Threshold-awareness in adaptive cancer therapy

MingYi Wang[1], Jacob G. Scott[2], and Alexander Vladimirsky[*,3]

[1]Center for Applied Mathematics, Cornell University, Ithaca, NY

[2]Dept. of Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH

[3]Dept. of Mathematics and Center for Applied Mathematics, Cornell University, Ithaca, NY

Abstract

Although adaptive cancer therapy shows promise in integrating evolutionary dynamics into treatment scheduling, the stochastic nature of cancer evolution has seldom been taken into account. Various sources of random perturbations can impact the evolution of heterogeneous tumors, making performance metrics of any treatment policy random as well. In this paper, we propose an efficient method for selecting optimal adaptive treatment policies under randomly evolving tumor dynamics. The goal is to improve the cumulative "cost" of treatment, a combination of the total amount of drugs used and the total treatment time. As this cost also becomes random in any stochastic setting, we maximize the probability of reaching the treatment goals (tumor stabilization or eradication) without exceeding a pre-specified threshold (or a "budget"). We use a novel Stochastic Optimal Control formulation and Dynamic Programming to find such "threshold-aware" optimal treatment policies. Our approach enables an efficient algorithm to compute these policies for a range of threshold values simultaneously. Compared to treatment plans shown to be optimal in a deterministic setting, the new "threshold-aware" policies significantly improve the chances of the therapy succeeding under the budget, which is correlated with a lower general drug usage. We illustrate this method using two specific examples, but our approach is far more general and provides a new tool for optimizing adaptive therapies based on a broad range of stochastic cancer models.

Keywords: adaptive therapy, tumor heterogeneity, evolutionary game theory, environmental stochasticity, stochastic optimal control, threshold-aware optimal treatment policy

Author Summary: Tumor heterogeneities provide an opportunity to improve therapies by leveraging complex (often competitive) interactions of different types of cancer cells. These interactions are usually stochastic due to both individual cell differences and random events affecting the patient as a whole. The new generation of cancer models strive to account for this inherent stochasticity, and *adaptive* treatment plans need to reflect it as well. In optimizing such treatment, the most common approach is to maximize the probability of eventually stabilizing or eradicating the tumor. In this paper, we consider a more nuanced version of success, maximizing the

[*]Corresponding author: Alexander Vladimirsky, 561 Malott Hall, Dept. of Mathematics, Cornell University, Ithaca, NY 14853-4201. Telephone: (607) 255-9871. Email: vladimirsky@cornell.edu

probability of reaching these therapy goals before the cumulative burden from the disease and treatment exceed a chosen threshold. Importantly, our method allows computing such optimal treatment plans efficiently and for a range of thresholds at once. If used on a high-fidelity personalized model, 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 (e.g., to include the trade-offs between toxicity and quality of life).

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 [55]. Most researchers using this perspective focus on deterministic models of tumor evolution, with typical optimization objectives of maximizing the survival time [46], minimizing the tumor size [6], or minimizing the time until the tumor size is stabilized [7]. 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., [10, 13, 37]) or minimize the likelihood of negative events (e.g., metastasis) after specified time [19]. 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 paper 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 [30]. 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 [24]. 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 [34, 45, 50, 51] as well as for a number of cancers [14, 65]. In the study of therapeutic scheduling, adaptive therapy, which uses mathematical tools from Evolutionary Game Theory (EGT), has shown promise not only in theory [23], but also in a phase 2 trial for men with metastatic prostate cancer [63]. Experimentally, there have been confirmations of EGT principles *in vivo*[17] as well as more quantitatively focused assay development *in vitro* [38], and observations of game interactions using these methods [18]. There are also many other models capturing

the competition within heterogeneous tumors without using game-theoretic derivations; e.g., [6, 9, 26, 42]. The majority of theoretical work in this space has focused on optimization of different drug regimens for *deterministic* models of cancer evolution [11, 12, 25, 60, 61]. 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 [27], varying reactions to drugs [41], differences in signalling, and small-scale variations in the tumor environment. Many of these are instances of *demographic stochasticity* [43], 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 [44, 62] and exist as distributions around a central genotype, with all cells in the group behaving in a similar manner despite random birth/death events [16, 43] and small within-the-group genetic heterogeneities [35]. 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 [16, 43]; 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) [2, 4], whose behavior can be optimized using Stochastic Optimal Control Theory [20]. 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 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 and §2.3.

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 a function $\boldsymbol{f}$, which 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). But the same framework can be also used for multiple drugs, with a separate upper bound specified for each element of $\boldsymbol{d}(t)$. Given an initial tumor state $\boldsymbol{x}(0) = \boldsymbol{\xi}$, a successful therapy aims to drive the tumor state to a set Δ_{succ} while ensuring that the set Δ_{fail} is avoided[I]. If the therapy accomplishes this, its overall "cost" $\mathcal{J}$ is assessed by integrating some running cost K 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 changing 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

[I] E.g., 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 §2.3. 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.

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: $$T(\boldsymbol{\xi}, \boldsymbol{d}(\cdot)) = \inf \Big\{ t \in \mathbb{R}_+ \ \mid \ \boldsymbol{x}(t) \in \Delta, \ \boldsymbol{x}(0) = \boldsymbol{\xi} \Big\}. \tag{2.2}$$

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

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 $$\min_{\boldsymbol{d} \in \mathcal{D}} \Big\{ K(\boldsymbol{\xi}, \boldsymbol{d}) + \nabla u(\boldsymbol{\xi}) \cdot \boldsymbol{f}(\boldsymbol{\xi}, \boldsymbol{d}) \Big\} = 0, \tag{2.4}$$

with the boundary condition $u = g$ on Δ. See SM §3S.1 for the derivation and §4S.2 for the numerics.

dynamic programming [3] 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. 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

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

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

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 [21] :

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

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 $\Delta = \Delta_{\mathrm{succ}} \cup \Delta_{\mathrm{fail}}$.

Note: If one uses $g(\boldsymbol{X}(T)) = +\infty$ when therapy fails (i.e., when $\boldsymbol{X}(T) \in \Delta_{\mathrm{fail}}$), 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: $$\tilde{w}(\boldsymbol{\xi}) = \sup_{\boldsymbol{d}(\cdot)} \mathbb{P}\Big(\boldsymbol{X}(T) \in \Delta_{\mathrm{succ}}\Big) \tag{2.8}$$

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}.

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 $\boldsymbol{d}_*(\boldsymbol{\xi})$ that minimizes the expected treatment cost $\mathbb{E}\big[\mathcal{J}\big]$. As explained in Box 3, the resulting 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 $\mathbb{P}\left(\mathcal{J} \le \bar{s}\right)$.

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[II] (2.11) numerically, and the optimal feedback policy $\boldsymbol{d}_*^{\bar{s}}(\boldsymbol{\xi}, s)$ is recovered in the process[III] for all $\bar{s} \in (0, \bar{\mathcal{S}}]$.

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 §3S.2 of Supplementary Materials.

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

While this threshold-aware framework has important advantages illustrated below, it also brings to the forefront

[II] 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 [5, 19, 52] 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 $s(t)$.

[III] The details of our numerical method are included in Supplementary Materials (SM) §4S.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 §3 at `https://github.com/eikonal-equation/Stochastic-Cancer`

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 ($s(t) = 0$) under any treatment policy. 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 ($\boldsymbol{X}(\tau) \notin \Delta_{\text{fail}},\ \forall \tau \in [0, t]$) 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) [54], 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 [66] or imunotherapy [29]. 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 [40]. 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 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 [39], which 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 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 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 Supplementary Materials (SM) §1S.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" for all cancer cells since it is damaging for the surrounding non-cancerous tissue. The base model in [39] 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) [33, 57] 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 [39], 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 [39] for the derivation. The goal of the drug therapy here is to drive the GLY fraction $p(t) = x_{\mathrm{G}}(t)$ down[IV] below a specified "stabilization barrier" γ_{r}. 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 [39]. 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 Figure 2 in [39].

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 [25] using the framework of deterministic optimal control theory [20]. A time-dependent intensity of the therapy $d(t)$ (ranging from 0 to the MTD level

[IV] In [39], 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.

Box 5: Example 1 (base model adopted from [25, 39])

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

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

The above reflects the formulas for subpopulation fitnesses $(\psi_{\mathrm{G}}, \psi_{\mathrm{D}}, \psi_{\mathrm{V}})$; see details in SM § 1S.1.

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

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

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

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

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

$d_{\max}$) was chosen to minimize the overall cost of treatment $\mathcal{J}\left(q_0, p_0, d(\cdot)\right) = \int\limits_0^T d(t)\, ds + \delta T + g\left(q(T), p(T)\right),$

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 [3] summarized in Box 1, this corresponds to minimizing the integral of the running cost $K = d(t)+\delta$. In [25], the deterministic-optimal policy is obtained in *feedback form* (i.e., $d = d_\star(q, p)$) by numerically solving the Hamilton-Jacobi-Bellman (HJB) PDE (2.4). Figure 1(a) summarizes this policy (showing in yellow the MTD region where $d_\star(q, p) = d_{\max}$) and illustrates the corresponding trajectory for one specific initial (q_0, p_0).

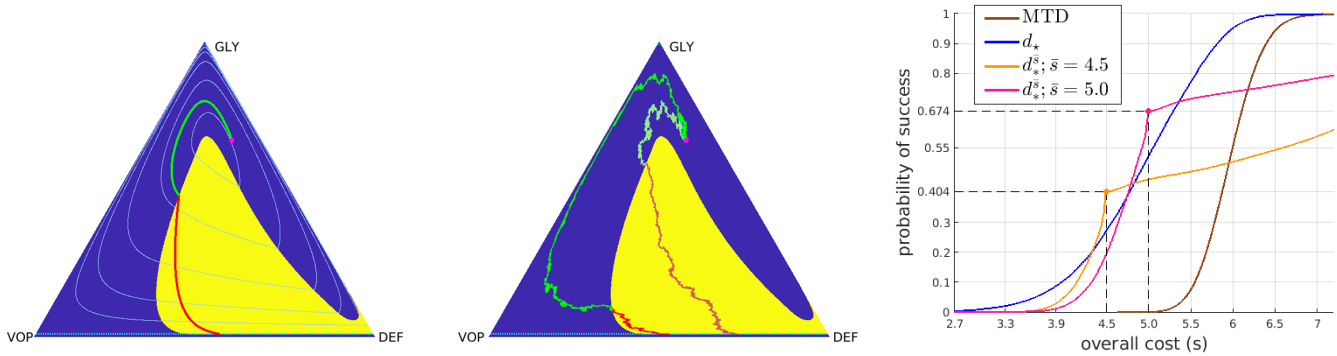


Figure 1: Deterministic-optimal policy in the EGT-model. The (GLY-VOP-DEF) triangle represents all possible relative abundances of respective subpopulations. Since the policy is bang-bang, we show it by using the yellow background where drugs should be used at the MTD rate and the blue background where 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 has a cost of 3.33 while the other has a cost of 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.

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 $(\psi_{\text{G}}, \psi_{\text{D}}, \psi_{\text{V}})$ encoding the expected values of those rates and the scale of random perturbations specified by $(\sigma_{\text{G}}, \sigma_{\text{D}}, \sigma_{\text{V}})$. This approach, originating from Fudenberg and Harris paper [22], is suitable for modeling heterogeneous tumors, in which subpopulations not only interact [47] but can also vary in their growth rates over time [58]. Adopting the usual probabilistic notation of using capital letters for random variables, we can again start with the subpopulation sizes $(Z_{\text{G}}, Z_{\text{D}}, Z_{\text{V}})$ 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 the fitness of the respective subpopulation. This can be used to derive the SDEs for the corresponding fractions[V] $(X_{\text{G}}, X_{\text{D}}, X_{\text{V}})$; see the summary in Box 5 for the reduced (Q, P) coordinates and

[V] We note that Related 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 [31, 56].

the derivation in SM §2S.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 $d(\cdot)$. Figure 1(b) shows one example of using the deterministic-optimal policy $d(t) = d_\star\left(Q(t), P(t)\right)$ 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:

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

whose graph is shown in blue in Figure 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 Figure 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 §2.1 and summarized in Box 4, this is accomplished for a range of threshold values and all initial cancer configurations simultaneously. Figure 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 §3.1.

2.3 Example 2: a Sensitive-Resistant competition model.

We also illustrate our approach by extending a model proposed by Carrère [6], 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". A series of *in vitro* experiments were conducted by Manon Carré at the CRO2[VI] where only the above two types of cells were present. 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 by the time t is $\dfrac{z_{\text{S}}(t) + m z_{\text{R}}(t)}{C}$. 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 [6]. To model this competitive advantage, they have used an additional competition term $-\beta z_{\text{S}} z_{\text{R}}$ to describe the rate of change of $z_{\text{R}}(t)$, 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

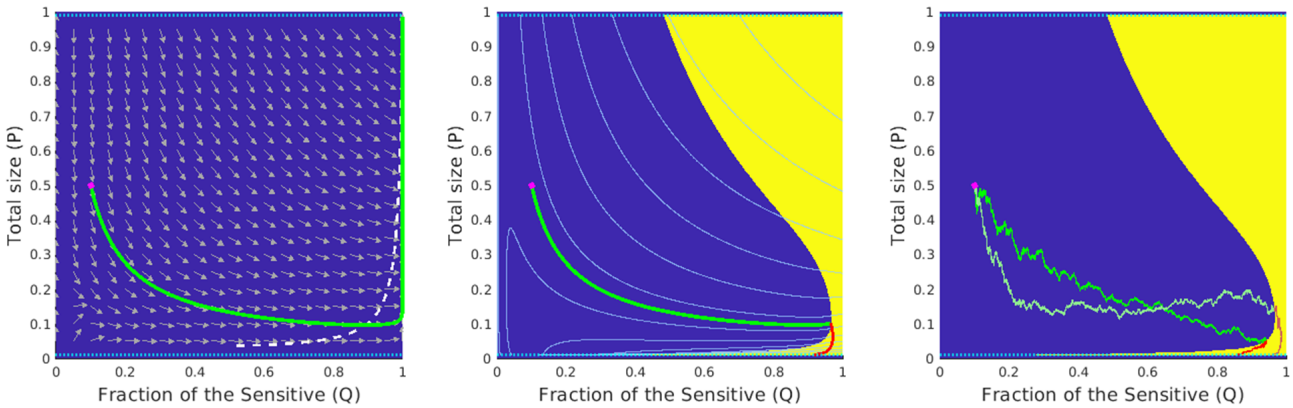
[VI] Center for Research in Oncobiology and Oncopharmacology, Aix-Marseille Université

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 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 $q(t) = z_{\mathrm{S}}(t)/p(t)$. This change of coordinates yields an ODE model (2.19) summarized in Box 6; see SM §1S.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 $\gamma_{\mathrm{f}} < 1$.

Figure 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 ($p(t) > \gamma_{\mathrm{f}}$). The 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). Figure 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.



(a) (b) (c)

Figure 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 Δ_{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 a stochastic perturbation in $(g_{\mathrm{S}}, g_{\mathrm{R}})$ (the brighter one has a cost of 70.45 while the other has a cost of 49.43); 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

Box 6: Example 2 (base model adopted from [6])

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

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

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

$$\begin{aligned} \boldsymbol{X} = \begin{bmatrix} Q \\ P \end{bmatrix} &:= \begin{bmatrix} \dfrac{Z_{\mathrm{S}}}{Z_{\mathrm{S}} + m Z_{\mathrm{R}}} \\ Z_{\mathrm{S}} + m Z_{\mathrm{R}} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} B_t \end{bmatrix}, \\ \boldsymbol{a}(\boldsymbol{X}, d) &= \begin{bmatrix} Q(1-Q)\Big\{(1-P)(g_{\mathrm{S}} - g_{\mathrm{R}}) - \alpha d + \beta C Q P + (1-P)^2[\sigma_{\mathrm{R}}^2(1-Q) - \sigma_{\mathrm{S}}^2 Q + \sigma_{\mathrm{S}}\sigma_{\mathrm{R}}]\Big\} \\ P(1-P)(g_{\mathrm{S}} Q + g_{\mathrm{R}}(1-Q)) - \alpha Q P d - \beta C P^2 Q(1-Q) \end{bmatrix}, \\ \boldsymbol{\Sigma}(\boldsymbol{X}, d) &= \begin{bmatrix} (1-P)Q(1-Q)(\sigma_{\mathrm{S}} - \sigma_{\mathrm{R}}) \\ P(1-P)[\sigma_{\mathrm{S}} Q + \sigma_{\mathrm{R}}(1-Q)] \end{bmatrix}. \end{aligned} \tag{2.20}$$

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 Supplementary Materials §5S.2.

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

actually random and normally distributed at any instant[VII]. 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

[VII] This approach was also used in modeling persistence strategies among bacteria in [5].

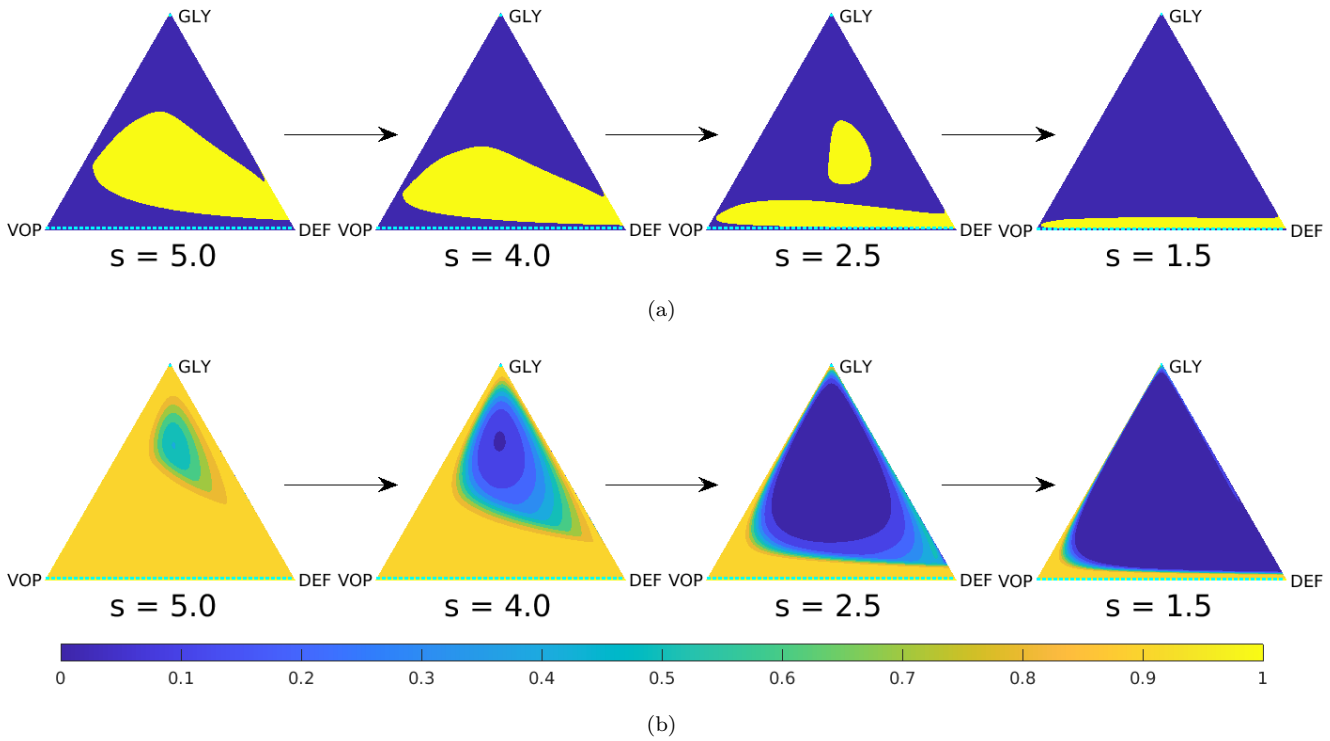
three-dimensional.) Depending on the nature of perturbations, a similar assumption might be reasonable in the current example as well. But this is not a necessary feature for the threshold-aware optimization approach to be applicable. To demonstrate this, we will instead assume in what follows that a single (1D) Brownian motion perturbs the intrinsic growth rates of both S and R, with $(g_{\mathrm{S}}, g_{\mathrm{R}})$ still representing the expected rates and $(\sigma_{\mathrm{S}}, \sigma_{\mathrm{R}})$ denoting their respective volatilities. This yields SDEs for the stochastic evolution of (Q, P), which are derived in §2S.3 of SM and summarized in Box 6. As shown in Figure 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 §3.2.

3 Results

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. 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. [39] and Gluzman et al. [25]. 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 [25]. 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 SM §4S.3. Additional examples, including those with higher volatilities, in which the threshold-performance advantages are even more significant, can be found in SM §5S.

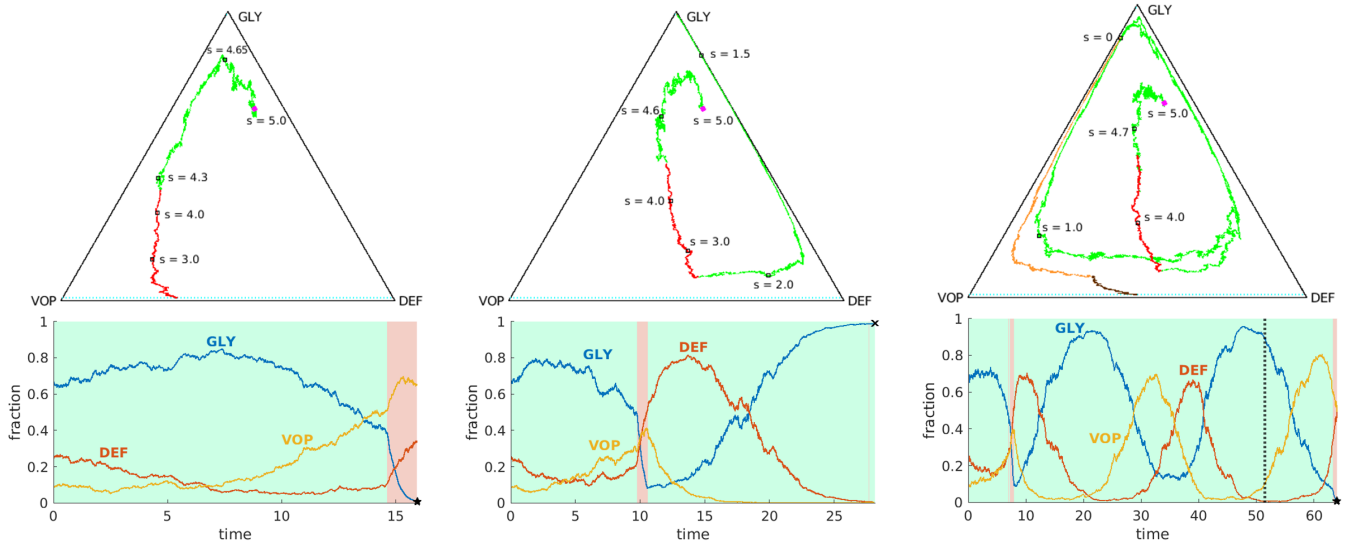
In Figure 3, we present some representative s-slices of threshold-aware optimal policies and their corresponding optimal probability of success for respective threshold values. To be consistent with [25], in all of our figures, the drugs-on region (at the MTD level) is shown in yellow and the drugs-off region is shown in blue. We observe that this drugs-on region is strongly s-dependent and completely different from the one in the deterministic-optimal case shown in Figure 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 Figure 4, starting from the same initial tumor configuration $(q_0, p_0) = (0.26, 0.665)$ already used in Figure 1 and focusing on a threshold $\bar{s} = 5$. We use the example from Figure 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 Figure 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 Figure 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



(a)

(b)

Figure 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.



(a) (b) (c)

Figure 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.

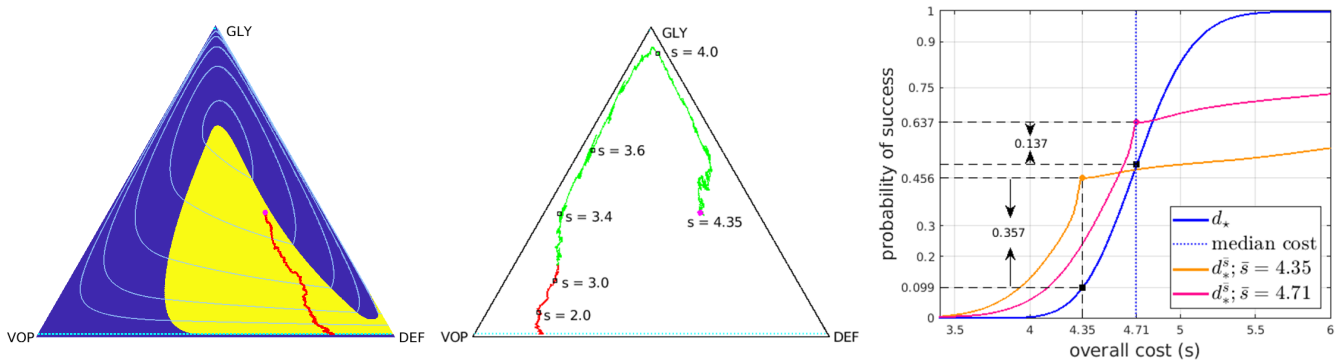
on d_* (as s changes continuously while Figure 3(a) presents just a few representative slices[VIII]) taking into account the changing tumor configuration $(Q(t), P(t))$.

In contrast to the success story in Figure 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 Figure 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 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 Figure 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 $d_\bigstar$ illustrated in Figure 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 $d_\bigstar$ typically yields smaller costs while only slightly increasing the chances of eventual failure (e.g., $\approx 0.45\%$ of crossings into Δ_{fail} using $d_\bigstar$ versus $\approx 0.1\%$ using the full MTD once the original budget of $\bar{s} = 4.71$ 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 Figure 3(b), we observe that v has particularly large gradient near the level curves of the deterministic-optimal value function u shown in Figure 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 Figure 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 $d_\bigstar = d_{\max}$ while $d_*^{\bar{s}} = 0$ for a range of $\bar{s}$ values; see Figures 5(a) and (b) for representative paths and 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 $\mathbb{P}(\mathcal{J}(d_\bigstar) \leq 4.35) < 10\%$, 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 Figure 5(a). In contrast, our threshold-aware policies tend not to prescribe drugs until GLY is relatively low and VOP is relatively high; see Figure 5(b). As a result, the patient would suffer less toxicity from drugs in most scenarios.

[VIII] Movies with additional information for Figures 3 and 4 are available at `https://eikonal-equation.github.io/Stochastic-Cancer/examples.html`



(a) (b) (c)

Figure 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 Figure 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 Figure 3(a).

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

We now turn to the SR model system described in §2.3. Our numerical experiments use $d_{\max} = 3$, $\gamma_{\mathrm{r}} = 1 - \gamma_{\mathrm{f}} = 10^{-2}$, $\delta = 0.05$, and volatilities $\sigma_{\mathrm{R}} = \sigma_{\mathrm{S}} = 0.15$. For other parameter values, see §5S.2 of SM.

We show the representative s-slices of threshold aware policies and the corresponding success probabilities in Figure 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 Figure 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 Δ_{succ} (where a short burst of drugs is likely sufficient).

In Figure 7, we consider a particular initial tumor configuration $(q_0, p_0) = (0.45, 0.9)$ and compare the performance of three different policies. Despite the fact that all three use no drugs at the very beginning, the deterministic-optimal policy typically starts prescribing drugs much earlier. See the comparison of sample tra-

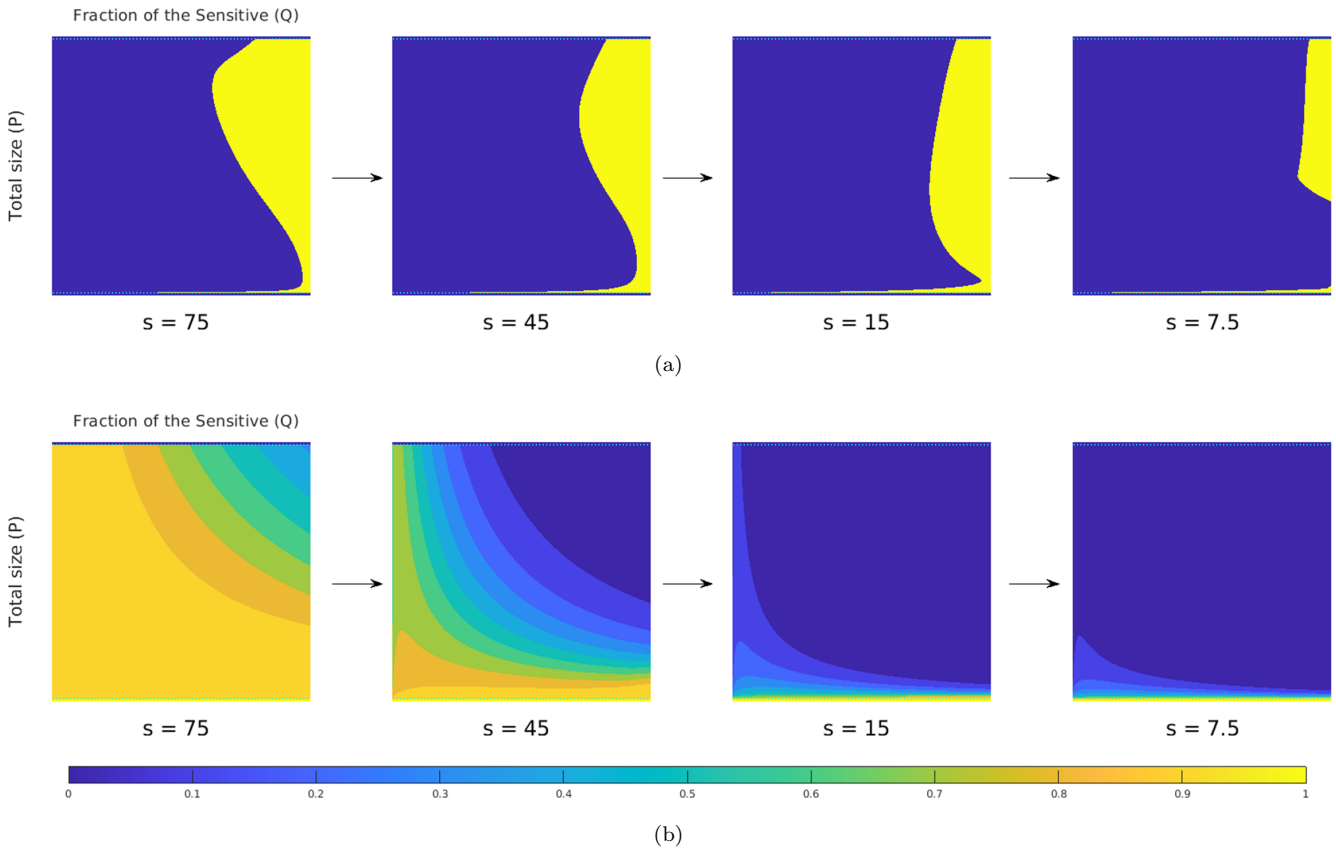


Figure 6: Representative slices of the threshold-aware optimal policy (top row) and the corresponding probability of success (bottom row) for the Carrère example. Each square represents all possible tumor states (sizes and compositions). The horizontal axis is the *fraction of the Sensitive (Q)* and 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.

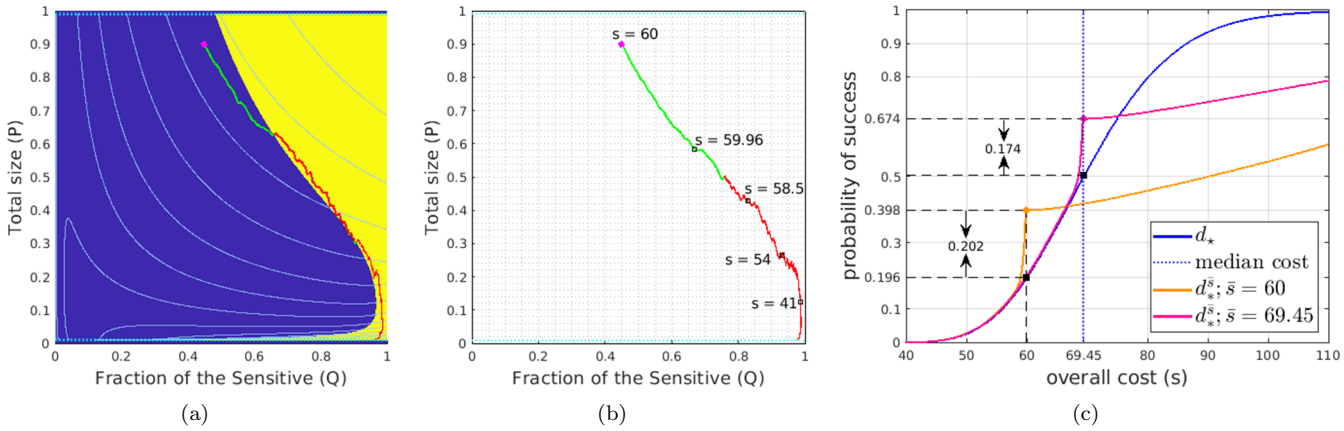


Figure 7: Comparison between threshold-aware policies and the deterministic-optimal policy. Starting from an initial state $(q_0, p_0) = (0.45, 0.9)$ (magenta dot): (a) a sample path with cost 57.3 under the deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 60$ with a total cost of 53.7 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 $\bar{s} = 69.45$.

jectories under $d_\bigstar$ and $d_*^{\bar{s}}$ in Figures 7(a) and 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 $d_\bigstar$. This advantage is even more significant with lower thresholds. E.g., $\mathbb{P}(\mathcal{J}(d_\bigstar) \leq 60)$ is only 19.6%, while our threshold-aware policy (CDF shown in orange) more than doubles this probability of under-threshold remission to 39.8%.

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*[38] and *in vivo*[17, 47], 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 [55], 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 [1, 7, 19, 25, 36, 42, 53]. 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 [48], tensor decompositions [15], or deep neural networks [64].

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 [28, 32]. We did not pursue such examples here primarily to make for an easier comparison with prior work [25] 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 [5, 22]. 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 becomes 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[IX], including jump-diffusion processes, which could be used to build future models that account for demographic stochasticity in these special small-subpopulation regimes. Dynamic programming is also used in discrete population models focused on demographic stochasticity [49]. 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 [8].

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 [19]. 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 [38]. 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.

Funding and Acknowledgements The authors are grateful to Roberto Ferretti and Lars Grüne for their advice on some aspects of numerical methods used in this project. The authors would also like to thank the anonymous reviewers for their very insightful and helpful comments. MW and AV are grateful to the National

[IX] A similar threshold-aware method has been developed in [8] for controlling "piecewise-deterministic" processes, where perturbations happen at discrete points in time and amount to abrupt switches in system dynamics. More recently, it has been 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 $\bar{s}$ [59].

Science Foundation (DMS-1645643, DMS-1738010, and DMS-2111522) for supporting this project. JGS would like to thank the National Institutes of Health (5R37CA244613, 5U54CA274513, U01CA280829) and the American Cancer Society (RSG-20-096-01) for their generous support.

References

[1] Alamir, M. (2014). Robust feedback design for combined therapy of cancer. *Optimal Control Applications and Methods*, 35(1):77–88.

[2] Allen, E. (2007). *Modeling with Itô stochastic differential equations*, volume 22. Springer Science & Business Media.

[3] Bardi, M. & Dolcetta, I. (1997). *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhäuser.

[4] Braumann, C. A. (2010). Environmental versus demographic stochasticity in population growth. In: *Workshop on Branching Processes and Their Applications*, pages 37–52. Springer.

[5] Browning, A. P., Sharp, J. A., Mapder, T., Baker, C. M., Burrage, K., & Simpson, M. J. (2021). Persistence as an optimal hedging strategy. *Biophysical Journal*, 120(1):133–142.

[6] Carrère, C. (2017). Optimization of an in vitro chemotherapy to avoid resistant tumours. *Journal of Theoretical Biology*, 413:24–33.

[7] Carrère, C. & Zidani, H. (2020). Stability and reachability analysis for a controlled heterogeneous population of cells. *Optimal Control Applications and Methods*, 41(5):1678–1704.

[8] Cartee, E., Nellis, A., Van Hook, J., Farah, A., & Vladimirsky, A. (2023). Quantifying and managing uncertainty in piecewise-deterministic Markov processes. *SIAM/ASA Journal on Uncertainty Quantification*, 11(3):814–847.

[9] Chisholm, R. H., Lorenzi, T., & Clairambault, J. (2016). Cell population heterogeneity and evolution towards drug resistance in cancer: biological and mathematical assessment, theoretical treatment optimisation. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(11):2627–2645.

[10] Coldman, A. J. & Murray, J. (2000). Optimal control for a stochastic model of cancer chemotherapy. *Mathematical Biosciences*, 168(2):187–200.

[11] Cunningham, J., Thuijsman, F., Peeters, R., Viossat, Y., Brown, J., Gatenby, R., & Staňková, K. (2020). Optimal control to reach eco-evolutionary stability in metastatic castrate-resistant prostate cancer. *PLoS One*, 15(12):e0243386.

[12] Cunningham, J. J., Brown, J. S., Gatenby, R. A., & Staňková, K. (2018). Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. *Journal of theoretical biology*, 459:67–78.

[13] Day, R. S. (1986). Treatment sequencing, asymmetry, and uncertainty: protocol strategies for combination chemotherapy. *Cancer Research*, 46(8):3876–3885.

[14] Dhawan, A., Nichol, D., Kinose, F., Abazeed, M. E., Marusyk, A., Haura, E. B., & Scott, J. G. (2017). Collateral sensitivity networks reveal evolutionary instability and novel treatment strategies in alk mutated non-small cell lung cancer. *Scientific reports*, 7(1):1–9.

[15] Dolgov, S., Kalise, D., & Kunisch, K. K. (2021). Tensor decomposition methods for high-dimensional Hamilton–Jacobi–Bellman equations. *SIAM Journal on Scientific Computing*, 43(3):A1625–A1650.

[16] Engen, S., Bakke, Ø., & Islam, A. (1998). Demographic and environmental stochasticity-concepts and definitions. *Biometrics*, pages 840–846.

[17] Enriquez-Navas, P. M., Kam, Y., Das, T., Hassan, S., Silva, A., Foroutan, P., Ruiz, E., Martinez, G., Minton, S., Gillies, R. J., et al. (2016). Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Science translational medicine*, 8(327):327ra24–327ra24.

[18] Farrokhian, N., Maltas, J., Dinh, M., Durmaz, A., Ellsworth, P., Hitomi, M., McClure, E., Marusyk, A., Kaznatcheev, A., & Scott, J. G. (2022). Measuring competitive exclusion in non–small cell lung cancer. *Science Advances*, 8(26):eabm7212.

[19] Fischer, A., Vázquez-García, I., & Mustonen, V. (2015). The value of monitoring to control evolving populations. *Proceedings of the National Academy of Sciences*, 112(4):1007–1012.

[20] Fleming, W. H. & Rishel, R. W. (2012). *Deterministic and stochastic optimal control*, volume 1. Springer Science & Business Media.

[21] Fleming, W. H. & Soner, H. M. (2006). *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media.

[22] Fudenberg, D. & Harris, C. (1992). Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57(2):420–441.

[23] Gatenby, R. A., Silva, A. S., Gillies, R. J., & Frieden, B. R. (2009). Adaptive therapy. *Cancer research*, 69(11):4894–4903.

[24] Gillies, R. J., Verduzco, D., & Gatenby, R. A. (2012). Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nature Reviews Cancer*, 12(7):487–493.

[25] Gluzman, M., Scott, J. G., & Vladimirsky, A. (2020). Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory. *Proceedings of the Royal Society B*, 287(1925):20192454.

[26] Greene, J. M., Gevertz, J. L., & Sontag, E. D. (2019). Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. *JCO clinical cancer informatics*, 3:1–20.

[27] Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., & Lander, E. S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4):633–644.

[28] Hadeler, K. P. (1981). Stable polymorphisms in a selection model with mutation. *SIAM Journal on Applied Mathematics*, 41(1):1–7.

[29] Hamdache, A., Elmouki, I., & Saadi, S. (2014). Optimal control with an isoperimetric constraint applied to cancer immunotherapy. *International Journal of Computer Applications*, 94(15).

[30] Haslam, A., Kim, M., & Prasad, V. (2021). 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.

[31] Hening, A., Nguyen, D. H., & Chesson, P. (2021). A general theory of coexistence and extinction for stochastic ecological communities. *Journal of Mathematical Biology*, 82(6):56.

[32] Hofbauer, J. (1985). The selection mutation equation. *Journal of mathematical biology*, 23:41–53.

[33] Hofbauer, J. & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge university press.

[34] Imamovic, L. & Sommer, M. O. (2013). Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. *Science translational medicine*, 5(204):204ra132–204ra132.

[35] Iram, S., Dolson, E., Chiel, J., Pelesko, J., Krishnan, N., Güngör, Ö., Kuznets-Speck, B., Deffner, S., Ilker, E., Scott, J. G., et al. (2021). Controlling the speed and trajectory of evolution with counterdiabatic driving. *Nature Physics*, 17(1):135–142.

[36] Jeong, Y. D., Kim, K. S., Roh, Y., Choi, S., Iwami, S., Jung, I. H., & Li, G. (2022). Optimal feedback control of cancer chemotherapy using Hamilton-Jacobi-Bellman equation. *Complexity*, 2022.

[37] Katouli, A. A. & Komarova, N. L. (2011). The worst drug rule revisited: mathematical modeling of cyclic cancer treatments. *Bulletin of mathematical biology*, 73:549–584.

[38] Kaznatcheev, A., Peacock, J., Basanta, D., Marusyk, A., & Scott, J. G. (2019). Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nature ecology & evolution*, 3(3):450–456.

[39] Kaznatcheev, A., Vander Velde, R., Scott, J. G., & Basanta, D. (2017). Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *British journal of cancer*, 116(6):785–792.

[40] Kumar, A. & Vladimirsky, A. (2010). An efficient method for multiobjective optimal control and optimal control subject to integral constraints. *Journal of Computational Mathematics*, 28(4):517–551.

[41] Kumar, N., Cramer, G. M., Dahaj, S. A. Z., Sundaram, B., Celli, J. P., & Kulkarni, R. V. (2019). Stochastic modeling of phenotypic switching and chemoresistance in cancer cell populations. *Scientific reports*, 9(1):1–10.

[42] Kuosmanen, T., Cairns, J., Noble, R., Beerenwinkel, N., Mononen, T., & Mustonen, V. (2021). Drug-induced resistance evolution necessitates less aggressive treatment. *PLoS computational biology*, 17(9):e1009418.

[43] Lande, R., Engen, S., Saether, B.-E., et al. (2003). *Stochastic population dynamics in ecology and conservation*. Oxford University Press on Demand.

[44] Lauring, A. S. & Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens*, 6(7):e1001005.

[45] Maltas, J. & Wood, K. B. (2019). Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLoS biology*, 17(10):e3000515.

[46] Martin, R., Fisher, M., Minchin, R., & Teo, K. (1992). Optimal control of tumor size used to maximize survival time when cells are resistant to chemotherapy. *Mathematical Biosciences*, 110(2):201–219.

[47] Marusyk, A., Tabassum, D. P., Altrock, P. M., Almendro, V., Michor, F., & Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520):54–58.

[48] Miksis, Z. M. & Zhang, Y.-T. (2022). Sparse-grid implementation of fixed-point fast sweeping WENO schemes for eikonal equations. *Communications on Applied Mathematics and Computation*, pages 1–27.

[49] Mononen, T., Kuosmanen, T., Cairns, J., & Mustonen, V. (2023). Understanding cellular growth strategies via optimal control. *Journal of the Royal Society Interface*, 20(198):20220744.

[50] Nichol, D., Jeavons, P., Fletcher, A. G., Bonomo, R. A., Maini, P. K., Paul, J. L., Gatenby, R. A., Anderson, A. R., & Scott, J. G. (2015). Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. *PLoS computational biology*, 11(9):e1004493.

[51] Nichol, D., Rutter, J., Bryant, C., Hujer, A. M., Lek, S., Adams, M. D., Jeavons, P., Anderson, A. R., Bonomo, R. A., & Scott, J. G. (2019). Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nature communications*, 10(1):1–10.

[52] Nourmohammad, A. & Eksin, C. (2021). Optimal evolutionary control for artificial selection on molecular phenotypes. *Physical Review X*, 11(1):011044.

[53] Nowakowski, A. & Popa, A. (2013). A dynamic programming approach for approximate optimal control for cancer therapy. *J. Optim. Theory Appl.*, 156(2):365–379.

[54] Pontryagin, L., Boltyanskii, V., Gamkrelidze, R., & Mishchenko, E. (1962). *The mathematical theory of optimal processes*. John Wiley & Sons, Inc.

[55] Schättler, H. & Ledzewicz, U. (2015). *Optimal Control for Mathematical Models of Cancer Therapies*, volume 42 of *Interdisciplinary Applied Mathematics*. Springer New York.

[56] Schreiber, S. J., Benaïm, M., & Atchadé, K. A. (2011). Persistence in fluctuating environments. *Journal of Mathematical Biology*, 62:655–683.

[57] Taylor, P. D. & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical biosciences*, 40(1-2):145–156.

[58] Vander Velde, R., Yoon, N., Marusyk, V., Durmaz, A., Dhawan, A., Miroshnychenko, D., Lozano-Peral, D., Desai, B., Balynska, O., Poleszhuk, J., et al. (2020). Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. *Nature communications*, 11(1):1–13.

[59] Wang, M., Patnaik, N., Somalwar, A., Wu, J., & Vladimirsky, A. (2023). Risk-aware stochastic control of a sailboat. *preprint:* `https://arxiv.org/abs/2309.13436`.

[60] West, J., You, L., Zhang, J., Gatenby, R. A., Brown, J. S., Newton, P. K., & Anderson, A. R. (2020). Towards multidrug adaptive therapy. *Cancer research*, 80(7):1578–1589.

[61] West, J. B., Dinh, M. N., Brown, J. S., Zhang, J., Anderson, A. R., & Gatenby, R. A. (2019). Multidrug cancer therapy in metastatic castrate-resistant prostate cancer: an evolution-based strategy. *Clinical Cancer Research*, 25(14):4413–4421.

[62] Wilke, C. O. (2005). Quasispecies theory in the context of population genetics. *BMC evolutionary biology*, 5(1):1–8.

[63] Zhang, J., Cunningham, J. J., Brown, J. S., & Gatenby, R. A. (2017). Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature communications*, 8(1):1–9.

[64] Zhang, X., Long, J., Hu, W., Han, J., et al. (2022). Initial value problem enhanced sampling for closed-loop optimal control design with deep neural networks. *arXiv preprint arXiv:2209.04078*.

[65] Zhao, B., Sedlak, J. C., Srinivas, R., Creixell, P., Pritchard, J. R., Tidor, B., Lauffenburger, D. A., & Hemann, M. T. (2016). Exploiting temporal collateral sensitivity in tumor clonal evolution. *Cell*, 165(1):234–246.

[66] Zouhri, S., El Baroudi, M., & Saadi, S. (2022). Optimal control with isoperimetric constraint for chemotherapy of tumors. *International Journal of Applied and Computational Mathematics*, 8(4):215.

SM Supplementary Materials. Mathematical details and additional computational experiments.