i,j} \approx \pi^n(x_i, y_j)$ be the discretized solution/policy at the n-th iteration at gridpoint (x_i, y_j). We use `err` to denote the L_∞-norm of W-change in the current value iteration. Whenever `err` stagnates, we proceed to the "*policy-evaluation*" (PE) step.

In the PE step, we compute the value function by solving a system of linear equations with a fixed policy $\hat{\pi}$ (recovered from the most recent value iteration). A first-order approximation of the system (5.2) starting from (x_i, y_j) with policy $\hat{\Pi}_{i,j} \approx \hat{\pi}(x_i, y_j)$ is $\big(\tilde{f}_{i,\hat{\Pi}},\ \tilde{N}_{j,\hat{\Pi}}\big) = \big(x_i + \Delta t * F(x_i, y_j, \hat{\Pi}_{i,j}),\ y_j + \Delta t * G(x_i, y_j, \hat{\Pi}_{i,j})\big)$. We thus solve a linear system of equations

$$\hat{W}^{\hat{\Pi}}_{i,j} = \big(1 - e^{-\lambda\Delta t}\big)\check{W}^{\hat{\Pi},\rho}_{i,j} + e^{-\lambda\Delta t}\tilde{W}^{\hat{\Pi}}_{i,j}, \tag{5.59}$$

where $\check{W}_{i,j}^{\hat{\Pi},\rho} \approx \hat{w}^{\hat{\pi}}\left(\tilde{f}_{i,\hat{\Pi}}, \rho\tilde{N}_{j,\hat{\Pi}}\right)$ and $\tilde{W}_{i,j}^{\hat{\Pi}} \approx \hat{w}^{\hat{\pi}}\left(\tilde{f}_{i,\hat{\Pi}}, \tilde{N}_{j,\hat{\Pi}}\right)$ are again computed through a bi-linear interpolation.

After obtaining the solution to (5.59), we return to the value iteration part and repeat the process until $\texttt{err} < \texttt{tol}$, where `tol` is a preset tolerance of convergence. In all of our numerical experiments, we have used $M_x = M_y = 1600$ on each side of the unit fN-square, $\Delta t = 0.025$, and $\texttt{tol} = 10^{-6}$. our full method is summarized in Algorithm 7.

Under mild technical assumptions, Kushner and Dupuis [104, Chapters 10&16] showed that the discretized solution derived from a general jump-diffusion process converges to the value function using standard iterative methods. Our model forms a PDMP, which is just a specific case of jump-diffusion processes.

Algorithm 7: Value-Policy Iterations for the non-local HJB equation (5.47)

Initialize W^1 and Π^1 based on the boundary condition (5.15);

Prob_not_arrival $= \exp(-\lambda \Delta t)$;

Prob_arrival $= 1 -$ Prob_not_arrival;

$n = 1$;

`err` = 1e6 ;

while *err* $>$ *tol* **do**

  for *every* $(x_i, y_j) \in \Xi$ **do**

    for $a \in \{0, 1\}$ **do**

      $\tilde{f}_{i,a} = x_i + \Delta t * F(x_i, y_j, a)$;

      $\tilde{N}_{j,a} = y_j + \Delta t * G(x_i, y_j, a)$;

      $\tilde{W}_{\text{temp}} \leftarrow w^n(\tilde{f}_{i,a}, \tilde{N}_{j,a})$ by interpolation;

      $\check{W}_{\text{temp}} \leftarrow w^n(\tilde{f}_{i,a}, \rho \tilde{N}_{j,a})$ by interpolation;

      $W_{i,j,a}^{n+1,\text{temp}} \leftarrow$ Prob_arrival $* \check{W}_{\text{temp}} +$ Prob_not_arrival $* \tilde{W}_{\text{temp}}$;

    $W_{i,j}^{n+1,\text{temp}} \leftarrow \max\limits_{a \in \{0,1\}} \left\{ W_{i,j,a}^{n+1,\text{temp}} \right\}$;

    $\Pi_{i,j}^{n+1,\text{temp}} \leftarrow \arg\max\limits_{a \in \{0,1\}} \left\{ W_{i,j,a}^{n+1,\text{temp}} \right\}$;

    if $W_{i,j}^{n+1,\text{temp}} > W_{i,j}^{n}$ **then**

      $W_{i,j}^{n+1} \leftarrow W_{i,j}^{n+1,\text{temp}}$;

      $\Pi_{i,j}^{n+1} \leftarrow \Pi_{i,j}^{n+1,\text{temp}}$;

    else

      $W_{i,j}^{n+1} \leftarrow W_{i,j}^{n}$;

      $\Pi_{i,j}^{n+1} \leftarrow \Pi_{i,j}^{n}$;

  `err` $= \|W^{n+1} - W^n\|$;

  if *err stagnates* **then**

    "Policy-Evaluation" with Π^{n+1}

  $n \leftarrow n + 1$;

$W = W^{n+1}$;

$\Pi = \Pi^{n+1}$;

Remark 5.7: This is a semi-Lagrangian-based method for solving the linear non-local probabilistic performance metric (5.9) in the main text, we will apply it to assess the probability performance of α_0 and α_1.

5.4.5 Population-dependent (hyperbolic) win/defeat boundaries

In the main text, we have focused on *fraction-dependent (vertical)* boundary arising from the definitions of killers' victory and defeat:

$$T_v(x,y,\alpha(\cdot)) = \inf\Big\{t>0 \mid f(t) > \gamma_{\text{v}};\ f(0)=x,\ N(0)=y, \text{ with many dilutions}\Big\}, \quad (5.60)$$
$$T_d(x,y,\alpha(\cdot)) = \inf\Big\{t>0 \mid f(t) < \gamma_{\text{d}};\ f(0)=x,\ N(0)=y, \text{ with many dilutions}\Big\},$$

where both stopping criteria depended on the *fraction* of the killer strain, $f(t)$, only. In the main text, we have used $\gamma_{\text{v}} = 0.99$ and $\gamma_{\text{d}} = 0.01$.

However, there are many other suitable ways of defining killers' victory/defeat. In this section, we explore *population-dependent (hyperbolic)* win and defeat boundaries and demonstrate that the results remain qualitatively similar to those obtained in the main text.

Mathematically, we now define the (random) victory/defeat time as

$$T_v(x,y,\alpha(\cdot)) = \inf\Big\{t>0 \mid f(t)N(t) > \gamma^h_{\text{v}};\ f(0)=x,\ N(0)=y, \text{ with many dilutions}\Big\}, \quad (5.61)$$
$$T_d(x,y,\alpha(\cdot)) = \inf\Big\{t>0 \mid f(t)N(t) < \gamma^h_{\text{d}};\ f(0)=x,\ N(0)=y, \text{ with many dilutions}\Big\},$$

so that the stopping criteria depend on the *population size* of the killer strain, $n_{\text{K}}(t) = f(t)N(t)$.

As a result, while the equations for the "ultimately smart" killer ($w(x,y)$) and the probabilistic performance metric ($\hat{w}^{\alpha}(x,y)$) remain unchanged, their respective boundary condi-

tions are now specified on two *hyperbolas*:

$$w(x,y) = \begin{cases} 1, & \text{if } xy > \gamma_{\text{v}}^h, \\ 0, & \text{if } xy < \gamma_{\text{d}}^h. \end{cases} \tag{5.62}$$

$$\hat{w}^{\alpha}(x,y) = \begin{cases} 1, & \text{if } xy > \gamma_{\text{v}}^h, \\ 0, & \text{if } xy < \gamma_{\text{d}}^h. \end{cases} \tag{5.63}$$

Let w_v denote the value function with vertical boundaries and w_h the one with hyperbolic boundaries. Using the same parameter values as in Fig 5.9 in the main text but with $\gamma_{\text{d}}^h = 0.005$ and $\gamma_{\text{v}}^h = 0.95$, we find that α_∞ computed with hyperbolic boundaries (Fig 5.15(b)) is mostly the same as in Fig 5.9(b), except for the bottom right corner of the orange region. Consequently, the value function remains qualitatively the same too. To quantify the difference between w_v (Fig 5.9(a)) and w_h (Fig 5.15(a)), we calculate the mean absolute difference

$$\bar{D}(y) = \int_0^1 |w_v(x,y) - w_h(x,y)|\, dx. \tag{5.64}$$

and plot it for all initial populations $y \in [0,1]$. Fig 5.15(c) shows that $\bar{D}$ is only large when $y < 0.05$ due to a more abrupt transition from zero to a positive winning probability in w_v compared to w_h when x is close to 1 and y is small. This is not surprising since these initial conditions are much closer to the hyperbolic defeat boundary $xy = \gamma_{\text{d}}^h$ than they are to the vertical defeat boundary $x = \gamma_{\text{d}}$. For $y > 0.05$, this x-averaged difference is very close to zero, suggesting that for most initial conditions the probability of killers' winning is largely insensitive to the type of boundary used.

This conclusion also generally holds true when using the above hyperbolic boundaries to compute $\hat{w}$ with α_0 and α_1. Analogously to Fig 5.10 in the main text, we again focus on the initial condition $(f(0), N(0)) = (0.5, 0.1)$ and compare the performance of these toxin-production policies for a range of (ρ, λ) values. Fig 5.16 shows that these heat maps are both

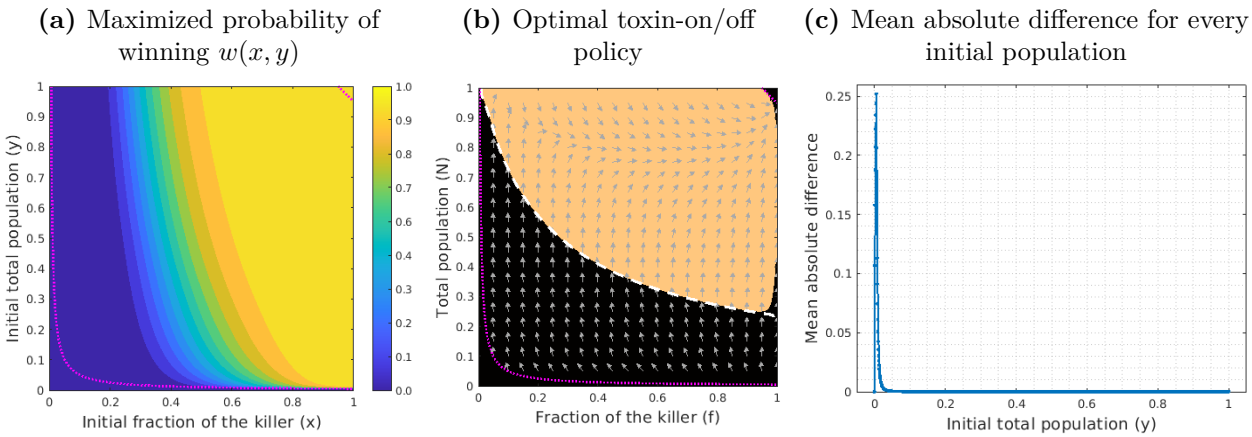


Figure 5.15. "Ultimately smart" killer with hyperbolic win/defeat boundaries. The optimal toxin-on region (orange in subfigure (b)) is almost the same as the one computed with vertical boundaries in Fig 5.9(b). (The toxin-on/off switch curve from the latter is shown here as a white-dashed line). As a result, the maximized probability of winning (subfigure (a)) is also very similar to the one computed with vertical boundaries in Fig 5.9(a). Subfigure (c) shows the x-averaged mean absolute difference between subfigure (a) and Fig 5.9(a) across all initial populations $y \in [0, 1]$. This difference is only noticeable when $y < 0.05$. In all subfigures, the victory and defeat barriers (γ_{v} and γ_{d}, respectively) are plotted with a magenta dotted line. All parameter values are the same as in Fig 5.9.

qualitatively and quantitatively similar to those in Fig 5.10.

We similarly quantify the boundary-related difference in the probabilistic performance of constitutive and stochastic-myopic killers in Fig 5.17. Let $\hat{w}_v$ and $\hat{w}_h$ denote the probabilistic performance with vertical and hyperbolic victory/defeat boundaries, respectively. Focusing on the same initial condition $(f(0), N(0)) = (0.5, 0.1)$, we observe that $|\hat{w}_v - \hat{w}_h|$ is negligible across half of the heat map ($\rho \geq 0.6$) for both α_∞ and α_1. In these two cases, a noticeable difference (with a maximum of approximately 0.017) is observed when the dilutions are strong ($\rho \leq 0.55$). For α_0, the maximum difference is slightly lower, around 0.014, while the region with noticeable differences is larger ($\rho \leq 0.6$). This is expected, as a stronger dilution more likely leads to a defeat under hyperbolic boundaries (due to a significantly larger Δ_{dft}) compared to vertical boundaries. These results indicate that our observations and conclusions in the main text remain largely unaffected by whether the victory or defeat

of the killer is defined by its fraction or population.

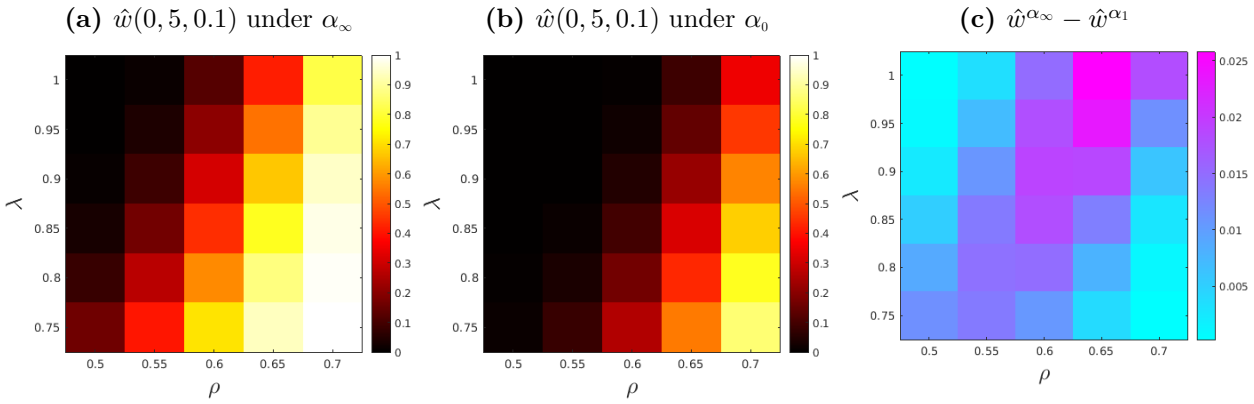


Figure 5.16. Hyperbolic boundaries: comparison of probabilistic performance for different types toxin-production policies starting from $(f(0), N(0)) = (0.5, 0.1)$ for a range of dilution strengths and frequencies. Policy α_1 is recomputed for each λ, while policy α_∞ is recomputed for each (ρ, λ) combination. The results remain both qualitatively and quantitatively similar to Fig 5.10 in the main text. A stronger survival rate (larger ρ) combined with less frequent dilutions (smaller λ) increases the chances of toxin-producers winning for all three policies. It is clear that the ultimately smart (subfigure (a)) and stochastic-myopic killers significantly outperform the constitutives (subfigure (b)). The differences in $\hat{w}(0.5, 0.1)$ between α_∞ and α_1 are still small, with the discrepancy increasing toward the upper right corner (subfigure (c)).

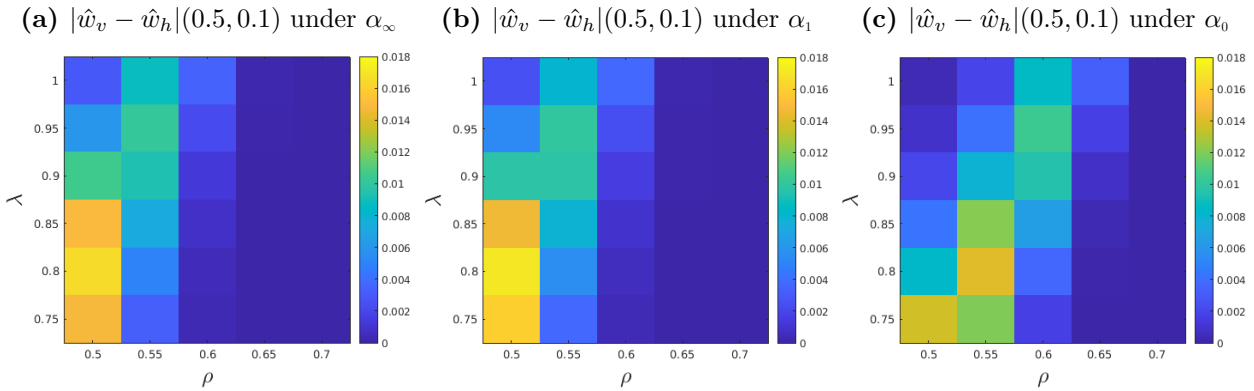


Figure 5.17. Hyperbolic boundaries: absolute difference in $\hat{w}(0.5, 0.1)$ resulting from two types of boundaries computed for a range of ρ and λ values. The differences under α_∞ (subfigure (a)) and α_1 (subfigure (b)) are again similar, with a maximum difference of around 0.017 when $\rho = 0.5$. For α_0 in subfigure (c), the region of noticeable differences is slightly larger ($\rho \leq 0.6$) although the maximum difference remains relatively small (≈ 0.014). All three subfigures share the same colorbar.

5.4.6 Monte Carlo simulations with "Binomial dilutions"

The results in the main text are all produced under "deterministic" dilution outcomes. That is, after each dilution, the relative abundances are preserved while only a ρ proportion of the total population survives. In this section, we present results under a specific form of *random* dilution outcomes and demonstrate, using Monte Carlo (MC) simulations, that they are qualitatively similar to the previous results.

In particular, we adopt a "Binomial Sampling" strategy to produce random dilution outcomes, which we refer to as "Binomial dilutions." Let f^- be the pre-dilution fraction of the killers, and N^- be the pre-dilution total population. Accordingly, the actual pre-dilution number of killer cells is $n_{\mathrm{K}}^- = f^- N^- C$, and the pre-dilution number of sensitive cells is $n_{\mathrm{S}}^- = (1 - f^-)N^- C$. We assume each cell has an independent survival probability ρ after each dilution. As a result, the post-dilution number of cells is a *Binomial random variable*:

- Post-dilution number of killer cells: $n_{\mathrm{K}}^+ = \mathrm{Bi}(n_{\mathrm{K}}^-, \rho)$;
- Post-dilution number of sensitive cells: $n_{\mathrm{S}}^+ = \mathrm{Bi}(n_{\mathrm{S}}^-, \rho)$.

Consequently, the random post-dilution (normalized) total population is $N^+ = \dfrac{n_{\mathrm{K}}^+ + n_{\mathrm{S}}^+}{C}$, and the random post-dilution fraction of killers is $f^+ = \dfrac{n_{\mathrm{K}}^+}{n_{\mathrm{K}}^+ + n_{\mathrm{S}}^+}$, which will serve as the initial condition for the next cycle.

We first conduct Monte Carlo simulations on a uniform grid. Starting from each $(x_i, y_j) = (i/10, j/10)$ with $i,\ j = 1, 2, \ldots 9$, each sample was simulated with $n = 200$ dilutions. Fig 5.19 shows the empirical distributions of $f((nT)^-)$ with $T = 1$ and Fig 5.18(a) shows their respective means. We observe that most distributions are unimodal, being either nearly 0 (all sensitives) or nearly 1 (all killers). The two exceptions with a bimodal distribution, peaking at 0 or 1, intersect precisely with the boundary (black-dashed line in

Fig 5.18(a)) that separates the initial conditions leading to a *deterministic* victory of the killers under proportional dilutions. Additionally, the killer-winning region (dark-red background in Fig 5.18(a)) under "Binomial dilutions" aligns well with this "deterministically-killer-winning" region shown in Fig 5.4(c) in the main text. This is not surprising since the stochastic fluctuations introduced by "Binomial dilutions" can be sufficiently large to cause samples starting near the boundary to drift towards either competitive exclusion over successive dilutions. See Fig 5.18(c) for such an example starting from $(x, y) = (0.5, 0.4)$. However, these fluctuations are typically insufficient to alter the fate of samples starting further from the boundary, where initial conditions strongly favor one strain.

Considering that focusing on a single initial condition for all samples might not capture enough information, we conducted additional Monte Carlo simulations using "Binomial dilutions" with *uniformly random in a cell* initial conditions. Specifically, for each grid cell centered at (x_i, y_j), the initial condition for each sample was chosen uniformly at random from the square $(x, y) \in [x_i - 0.05, x_i + 0.05] \times [y_j - 0.05, y_j + 0.05]$. This strategy diversifies the range of initial conditions, increasing the likelihood of intersecting the boundary of the "deterministically-killer-winning" region. As a result, we see from Fig 5.18(b) that almost all cells intersecting or near the dashed-line boundary now exhibit intermediate mean values. Moreover, Fig 5.20 shows that these cells again have a bimodal distribution, with random dilution outcomes pushing the dynamics towards one of the two competitive exclusions. Comparing Fig 5.20 with Fig 5.19, we further observe that the median at each grid cell remains unchanged. This consistency underscores the robustness of our simulation results, indicating that the conclusions in the main text would largely remain valid even under "Binomial dilutions."

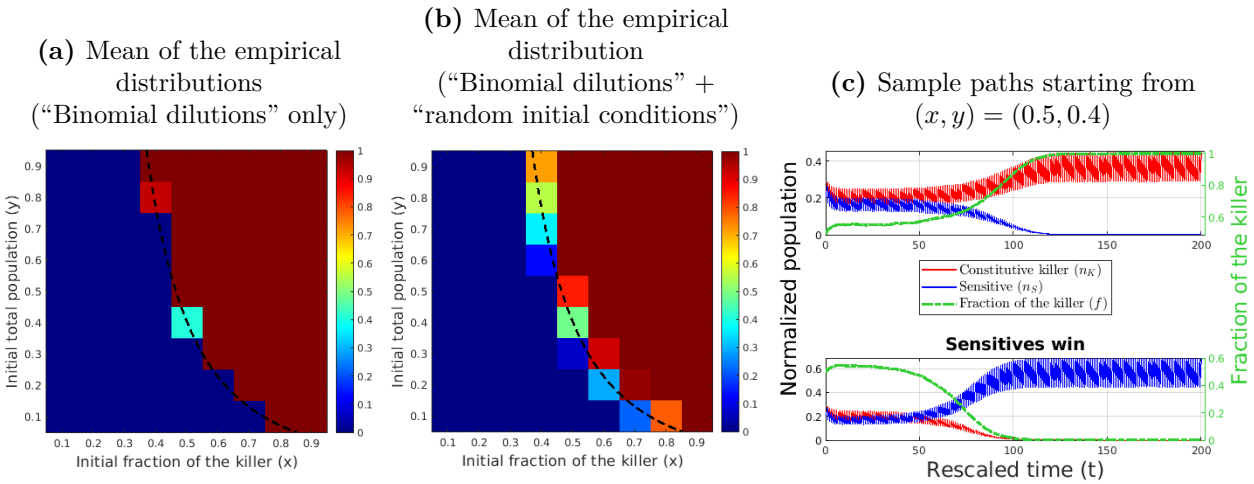


Figure 5.18. Monte Carlo simulations with "Binomial dilutions" on a uniform cell grid. (a) The mean of the empirical distribution, sampled with "Binomial dilutions" starting from the center of each cell in (x, y) space. The resulting distribution is almost always unimodal (dark blue - all sensitives; dark red – all killers). The exceptions are seen in only two cells among those intersected by the boundary (shown by a black dashed line) that separates the initial conditions leading to a deterministic victory of the killers under proportional dilutions (cf. Fig 5.4(c) in the main text). (b) Most means of the empirical distribution, sampled with both "Binomial dilutions" and "uniformly random in a cell" initial conditions, are also close to 0 or close to 1 in most cells. However, most cells that intersect or are close to that dashed line boundary now have more diverse intermediate mean values. In both cases, such cells exhibit a bimodal distribution with peaks at 0 and 1; see Figs 5.19&5.20 for the actual distributions. Subfigure (c) shows two sample trajectories starting from $(x, y) = (0.5, 0.4)$ resulting in different competitive exclusion outcomes due to the randomness in Binomial dilutions. All Monte Carlo simulations were conducted with 10^5 samples and 200 dilutions using parameter values $T = 1$, $\varepsilon = 0.2$, $r_{\text{KS}} = 0.85$, $\gamma = 1$, and $\rho = 0.65$. In (a), all samples for each grid cell start from the same initial condition $(x_i, y_j) = (i/10, j/10)$ with $i, j = 1, \ldots, 9$. In (b), for each grid cell centered at (x_i, y_j), the initial condition for each sample was chosen uniformly at random from the square $(x, y) \in [x_i - 0.05, x_i + 0.05] \times [y_j - 0.05, y_j + 0.05]$.

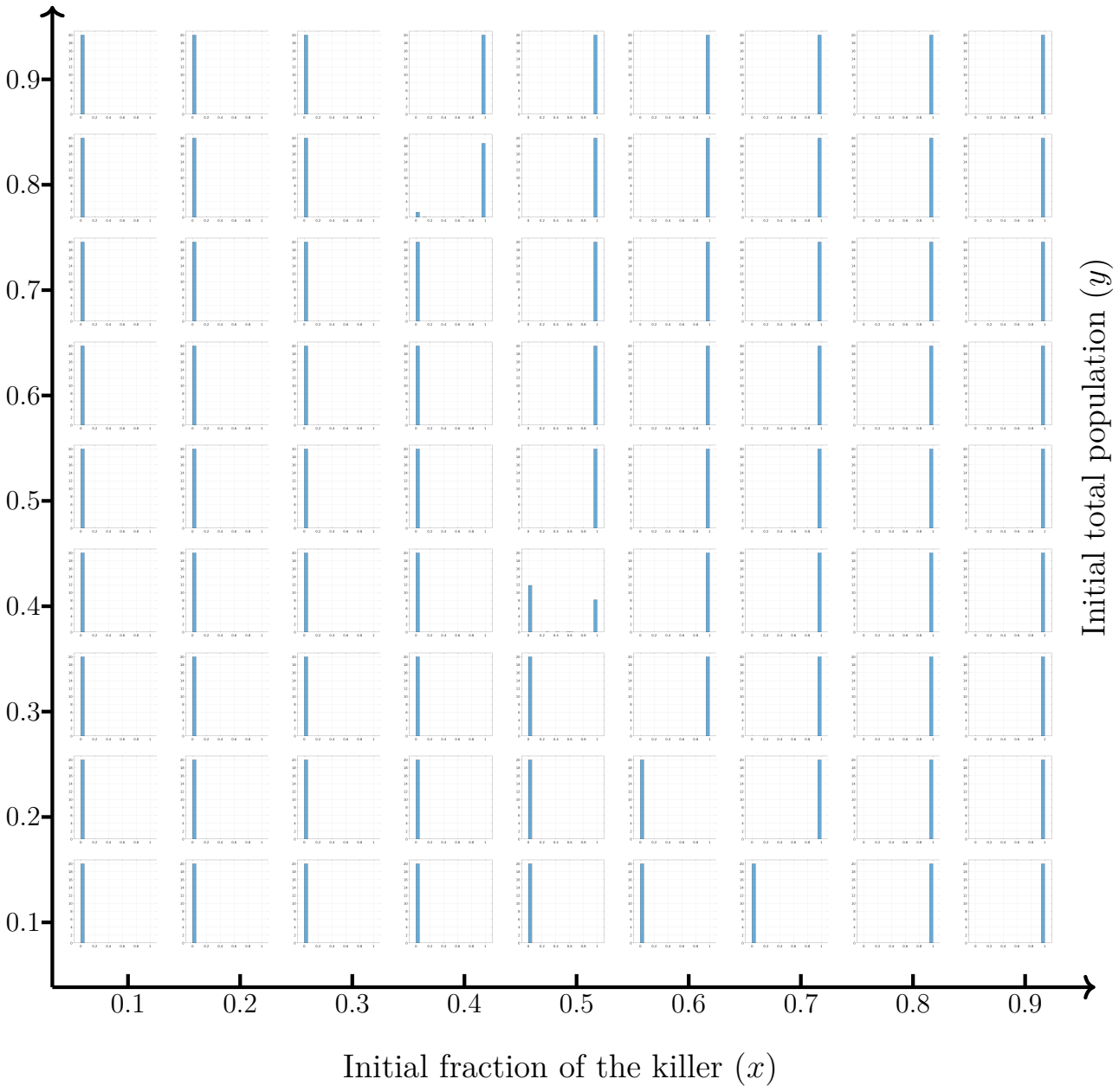


Figure 5.19. Empirical distributions of the fraction of killers with "Binomial dilutions" on a uniform grid. Most of the distributions are unimodal (either almost entirely 0 or almost entirely 1), except for two that are bimodal. The horizontal axis (of the entire figure) represents the initial fraction of the killer while the vertical axis encodes the initial total population for the simulations. For each subfigure, an empirical distribution of f (starting from the same $(x_i, y_j) = (i/10, j/10)$ with $i, j = 1, \ldots, 9$) after 200 dilutions is shown by a histogram. All subfigures share the same horizontal and vertical axes. The parameter values are the same as in Fig 5.18(a).

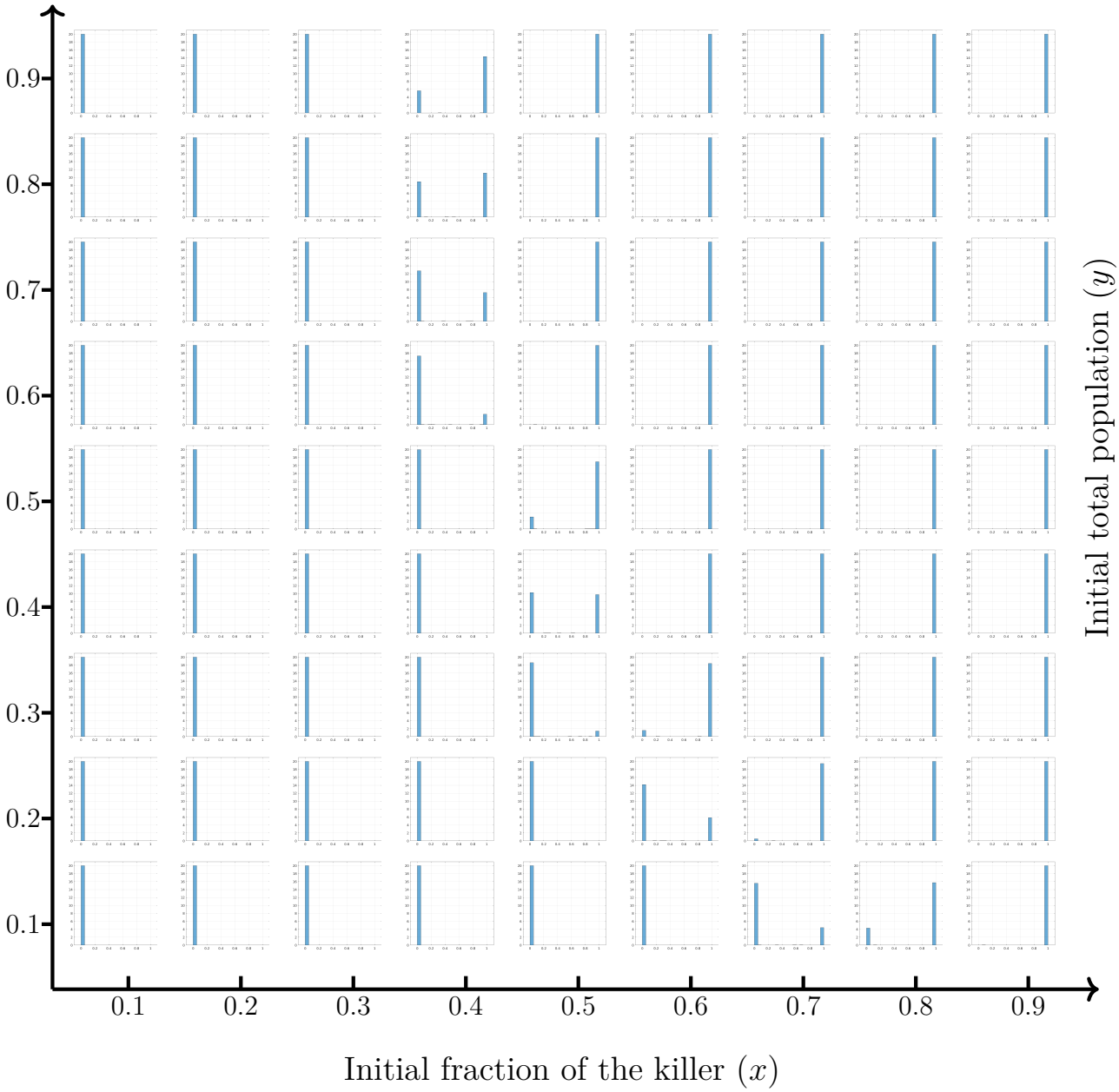


Figure 5.20. Empirical distributions of the fraction of killers with "Binomial dilutions" and "uniformly random in a cell" initial conditions on a cell grid. Most of the distributions are unimodal (either almost entirely 0 or almost entirely 1). However, the ones near the boundary of the "deterministically-winning" region (depicted as a black-dashed line) are *bimodal*. The horizontal axis (of the entire figure) represents the initial fraction of the killer while the vertical axis encodes the initial total population for the simulations. For each subfigure, an empirical distribution of $f(t)$ after 200 dilutions, with the initial condition chosen uniformly at random within the grid cell centered at $(x_i, y_j) = (i/10, j/10)$ with $i, j = 1, \ldots, 9$, is shown by a histogram. All subfigures share the same horizontal and vertical axes. The parameter values are the same as in Fig 5.18(b).

5.4.7 Strains, oligos, and plasmids

Tables 5.1, 5.2, and 5.3 report oligos, plasmids and strains used for the experiments. We substituted ymCherry in pAG11 with CyOFP1opt from pRP008 using Gibson assembly, producing pAG134, whose sequence was verified via long-read sequencing. Strain yAG171 was obtained by digesting pAG134 with PpuMI and transforming it into yAG75 [72] for integration of pAG134 in T_{CYC1}. Strain yAG177 was obtained by digesting pAG5 [72] with PpuMI and transforming it into yAG74 [72] for integration of pAG5 in P_{ACT1}.

Oligo name	Oligo sequence
oAG187	cgctgaggacTTGACCACACCTCTACCGG
oAG188	gctgaggcatTCGACCTGCAGCGTACGAAG
oAG189	tgcaggtcgaATGCCTCAGCACTAGTCCTG
oAG190	tgcaggtcgaATGCCTCAGCACTAGTCCTG

Table 5.1. DNA oligos used to assemble pAG134 via Gibson assembly. Bases in capital letters represent homology to the PCR template (pAG11 for yAG187/188 and pRP008 for yAG189/190) Lowercase bases represent homology to the backbone or fragment for Gibson assembly.

Plasmid name	Relevant transcriptional units	Sourced from
pAG5	P_{ACT1}-ymCitrine-T_{ADH1} P_{TEF}-KanMX6-T_{TEF}	[72]
pAG11	P_{ACT1}-ymCherry-T_{ADH1} P_{TEF}-KanMX6-T_{TEF} P_{GAL1}-K1-T_{CYC1}	[72]
pRP008	P_{TEF1}-CyOFP1opt-T_{ADH1}	[129]
pAG134	P_{ACT1}-CyOFP1opt-T_{ADH1} P_{TEF}-KanMX6-T_{TEF} P_{GAL1}-K1-T_{CYC1}	This study

Table 5.2. Plasmids used for strain construction.

Strain name	Genotype
yAG171	can1-100, his3-11,15, ura3Δ0, BUD4-S288C, gal1/10Δ::LEU2, prGAL3Δ::His3MX6-P_{ACT1}-GAL3, hxk2Δ::HphMX4, P_{TEF1}-CyOFP1Opt-T_{ADH1}, P_{TEF1}-KanMX6-T_{TEF}, prGAL1-K1-T_{CYC1}
yAG177	can1-100, his3-11,15, ura3Δ0, BUD4-S288C, gal1/10Δ::LEU2, P_{GAL3}Δ::His3MX6-P_{ACT1}-GAL3, P_{ACT1}-ymCitrine-T_{ADH1}, P_{TEF1}-KanMX6-T_{TEF}

Table 5.3. Strains used for the experiments.

CHAPTER 6

CONCLUSION

This thesis explores various topics within stochastic optimal control. In Chapter 2, we introduced a novel threshold (risk)-aware robust control framework, tailored for adaptive cancer therapy under a drift-diffusion process modeling the cancer dynamics. We devised an efficient semi-Lagrangian-based scheme to solve the Hamilton-Jacobi-Bellman PDE that arises from this framework. We provided detailed mathematical derivations and implementation of such algorithm (including control synthesis) in Chapter 3. Chapter 4 extended this approach to hybrid control problems, illustrated by sailboat routing under wind direction uncertainty. We again exploited the causality inherent in this problem to develop another efficient numerical scheme to solve a pair of quasi-variational inequalities. In Chapter 5, we shifted to the impact of environmental extreme events (dilutions) on bacterial competition. We analyzed both deterministic periodic events and randomly-timed events modeled by a Poisson process. For each scenario, we presented an effective algorithm to efficiently solve the corresponding equations, including a non-local Hamilton-Jacobi-type equation.

Although we have discussed specific extensions for each problem in the corresponding chapters, there are several common extensions we would like to highlight. First, the drift-diffusion processes discussed in Chapters 2 and 4 are extensively used in the literature but are relatively limited. Expanding our analysis to include a broader range of stochastic processes, such as a general jump-diffusion process [127, 151, 67], would be crucial to ensure our framework's broader applicability. On the numerical side, our algorithm benefits significantly from the inherent causality of the problem. It would be interesting to develop algorithms that can handle scenarios where the budget is not strictly decreasing; e.g., a non-negative running cost $K \geq 0$ (instead of $K > 0$) along the path to the target. This development would involve creating a numerical method that integrates the "s-marching" technique from

Chapter 3 with Value Iterations or Value-Policy Iterations. Although we have obtained preliminary results under a bang-bang control framework, we chose not to include them in this thesis as they are still in the early stages of research. Additionally, exploring situations where budget consumption is random or where the budget can be "re-filled" is definitely of interest. However, it would significantly complicate the mathematical formulations and increase computational time and complexities.

In our analysis across various chapters, we have assumed perfect information availability at any given moment, which is unrealistic in many situations. For example, in Chapter 2, biopsies for cancer patients are only conducted every few weeks or months, and tumors may even be undetectable due to physical constraints. In Chapter 4, wind conditions might be assessed intermittently, rendering the precise upwind direction imperfect at a given moment. Similarly, in Chapter 5, toxin-producing bacteria can only sense partial information about their environment. These scenarios involve dealing with partially-observable states or imperfect and intermittent state information, making them mathematically and computationally more complex. The former requires managing a "belief state" [17], where the probabilistic estimation process is challenging [92, 176]. The latter encompasses a broader range of topics including robust control [98, 109], adaptive control [21], and model predictive control [100].

Finally, although we have extensively utilized the Value-Policy Iterations (VPI) method [78] in this thesis, it has certain limitations. Primarily, as the dimensionality of the state space increases, the corresponding sparse linear systems become excessively large to be physically handled. While the algebraic multigrid (AMG) method [141] provides some relief, its effectiveness is still limited. Moreover, although VPI guarantees convergence, determining the actual rate of convergence remains an open question. It is also desirable to develop a more sophisticated criterion for transitioning to the "Policy-Evaluation" step once the change per iteration stagnates. Many of these challenges mentioned here are primarily due to the

curse of dimensionality inherent in dynamic programming. Moving forward, we are eager to explore approximate dynamic programming and reinforcement learning as potential avenues to address these limitations.

BIBLIOGRAPHY

[1] Mazen Alamir. Robust feedback design for combined therapy of cancer. *Optimal Control Applications and Methods*, 35(1):77–88, 2014.

[2] Edward Allen. *Modeling with Itô stochastic differential equations*, volume 22. Springer Science & Business Media, 2007.

[3] Daniel R Amor, Christoph Ratzke, and Jeff Gore. Transient invaders can induce shifts between alternative stable states of microbial communities. *Science Advances*, 6(8):eaay8676, 2020.

[4] June Andrews and Alexander Vladimirsky. Deterministic control of randomly-terminated processes. *Interfaces and Free Boundaries*, 16(1):1–40, 2014.

[5] Sara Arbulu and Morten Kjos. Revisiting the multifaceted roles of bacteriocins. *Microbial Ecology*, 87(1):1–14, 2024.

[6] Mohamed Assellaou, Olivier Bokanowski, and Hasnaa Zidani. Error estimates for second order Hamilton-Jacobi-Bellman equations. approximation of probabilistic reachable sets. *Discrete and Continuous Dynamical Systems-Series A*, 35(9):3933–3964, 2015.

[7] Nina S Atanasova, Maija K Pietilä, and Hanna M Oksanen. Diverse antimicrobial interactions of halophilic archaea and bacteria extend over geographical distances and cross the domain barrier. *MicrobiologyOpen*, 2(5):811–825, 2013.

[8] Martino Bardi and Italo Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhäuser, Boston, MA, 1997.

[9] Martino Bardi, Maurizio Falcone, and Pierpaolo Soravia. Numerical methods for pursuit-evasion games via viscosity solutions. In *Stochastic and differential games: theory and numerical methods*, pages 105–175. Springer, 1999.

[10] Martino Bardi, P Soravia, and M Falcone. Fully discrete schemes for the value function of pursuit-evasion games. In *Advances in dynamic games and applications*, pages 89–105. Springer, 1994.

[11] Belinda Barnes, Harvinder Sidhu, and David M Gordon. Host gastro-intestinal dynamics and the frequency of colicin production by escherichia coli. *Microbiology*, 153(9):2823–2827, 2007.

[12] Samuel E Barnett, Nicholas D Youngblut, Chantal N Koechli, and Daniel H Buckley. Multisubstrate dna stable isotope probing reveals guild structure of bacteria that mediate soil carbon cycling. *Proceedings of the National Academy of Sciences*, 118(47):e2115292118, 2021.

[13] Tamer Başar and Pierre Bernhard. *H^∞-optimal control and related minimax design problems*. Birkhauser, 1995.

[14] Joachim Becker, Nico Eisenhauer, Stefan Scheu, and Alexandre Jousset. Increasing antagonistic interactions cause bacterial communities to collapse at high diversity. *Ecology Letters*, 15(5):468–474, 2012.

[15] Michael J Behrenfeld, Yongxiang Hu, Robert T O'Malley, Emmanuel S Boss, Chris A Hostetler, David A Siegel, Jorge L Sarmiento, Jennifer Schulien, Johnathan W Hair, Xiaomei Lu, et al. Annual boom–bust cycles of polar phytoplankton biomass revealed by space-based lidar. *Nature Geoscience*, 10(2):118–122, 2017.

[16] Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[17] Alain Bensoussan. Stochastic control of partially observable systems. *(No Title)*, 1992.

[18] D. Bertsekas. *Dynamic Programming and Optimal Control: Volume II; Approximate Dynamic Programming*. Athena Scientific optimization and computation series. Athena Scientific, 2012.

[19] Dimitri Bertsekas. *Abstract dynamic programming*. Athena Scientific, 2022.

[20] Dimitri P Bertsekas. Value and policy iterations in optimal control and adaptive dynamic programming. *IEEE transactions on neural networks and learning systems*, 28(3):500–509, 2015.

[21] Robert R Bitmead, Michel Gevers, and Vincent Wertz. Adaptive optimal control the thinking man's gpc. 1990.

[22] Sean C Booth, William PJ Smith, and Kevin R Foster. The evolution of short-and long-range weapons for bacterial competition. *Nature Ecology & Evolution*, 7(12):2080–2091, 2023.

[23] Vivek S Borkar. Controlled diffusion processes. *Probability surveys*, 2:213–244, 2005.

[24] Michelle Boué and Paul Dupuis. Markov Chain Approximations for Deterministic

Control Problems with Affine Dynamics and Quadratic Cost in the Control. *SIAM Journal on Numerical Analysis*, 36(3):667–695, 1999.

[25] Carlos A Braumann. Environmental versus demographic stochasticity in population growth. In *Workshop on Branching Processes and Their Applications*, pages 37–52. Springer, 2010.

[26] MR Brown, JC Baptista, M Lunn, DL Swan, SJ Smith, RJ Davenport, BD Allen, WT Sloan, and TP Curtis. Coupled virus-bacteria interactions and ecosystem function in an engineered microbial system. *Water Research*, 152:264–273, 2019.

[27] Alexander P Browning, Jesse A Sharp, Tarunendu Mapder, Christopher M Baker, Kevin Burrage, and Matthew J Simpson. Persistence as an optimal hedging strategy. *Biophysical Journal*, 120(1):133–142, 2021.

[28] Sean W Buskirk, Alecia B Rokes, and Gregory I Lang. Adaptive evolution of nontransitive fitness in yeast. *Elife*, 9:e62238, 2020.

[29] S. Cacace, R. Ferretti, and A. Festa. Stochastic hybrid differential games and match race problems. *Applied Mathematics and Computation*, 372:124966, 2020.

[30] Cécile Carrère. Optimization of an in vitro chemotherapy to avoid resistant tumours. *Journal of Theoretical Biology*, 413:24–33, 2017.

[31] Cécile Carrère and Hasnaa Zidani. Stability and reachability analysis for a controlled heterogeneous population of cells. *Optimal Control Applications and Methods*, 41(5):1678–1704, 2020.

[32] Elliot Cartee, April Nellis, Jacob Van Hook, Antonio Farah, and Alexander Vladimirsky. Quantifying and managing uncertainty in piecewise-deterministic Markov processes. *SIAM/ASA Journal on Uncertainty Quantification*, 11(3):814–847, 2023.

[33] Lin Chao and Bruce R Levin. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proceedings of the National Academy of Sciences*, 78(10):6324–6328, 1981.

[34] Peter Chesson. Multispecies competition in variable environments. *Theoretical population biology*, 45(3):227–276, 1994.

[35] Rebecca H Chisholm, Tommaso Lorenzi, and Jean Clairambault. Cell population heterogeneity and evolution towards drug resistance in cancer: biological and mathe-

matical assessment, theoretical treatment optimisation. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(11):2627–2645, 2016.

[36] Andrew J. Coldman and J.M. Murray. Optimal control for a stochastic model of cancer chemotherapy. *Mathematical Biosciences*, 168(2):187–200, 2000.

[37] Joseph H Connell. Diversity in tropical rain forests and coral reefs: high diversity of trees and corals is maintained only in a nonequilibrium state. *Science*, 199(4335):1302–1310, 1978.

[38] Raymond Copeland, Christopher Zhang, Brian K Hammer, and Peter J Yunker. Spatial constraints and stochastic seeding subvert microbial arms race. *PLOS Computational Biology*, 20(1):e1011807, 2024.

[39] Michael G Crandall and Pierre-Louis Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–42, 1983.

[40] Jessica Cunningham, Frank Thuijsman, Ralf Peeters, Yannick Viossat, Joel Brown, Robert Gatenby, and Kateřina Staňková. Optimal control to reach eco-evolutionary stability in metastatic castrate-resistant prostate cancer. *PLoS One*, 15(12):e0243386, 2020.

[41] Jessica J Cunningham, Joel S Brown, Robert A Gatenby, and Kateřina Staňková. Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. *Journal of theoretical biology*, 459:67–78, 2018.

[42] Manoshi S Datta, Elzbieta Sliwerska, Jeff Gore, Martin F Polz, and Otto X Cordero. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature communications*, 7(1):11965, 2016.

[43] M. H. A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):353–388, 1984.

[44] Mark H. A. Davis and Mohammad Farid. *Piecewise-Deterministic Processes and Viscosity Solutions*, pages 249–268. Birkhäuser Boston, Boston, MA, 1999.

[45] Roger S Day. Treatment sequencing, asymmetry, and uncertainty: protocol strategies for combination chemotherapy. *Cancer Research*, 46(8):3876–3885, 1986.

[46] Andrew Dhawan, Daniel Nichol, Fumi Kinose, Mohamed E Abazeed, Andriy Marusyk, Eric B Haura, and Jacob G Scott. Collateral sensitivity networks reveal evolutionary

instability and novel treatment strategies in alk mutated non-small cell lung cancer. *Scientific reports*, 7(1):1–9, 2017.

[47] Sergey Dolgov, Dante Kalise, and Karl K Kunisch. Tensor decomposition methods for high-dimensional Hamilton–Jacobi–Bellman equations. *SIAM Journal on Scientific Computing*, 43(3):A1625–A1650, 2021.

[48] Rick Durrett and Simon Levin. Allelopathy in spatially distributed populations. *Journal of theoretical biology*, 185(2):165–171, 1997.

[49] Michaela J Eickhoff and Bonnie L Bassler. Snapshot: bacterial quorum sensing. *Cell*, 174(5):1328–1328, 2018.

[50] Steinar Engen, Øyvind Bakke, and Aminul Islam. Demographic and environmental stochasticity-concepts and definitions. *Biometrics*, pages 840–846, 1998.

[51] Pedro M Enriquez-Navas, Yoonseok Kam, Tuhin Das, Sabrina Hassan, Ariosto Silva, Parastou Foroutan, Epifanio Ruiz, Gary Martinez, Susan Minton, Robert J Gillies, et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Science translational medicine*, 8(327):327ra24–327ra24, 2016.

[52] Peyman Mohajerin Esfahani, Debasish Chatterjee, and John Lygeros. The stochastic reach-avoid problem and set characterization for diffusions. *Automatica*, 70:43–56, 2016.

[53] Giorgio Fabbri, Fausto Gozzi, and Andrzej Swiech. *Stochastic optimal control in infinite dimension : dynamic programming and HJB equations*. Springer, 2017.

[54] Maurizio Falcone and Roberto Ferretti. *Semi-Lagrangian approximation schemes for linear and Hamilton—Jacobi equations*. SIAM, 2013.

[55] Nathan Farrokhian, Jeff Maltas, Mina Dinh, Arda Durmaz, Patrick Ellsworth, Masahiro Hitomi, Erin McClure, Andriy Marusyk, Artem Kaznatcheev, and Jacob G Scott. Measuring competitive exclusion in non–small cell lung cancer. *Science Advances*, 8(26):eabm7212, 2022.

[56] Judith Feichtmayer, Li Deng, and Christian Griebler. Antagonistic microbial interactions: contributions and potential applications for controlling pathogens in the aquatic systems. *Frontiers in microbiology*, 8:281883, 2017.

[57] Michael Feldgarden and Margaret A Riley. The phenotypic and fitness effects of colicin resistance in escherichia coli k-12. *Evolution*, 53(4):1019–1027, 1999.

[58] Roberto Ferretti. A technique for high-order treatment of diffusion terms in semi-Lagrangian schemes. *Communications in Computational Physics*, 8(2):445–70, 2010.

[59] Roberto Ferretti and Adriano Festa. Optimal route planning for sailing boats: A hybrid formulation. *Journal of Optimization Theory and Applications*, 181(3):1015–1032, 2019.

[60] Andrej Fischer, Ignacio Vázquez-García, and Ville Mustonen. The value of monitoring to control evolving populations. *Proceedings of the National Academy of Sciences*, 112(4):1007–1012, 2015.

[61] Wendell H Fleming and Raymond W Rishel. *Deterministic and stochastic optimal control*, volume 1. Springer Science & Business Media, 2012.

[62] Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.

[63] Jason J Flowers, Tracey A Cadkin, and Katherine D McMahon. Seasonal bacterial community dynamics in a full-scale enhanced biological phosphorus removal plant. *Water research*, 47(19):7019–7031, 2013.

[64] Andrew C Fowler. Atto-foxes and other minutiae. *Bulletin of Mathematical Biology*, 83(10):104, 2021.

[65] Drew Fudenberg and Christopher Harris. Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57(2):420–441, 1992.

[66] Leonor García-Bayona and Laurie E Comstock. Bacterial antagonism in host-associated microbial communities. *Science*, 361(6408):eaat2456, 2018.

[67] Crispin Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.

[68] Robert A Gatenby, Ariosto S Silva, Robert J Gillies, and B Roy Frieden. Adaptive therapy. *Cancer research*, 69(11):4894–4903, 2009.

[69] Carmen M Gayoso, Jesús Mateos, José A Méndez, Patricia Fernández-Puente, Carlos Rumbo, Maria Tomas, Oskar Martinez de Ilarduya, and Germán Bou. Molecular mechanisms involved in the response to desiccation stress and persistence in acinetobacter baumannii. *Journal of proteome research*, 13(2):460–476, 2014.

[70] Sean M Gibbons, Monika Scholz, Alan L Hutchison, Aaron R Dinner, Jack A Gilbert,

and Maureen L Coleman. Disturbance regimes predictably alter diversity in an ecologically complex bacterial system. *MBio*, 7(6):10–1128, 2016.

[71] Robert J Gillies, Daniel Verduzco, and Robert A Gatenby. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nature Reviews Cancer*, 12(7):487–493, 2012.

[72] Andrea Giometto, David R Nelson, and Andrew W Murray. Antagonism between killer yeast strains as an experimental model for biological nucleation dynamics. *Elife*, 10:e62932, 2021.

[73] Mark Gluzman, Jacob G Scott, and Alexander Vladimirsky. Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory. *Proceedings of the Royal Society B*, 287(1925):20192454, 2020.

[74] David M Gordon and Claire L O'Brien. Bacteriocin diversity and the frequency of multiple bacteriocin production in escherichia coli. *Microbiology*, 152(11):3239–3244, 2006.

[75] Elisa T Granato, Thomas A Meiller-Legrand, and Kevin R Foster. The evolution and ecology of bacterial warfare. *Current biology*, 29(11):R521–R537, 2019.

[76] James M Greene, Jana L Gevertz, and Eduardo D Sontag. Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. *JCO clinical cancer informatics*, 3:1–20, 2019.

[77] Major Greenwood et al. A report on the natural duration of cancer. *A Report on the Natural Duration of Cancer.*, (33), 1926.

[78] Lars Grüne and Willi Semmler. Using dynamic programming with adaptive grid scheme for optimal control problems in economics. *J. Econ. Dyn. Control*, 28(12):2427 – 2456, 2004.

[79] Piyush B Gupta, Christine M Fillmore, Guozhi Jiang, Sagi D Shapira, Kai Tao, Charlotte Kuperwasser, and Eric S Lander. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4):633–644, 2011.

[80] Karl P Hadeler. Stable polymorphisms in a selection model with mutation. *SIAM Journal on Applied Mathematics*, 41(1):1–7, 1981.

[81] Amine Hamdache, Ilias Elmouki, and Smahane Saadi. Optimal control with an isoperi-

metric constraint applied to cancer immunotherapy. *International Journal of Computer Applications*, 94(15), 2014.

[82] A Haslam, MS Kim, and V Prasad. Updated estimates of eligibility for and response to genome-targeted oncology drugs among us cancer patients, 2006-2020. *Annals of Oncology*, 32(7):926–932, 2021.

[83] Simon Heilbronner, Bernhard Krismer, Heike Brötz-Oesterhelt, and Andreas Peschel. The microbiome-shaping roles of bacteriocins. *Nature Reviews Microbiology*, 19(11):726–739, 2021.

[84] Alexandru Hening, Dang H Nguyen, and Peter Chesson. A general theory of coexistence and extinction for stochastic ecological communities. *Journal of Mathematical Biology*, 82(6):56, 2021.

[85] Josef Hofbauer. The selection mutation equation. *Journal of mathematical biology*, 23:41–53, 1985.

[86] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[87] Elyse A Hope, Clara J Amorosi, Aaron W Miller, Kolena Dang, Caiti Smukowski Heil, and Maitreya J Dunham. Experimental evolution reveals favored adaptive routes to cell aggregation in yeast. *Genetics*, 206(2):1153–1167, 2017.

[88] Michael Huston. A general hypothesis of species diversity. *The American Naturalist*, 113(1):81–101, 1979.

[89] Lejla Imamovic and Morten OA Sommer. Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. *Science translational medicine*, 5(204):204ra132–204ra132, 2013.

[90] Shamreen Iram, Emily Dolson, Joshua Chiel, Julia Pelesko, Nikhil Krishnan, Özenç Güngör, Benjamin Kuznets-Speck, Sebastian Deffner, Efe Ilker, Jacob G Scott, et al. Controlling the speed and trajectory of evolution with counterdiabatic driving. *Nature Physics*, 17(1):135–142, 2021.

[91] Yong Dam Jeong, Kwang Su Kim, Yunil Roh, Sooyoun Choi, Shingo Iwami, Il Hyo Jung, and Guang Li. Optimal feedback control of cancer chemotherapy using Hamilton-Jacobi-Bellman equation. *Complexity*, 2022, jan 2022.

[92] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

[93] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[94] Allen A Katouli and Natalia L Komarova. The worst drug rule revisited: mathematical modeling of cyclic cancer treatments. *Bulletin of mathematical biology*, 73:549–584, 2011.

[95] Artem Kaznatcheev, Jeffrey Peacock, David Basanta, Andriy Marusyk, and Jacob G Scott. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nature ecology & evolution*, 3(3):450–456, 2019.

[96] Artem Kaznatcheev, Robert Vander Velde, Jacob G Scott, and David Basanta. Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *British journal of cancer*, 116(6):785–792, 2017.

[97] Benjamin Kerr, Margaret A Riley, Marcus W Feldman, and Brendan JM Bohannan. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature*, 418(6894):171–174, 2002.

[98] IS Khalil, JC Doyle, and K Glover. *Robust and optimal control*. Prentice hall, 1996.

[99] Peter E Kloeden and Eckhard Platen. Stochastic differential equations. In *Numerical Solution of Stochastic Differential Equations*, pages 103–160. Springer, 1992.

[100] Basil Kouvaritakis and Mark Cannon. Model predictive control. *Switzerland: Springer International Publishing*, 38:13–56, 2016.

[101] Ajeet Kumar and Alexander Vladimirsky. An efficient method for multiobjective optimal control and optimal control subject to integral constraints. *Journal of Computational Mathematics*, 28(4):517–551, 2010.

[102] Niraj Kumar, Gwendolyn M Cramer, Seyed Alireza Zamani Dahaj, Bala Sundaram, Jonathan P Celli, and Rahul V Kulkarni. Stochastic modeling of phenotypic switching and chemoresistance in cancer cell populations. *Scientific reports*, 9(1):1–10, 2019.

[103] Teemu Kuosmanen, Johannes Cairns, Robert Noble, Niko Beerenwinkel, Tommi Mononen, and Ville Mustonen. Drug-induced resistance evolution necessitates less aggressive treatment. *PLoS computational biology*, 17(9):e1009418, 2021.

[104] Harold J. Kushner. *Numerical methods for stochastic control problems in continuous time*. Applications of mathematics. Springer, New York, second edition edition, 2001. Hier auch später erschienene, unveränderte Nachdrucke.

[105] Marina V Kuznetsova, Veronika S Mihailovskaya, Natalia B Remezovskaya, and Marjanca Starčič Erjavec. Bacteriocin-producing escherichia coli isolated from the gastrointestinal tract of farm animals: Prevalence, molecular characterization and potential for application. *Microorganisms*, 10(8):1558, 2022.

[106] Russell Lande, Steinar Engen, Bernt-Erik Saether, et al. *Stochastic population dynamics in ecology and conservation*. Oxford University Press on Demand, 2003.

[107] Adam S Lauring and Raul Andino. Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens*, 6(7):e1001005, 2010.

[108] Richard Levins. Coexistence in a variable environment. *The American Naturalist*, 114(6):765–783, 1979.

[109] Feng Lin. *Robust control design: an optimal control approach*. John Wiley & Sons, 2007.

[110] Richard A Long and Farooq Azam. Antagonistic interactions among marine pelagic bacteria. *Applied and environmental microbiology*, 67(11):4975–4983, 2001.

[111] Jeff Maltas and Kevin B Wood. Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLoS biology*, 17(10):e3000515, 2019.

[112] Christopher P Mancuso, Hyunseok Lee, Clare I Abreu, Jeff Gore, and Ahmad S Khalil. Environmental fluctuations reshape an unexpected diversity-disturbance relationship in a microbial community. *Elife*, 10:e67175, 2021.

[113] R.B. Martin, M.E. Fisher, R.F. Minchin, and K.L. Teo. Optimal control of tumor size used to maximize survival time when cells are resistant to chemotherapy. *Mathematical Biosciences*, 110(2):201–219, 1992.

[114] Andriy Marusyk, Doris P Tabassum, Philipp M Altrock, Vanessa Almendro, Franziska Michor, and Kornelia Polyak. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520):54–58, 2014.

[115] Luke McNally, Eryn Bernardy, Jacob Thomas, Arben Kalziqi, Jennifer Pentz, Sam P Brown, Brian K Hammer, Peter J Yunker, and William C Ratcliff. Killing by type

vi secretion drives genetic phase separation and correlates with increased cooperation. *Nature communications*, 8(1):14371, 2017.

[116] Zachary M Miksis and Yong-Tao Zhang. Sparse-grid implementation of fixed-point fast sweeping WENO schemes for eikonal equations. *Communications on Applied Mathematics and Computation*, pages 1–27, 2022.

[117] Cole Miles and Alexander Vladimirsky. Stochastic optimal control of a sailboat. *IEEE Control Systems Letters*, 6:2048–2053, 2021.

[118] Eric L Miller, Morten Kjos, Monica I Abrudan, Ian S Roberts, Jan-Willem Veening, and Daniel E Rozen. Eavesdropping and crosstalk between secreted quorum sensing peptide signals that regulate bacteriocin production in streptococcus pneumoniae. *The ISME journal*, 12(10):2363–2375, 2018.

[119] Tommi Mononen, Teemu Kuosmanen, Johannes Cairns, and Ville Mustonen. Understanding cellular growth strategies via optimal control. *Journal of the Royal Society Interface*, 20(198):20220744, 2023.

[120] Carey D Nadell, Knut Drescher, and Kevin R Foster. Spatial structure, cooperation and competition in biofilms. *Nature Reviews Microbiology*, 14(9):589–600, 2016.

[121] Jen Nguyen, Juanita Lara-Gutiérrez, and Roman Stocker. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS microbiology reviews*, 45(4):fuaa068, 2021.

[122] Daniel Nichol, Peter Jeavons, Alexander G Fletcher, Robert A Bonomo, Philip K Maini, Jerome L Paul, Robert A Gatenby, Alexander RA Anderson, and Jacob G Scott. Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. *PLoS computational biology*, 11(9):e1004493, 2015.

[123] Daniel Nichol, Joseph Rutter, Christopher Bryant, Andrea M Hujer, Sai Lek, Mark D Adams, Peter Jeavons, Alexander RA Anderson, Robert A Bonomo, and Jacob G Scott. Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nature communications*, 10(1):1–10, 2019.

[124] Rene Niehus, Nuno M Oliveira, Aming Li, Alexander G Fletcher, and Kevin R Foster. The evolution of strategy in bacterial warfare via the regulation of bacteriocins and antibiotics. *Elife*, 10:e69756, 2021.

[125] Armita Nourmohammad and Ceyhun Eksin. Optimal evolutionary control for artificial selection on molecular phenotypes. *Physical Review X*, 11(1):011044, 2021.

[126] A. Nowakowski and A. Popa. A dynamic programming approach for approximate optimal control for cancer therapy. *J. Optim. Theory Appl.*, 156(2):365–379, feb 2013.

[127] Bernt Øksendal and Agnes Sulem. Stochastic control of jump diffusions. In *Applied Stochastic Control of Jump Diffusions*, pages 93–155. Springer, 2019.

[128] Rocío-Anaís Pérez-Gutiérrez, Varinia López-Ramírez, Africa Islas, Luis David Alcaraz, Ismael Hernández-González, Beatriz Carely Luna Olivera, Moisés Santillán, Luis E Eguiarte, Valeria Souza, Michael Travisano, et al. Antagonism influences assembly of a bacillus guild in a local community and is depicted as a food-chain network. *The ISME Journal*, 7(3):487–497, 2013.

[129] Raquel Perruca-Foncillas, Johan Davidsson, Magnus Carlquist, and Marie F Gorwa-Grauslund. Assessment of fluorescent protein candidates for multi-color flow cytometry analysis of saccharomyces cerevisiae. *Biotechnology Reports*, 34:e00735, 2022.

[130] S Brook Peterson, Savannah K Bertolli, and Joseph D Mougous. The central role of interbacterial antagonism in bacterial life. *Current Biology*, 30(19):R1203–R1214, 2020.

[131] Olga E Petrova and Karin Sauer. Sticky situations: key components that control bacterial surface attachment. *Journal of bacteriology*, 194(10):2413–2425, 2012.

[132] A. B. Philpott, S. G. Henderson, and D. Teirney. A simulation model for predicting yacht match race outcomes. *Operations Research*, 52(1):1–16, February 2004.

[133] Magdalena D Pieczynska, Dominika Wloch-Salamon, Ryszard Korona, and J Arjan GM de Visser. Rapid multiple-level coevolution in experimental populations of yeast killer and nonkiller strains. *Evolution*, 70(6):1342–1353, 2016.

[134] Jason Pintar and William T Starmer. The costs and benefits of killer toxin production by the yeast pichia kluyveri. *Antonie van Leeuwenhoek*, 83:89–97, 2003.

[135] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko. *The mathematical theory of optimal processes*. John Wiley & Sons, Inc., New York, 1962.

[136] Dongping Qi, Adam Dhillon, and Alexander Vladimirsky. Optimality and robustness in path-planning under initial uncertainty. *arXiv preprint arXiv:2106.11405*, 2021.

[137] Jianliang Qian, Yong-Tao Zhang, and Hong-Kai Zhao. Fast Sweeping Methods for Eikonal Equations on Triangular Meshes. *SIAM Journal on Numerical Analysis*, 45(1):83–107, 2007.

[138] David A Relman. The human microbiome: ecosystem resilience and health. *Nutrition reviews*, 70(suppl_1):S2–S9, 2012.

[139] Margaret A Riley and John E Wertz. Bacteriocins: evolution, ecology, and application. *Annual Reviews in Microbiology*, 56(1):117–137, 2002.

[140] Emily SC Rittershaus, Seung-Hun Baek, and Christopher M Sassetti. The normalcy of dormancy: common themes in microbial quiescence. *Cell host & microbe*, 13(6):643–651, 2013.

[141] John W Ruge and Klaus Stüben. Algebraic multigrid. In *Multigrid methods*, pages 73–130. SIAM, 1987.

[142] Jakob Russel, Henriette L Røder, Jonas S Madsen, Mette Burmølle, and Søren J Sørensen. Antagonism correlates with metabolic similarity in diverse bacteria. *Proceedings of the National Academy of Sciences*, 114(40):10684–10688, 2017.

[143] Nazanin Saeidi, Choon Kit Wong, Tat-Ming Lo, Hung Xuan Nguyen, Hua Ling, Susanna Su Jan Leong, Chueh Loo Poh, and Matthew Wook Chang. Engineering microbes to sense and eradicate pseudomonas aeruginosa, a human pathogen. *Molecular systems biology*, 7(1):521, 2011.

[144] Maria Santagati, Marina Scillato, Francesco Patane, Caterina Aiello, and Stefania Stefani. Bacteriocin-producing oral streptococci and inhibition of respiratory pathogens. *FEMS Immunology & Medical Microbiology*, 65(1):23–31, 2012.

[145] Heinz Schättler and Urszula Ledzewicz. *Optimal Control for Mathematical Models of Cancer Therapies*, volume 42 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY, 2015.

[146] Severin J Schink, Elena Biselli, Constantin Ammar, and Ulrich Gerland. Death rate of e. coli during starvation is set by maintenance cost and biomass recycling. *Cell systems*, 9(1):64–73, 2019.

[147] Ruth Schmidt, Viviane Cordovez, Wietse De Boer, Jos Raaijmakers, and Paolina Garbeva. Volatile affairs in microbial interactions. *The ISME journal*, 9(11):2329–2335, 2015.

[148] Sijmen E Schoustra, Jonathan Dench, Rola Dali, Shawn D Aaron, and Rees Kassen. Antagonistic interactions peak at intermediate genetic distance in clinical and laboratory strains of pseudomonas aeruginosa. *Bmc Microbiology*, 12:1–9, 2012.

[149] Sebastian J Schreiber, Michel Benaïm, and Kolawolé AS Atchadé. Persistence in fluctuating environments. *Journal of Mathematical Biology*, 62:655–683, 2011.

[150] Orr H Shapiro, Ariel Kushmaro, and Asher Brenner. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *The ISME journal*, 4(3):327–336, 2010.

[151] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[152] Chi-Wang Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced numerical approximation of nonlinear hyperbolic equations*, pages 325–432. Springer, 1998.

[153] Kamilla S Sjøgaard, Thomas B Valdemarsen, and Alexander H Treusch. Responses of an agricultural soil microbiome to flooding with seawater after managed coastal realignment. *Microorganisms*, 6(1):12, 2018.

[154] Tami L Swenson, Ulas Karaoz, Joel M Swenson, Benjamin P Bowen, and Trent R Northen. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nature communications*, 9(1):19, 2018.

[155] John R Tagg, Liam K Harold, Rohit Jain, and John DF Hale. Beneficial modulation of human health in the oral cavity and beyond using bacteriocin-like inhibitory substance-producing streptococcal probiotics. *Frontiers in Microbiology*, 14:1161155, 2023.

[156] Peter D Taylor and Leo B Jonker. Evolutionary stable strategies and game dynamics. *Mathematical biosciences*, 40(1-2):145–156, 1978.

[157] Christoph A Thaiss, Maayan Levy, Tal Korem, Lenka Dohnalová, Hagit Shapiro, Diego A Jaitin, Eyal David, Deborah R Winter, Meital Gury-BenAri, Evgeny Tatirovsky, et al. Microbiota diurnal rhythmicity programs host transcriptome oscillations. *Cell*, 167(6):1495–1510, 2016.

[158] TF Thingstad, G Bratbak, and M Heldal. Aquatic phage ecology. *Bacteriophage ecology*, pages 251–280, 2008.

[159] Thomas J Travers-Cook, Jukka Jokela, and Claudia C Buser. The evolutionary ecology of fungal killer phenotypes. *Proceedings of the Royal Society B*, 290(2005):20231108, 2023.

[160] Racheal N Upton, Elizabeth M Bach, and Kirsten S Hofmockel. Spatio-temporal microbial community dynamics within soil aggregates. *Soil Biology and Biochemistry*, 132:58–68, 2019.

[161] Marc W Van Goethem, Tami L Swenson, Gareth Trubl, Simon Roux, and Trent R Northen. Characteristics of wetting-induced bacteriophage blooms in biological soil crust. *MBio*, 10(6):10–1128, 2019.

[162] WF Van Zyl, SM Deane, and LMT Dicks. Bacteriocin production and adhesion properties as mechanisms for the anti-listerial activity of lactobacillus plantarum 423 and enterococcus mundtii st4sa. *Beneficial microbes*, 10(3):329–349, 2019.

[163] Robert Vander Velde, Nara Yoon, Viktoriya Marusyk, Arda Durmaz, Andrew Dhawan, Daria Miroshnychenko, Diego Lozano-Peral, Bina Desai, Olena Balynska, Jan Poleszhuk, et al. Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. *Nature communications*, 11(1):1–13, 2020.

[164] Caroline Vincent, Mark A Miller, Thaddeus J Edens, Sudeep Mehrotra, Ken Dewar, and Amee R Manges. Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and clostridium difficile colonization or infection. *Microbiome*, 4:1–11, 2016.

[165] Flora Vincent, Matti Gralka, Guy Schleyer, Daniella Schatz, Miguel Cabrera-Brufau, Constanze Kuhlisch, Andreas Sichert, Silvia Vidal-Melgosa, Kyle Mayers, Noa Barak-Gavish, et al. Viral infection switches the balance between bacterial and eukaryotic recyclers of organic matter during coccolithophore blooms. *Nature communications*, 14(1):510, 2023.

[166] Laura Vinckenbosch. Stochastic Control and Free Boundary Problems for Sailboat Trajectory Optimization. 2012. Publisher: Lausanne, EPFL.

[167] A. Vladimirsky and C. Zheng. A fast implicit method for time-dependent Hamilton-Jacobi PDEs. *preprint:* `https://arxiv.org/abs/2008.00555`.

[168] MingYi Wang, Natasha Patnaik, Anne Somalwar, Jingyi Wu, and Alexander Vladimirsky. Risk-aware stochastic control of a sailboat. *ACC-2024; preprint:* `https://arxiv.org/abs/2309.13436`.

[169] MingYi Wang, Jacob G. Scott, and Alexander Vladimirsky. Threshold-awareness in adaptive cancer therapy. *PLOS Computational Biology*, 20(6):e1012165, June 2024.

[170] Yuheng Wang and Margaret P. Chapman. Risk-averse autonomous systems: A brief history and recent developments from the perspective of optimal control. *Artificial Intelligence*, 311:103743, 2022.

[171] Zhijun Wang, Li Xiang, Junjie Shao, Alicja Węgrzyn, and Grzegorz Węgrzyn. Effects of the presence of cole1 plasmid dna in escherichia coli on the host cell metabolism. *Microbial Cell Factories*, 5:1–18, 2006.

[172] Anna S Weiß, Alexandra Götz, and Madeleine Opitz. Dynamics of colicine2 production and release determine the competitive success of a toxin-producing bacterial population. *Scientific reports*, 10(1):4052, 2020.

[173] Jeffrey West, Li You, Jingsong Zhang, Robert A Gatenby, Joel S Brown, Paul K Newton, and Alexander RA Anderson. Towards multidrug adaptive therapy. *Cancer research*, 80(7):1578–1589, 2020.

[174] Jeffrey B West, Mina N Dinh, Joel S Brown, Jingsong Zhang, Alexander R Anderson, and Robert A Gatenby. Multidrug cancer therapy in metastatic castrate-resistant prostate cancer: an evolution-based strategy. *Clinical Cancer Research*, 25(14):4413–4421, 2019.

[175] Claus O Wilke. Quasispecies theory in the context of population genetics. *BMC evolutionary biology*, 5(1):1–8, 2005.

[176] Paul Zarchan. *Progress in astronautics and aeronautics: fundamentals of Kalman filtering: a practical approach*, volume 208. Aiaa, 2005.

[177] Jingsong Zhang, Jessica J Cunningham, Joel S Brown, and Robert A Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature communications*, 8(1):1–9, 2017.

[178] Xuanxi Zhang, Jihao Long, Wei Hu, Weinan E, and Jiequn Han. Initial value problem enhanced sampling for closed-loop optimal control design with deep neural networks. *preprint:* `https://arxiv.org/abs/2209.04078`.

[179] Boyang Zhao, Joseph C Sedlak, Raja Srinivas, Pau Creixell, Justin R Pritchard, Bruce Tidor, Douglas A Lauffenburger, and Michael T Hemann. Exploiting temporal collateral sensitivity in tumor clonal evolution. *Cell*, 165(1):234–246, 2016.

[180] Hongkai Zhao. A fast sweeping method for Eikonal equations. *Mathematics of Computation*, 74(250):603–628, 2004.

[181] Jinshui Zheng, Michael G Gänzle, Xiaoxi B Lin, Lifang Ruan, and Ming Sun. Diversity and dynamics of bacteriocins from human microbiome. *Environmental microbiology*, 17(6):2133–2143, 2015.

[182] Samira Zouhri, Mohcine El Baroudi, and Smahane Saadi. Optimal control with isoperimetric constraint for chemotherapy of tumors. *International Journal of Applied and Computational Mathematics*, 8(4):215, 2022.