Stochastic optimal control to guide 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

While adaptive cancer therapy is beginning to prove a promising approach of building evolutionary dynamics into therapeutic scheduling, the stochastic nature of cancer evolution has rarely been incorporated. Various sources of random perturbations can impact the evolution of heterogeneous tumors. In this paper, we propose a method that can effectively select optimal adaptive treatment policies under randomly evolving tumor dynamics based on Stochastic Optimal Control theory. We first construct a stochastic model of cancer dynamics under drug therapy based on Evolutionary Game theory. That model is then used to improve the cumulative "cost", a combination of the total amount of drugs used and the time to recovery. As this cost becomes random in a stochastic setting, we maximize the probability of recovery under a pre-specified cost threshold (or a "budget"). We can achieve our goal for a range of threshold values simultaneously using the tools of dynamic programming. We then compare our threshold-aware policies with the policies previously shown to be optimal in the deterministic setting. We show that this threshold-awareness yields a significant improvement in the probability of under-the-budget recovery, which is correlated with a lower general drug usage. The particular model underlying our discussion has originated in [22], but the presented 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, tumour heterogeneity, evolutionary game theory, environmental stochasticity, stochastic optimal control, threshold-aware optimal treatment policy

1 Background

The advent of personalized medicine in cancer has changed the way we think about chemotherapy 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

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

still small; it was estimated in 2020 that only $\approx 5\%$ of patients benefit from these targeted therapies [17]. 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 [14].

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 [19, 26, 29, 30] as well as for a number of cancers [7, 37]. In the study of therapeutic scheduling, adaptive therapy, which uses mathematical tools from Evolutionary Game Theory (EGT), has shown promise not only in theory [13], but also in a phase 2 trial for men with metastatic prostate cancer [36]. Experimentally, there have been confirmations of EGT principles *in vivo*[9] as well as more quantitatively focused assay development *in vitro* [21], and observations of game interactions using these methods [10]. The majority of theoretical work in this space has focused on simulation-based numerical optimization of different drug regimens for *deterministic* models of cancer evolution [5, 6, 33, 34]. In this paper, our goals are (a) to extend such game-theoretic models to account for stochastic perturbations and (b) to provide efficient computational tools for this more general setting.

Our previous paper [15] has focused on a deterministic model [22] and showed the advantages of performing such optimization in the framework of Dynamic Programming (DP) [2]. The key idea was to leverage the competition among different sub-populations of cancer cells to improve the timing and duration of the drug therapy. We showed how DP could be used by a clinician, with appropriate data, to formally optimize outcomes given a discussion about patient's goals (to include tradeoffs between toxicity and quality of life, for example). While our approach represented a meaningful step toward the application of DP to cancer therapy, it relied on several simplifying modeling assumptions. One of the more significant (and more limiting) among them was due to our ignoring stochasticity as an important aspect of cancer biology.

Cancer (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. These include individual genetic differences, fate transitions [16], varying reactions to drugs [23], differences in signalling, and small-scale variations in tumor microenvironment. Many instances of such *demographic stochasticity* [24] can be safely "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 [25, 35] and exist as distributions around a central genotype, with all cells in the group behaving in a similar manner despite random birth/death events [8, 24] and small within-the-group genetic heterogeneities [20]. 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 [8, 24]; e.g., for cancer they might result from therapy-unrelated drugs or from frequent small changes in the host's physiology.

Modeling such perturbations in continuous-time usually results in *Stochastic Differential Equations* (SDEs) [1, 3], whose behavior can be optimized using Stochastic Optimal Control Theory [11]. 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 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. 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., patient's recovery without exceeding a specific cost threshold). Our choice of optimization objective is primarily due to a possibility of failure (e.g., a patient dying despite the therapy), which makes any description of expected cost heavily dependent on the perceived cost of such failure – a quantity that would be hard to justify in many biomedical applications.

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 our main contribution is a new computational tool for the evolutionary therapy toolbox, with a broad applicability to stochastic cancer models, which will become increasingly realistic in the future.

2 Methods

2.1 Base deterministic model and effects of stochastic perturbations

We adopt the base model of cancer evolution proposed by Kaznatcheev et al. in [22], which describes a competition of 3 types of cancer cells. Glycoltyic 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. The respective proportions of these types in the overall population of cancer cells is encoded by $(x_{\mathrm{G}}, x_{\mathrm{D}}, x_{\mathrm{V}})$. 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" for all cancer cells since it is damaging for the surrounding non-cancerous tissue. The fundamental model in [22] is based on a game of $(n+1)$ locally interacting cells, and the fitness of each of them depends on the current x_{G}, x_{D}, and x_{V}. It was observed that, for a range of parameter values, these subpopulations exhibit

"cyclic dominance": when GLY are in the majority, VOP have the highest fitness; when VOP are in the majority, DEF have the highest fitness; and when DEF are in the majority, GLY are the fittest. A *replicator equation* [18, 31] is a standard EGT model for predicting the changes in type-proportions as a function of time. For the cyclic dominance described above, it predicts periodic orbits, in which $x_{\mathrm{G}}(t)$, $x_{\mathrm{D}}(t)$, and $x_{\mathrm{V}}(t)$ alternate in being dominant in the tumor, with the amplitude of oscillations determined by the initial conditions; see Figure 1(A). However, this periodicity is destroyed when one recognizes that fitness functions are also subject to stochastic perturbations. The nature of perturbations is detailed below, but the result is illustrated in Figure 1(B).

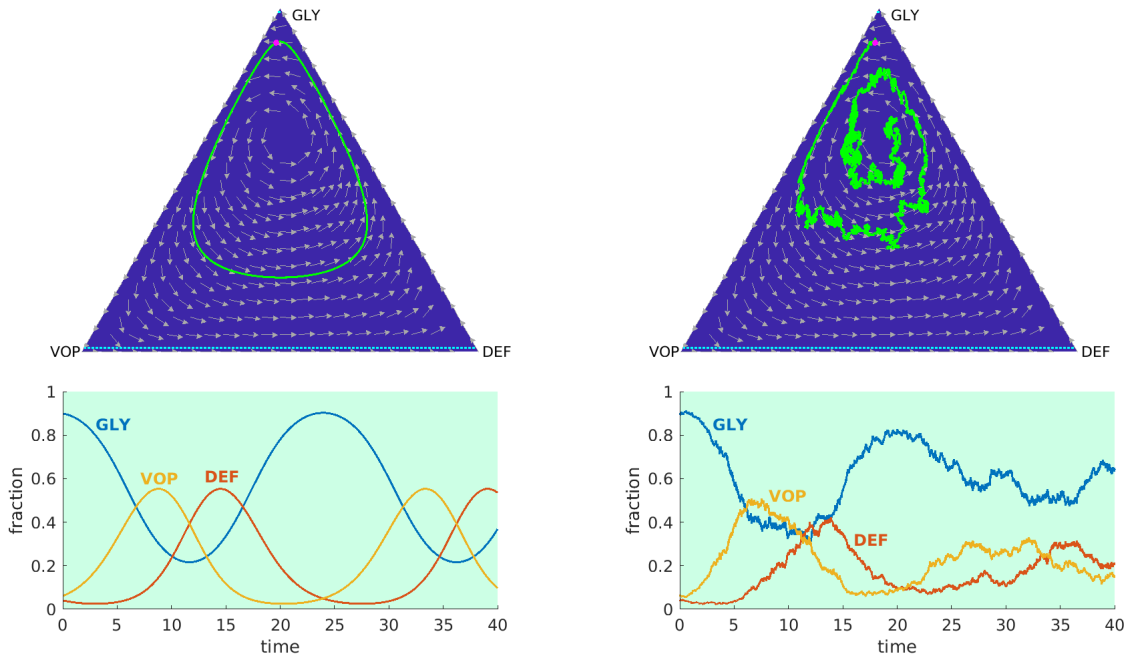


Figure 1: A comparison between natural (i.e., without therapy) deterministic and stochastic dynamics starting from an initial state $(x_{\mathrm{G}}, x_{\mathrm{D}}, x_{\mathrm{V}}) = (0.9, 0.04, 0.06)$ (magenta dot): (a) deterministic; (b) stochastic. We adopt the settings in [15] to use *blue background* and *green trajectory* as indications of not prescribing any drugs. The vector field is shown by *gray arrows*. Here we use *dashed cyan lines* to indicate recovery and failure barriers. **Top row:** sample trajectories on a GLY-DEF-VOP triangle. **Bottom row:** evolution of sub-populations with respect to time based on the reference trajectories from the top row.

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 [22], we use a slightly different reduction, rewriting everything in terms of the proportion of glycolitic cells in the tumor $p(t) = x_{\mathrm{G}}(t)$ and the proportion of VOP among aerobic cells $q(t) = x_{\mathrm{V}}(t)/(x_{\mathrm{V}}(t) + x_{\mathrm{D}}(t))$. A drug therapy (e.g. affecting the fitness of GLY cells only) is similarly easy to encode by modifying the replicator equation; see equation (2.2) in Box 1 and the Supplementary Materials in [22] for the derivation[I]. This original deterministic model provided a simple and very convincing illustration for 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 cure ($p(t)$ falling below the

[I]The goal of driving the GLY fraction $p(t) = x_{\mathrm{G}}(t)$ down below some "recovery threshold" r_b is justified in [22] by noting that, with GLY gone, the DEF cells will then quickly overcome VOP, leading to "an aerobic tumour with no - or significantly diminished - ability to recruit blood vessels." Our approach can be equally well used with other therapy goals as well. We adopt this one primarily to illustrate the method and simplify the comparison with prior results.

specified "recovery barrier" r_b) or a death ($p(t)$ rising above the specified "failure barrier" 1 - f_b) depending on how long we wait until this therapy starts.

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 $p(t) < r_b$. Gluzman et al. have addressed this in [15] using the framework of optimal control theory [11]. 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 $\mathcal{J}(q_0, p_0, d(\cdot)) = \int_0^T d(t)\, ds + \delta T$, where T is the time till recovery and the treatment time penalty $\delta > 0$ encodes the balance between the two optimization goals (total drugs vs total time). They obtained the deterministic-optimal policy in *feedback form* (i.e., $d = d_*(q, p)$) using the tools of dynamic programming [2]. The mathematical structure of this problem makes it easy to show that an optimal therapy will be "bang-bang" – i.e., $d_*(q, p)$ prescribes either zero drugs or drugs at the MTD-rate $d_{\max}$ for every possible tumor composition. Figure 2(A) summarizes that deterministic-optimal policy and shows an example of the corresponding trajectory for one specific initial (q_0, p_0). However, if the fitnesses of subpopulations are subject to stochastic perturbations detailed in subsection 2.2, then the resulting trajectory becomes random, and the actual cost $\mathcal{J}$ incurred along it can vary significantly; see the example in Figure 2(B). In fact, some samples might actually end up in failure (with stochastic perturbations pushing $p(t)$ above $1 - f_b$). 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:

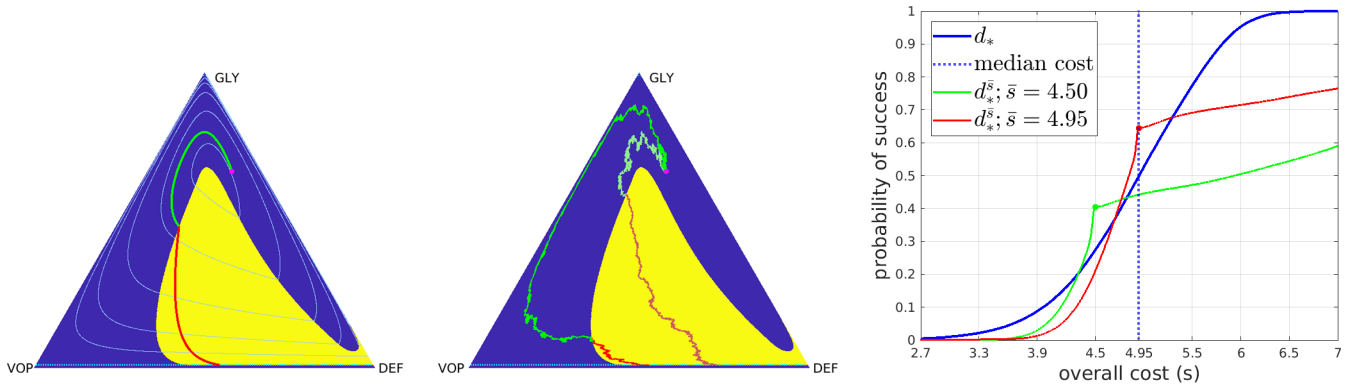
$$F_{d_*}(s) = \mathbb{P}(\mathcal{J} \leq s),$$

whose graph is shown in Figure 2(C). It shows, for example, that even though $\mathcal{J}$ averaged over all simulations that resulted in recovery was 4.91, which is not far from 5.13 predicted in deterministic case, at least 23% of simulations resulted in a significantly higher $\mathcal{J} > 5.5$. This motivates our optimization strategy for the stochastic problem: deriving a *threshold-aware optimal policy*[II] $d_*^{\bar{s}}$ to maximize the probability of recovery without exceeding a specific threshold cost $\bar{s}$. As shown in Figure 2(C), such policies can provide a significant threshold-advantage over the deterministic-optimal therapy.

2.2 The stochastic model and the optimization problem

In the deterministic model described above, the discussion typically starts by considering the actual *sizes* of cancer subpopulations $(z_{\mathrm{G}}, z_{\mathrm{D}}, z_{\mathrm{V}})$, whose respective rates of growth/decay are assumed to be the respective subpopulation fitnesses $(\psi_{\mathrm{G}}, \psi_{\mathrm{D}}, \psi_{\mathrm{V}})$. Since these fitnesses are determined by the current subpopulation *fractions* $\left(x_{\mathrm{G}} = \frac{z_{\mathrm{G}}}{z_{\mathrm{G}}+z_{\mathrm{D}}+z_{\mathrm{V}}}, x_{\mathrm{D}} = \frac{z_{\mathrm{D}}}{z_{\mathrm{G}}+z_{\mathrm{D}}+z_{\mathrm{V}}}, x_{\mathrm{V}} = \frac{z_{\mathrm{V}}}{z_{\mathrm{G}}+z_{\mathrm{D}}+z_{\mathrm{V}}}\right)$, it is natural to derive equations that model the evolution of the latter. This is precisely the idea of the "replicator" *Ordinary Differential Equations* (see Supplementary Materials section 1S), which is stated in the reduced (q, p) coordinates in Box 1.

[II] In the interest of computational reproducibility, we provide the full source code for computing threshold-aware policies at `https://github.com/eikonal-equation/Stochastic-Cancer`



(a) (b) (c)

Figure 2: Stochastic driven system with deterministic-optimal policy starting from an initial state $(q_0, p_0) = (0.26, 0.665)$ (magenta dot): (a) the optimal trajectory found by deterministic driven system (2.2) 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), *green part of a trajectory/path* means not prescribing drugs and *red part of a trajectory/path* means prescribing drugs at MTD rate. The level sets of the value function in the deterministic case are shown in *light blue*. In (c), the *solid blue* curve is the CDF generated with the deterministic-optimal policy. Its median (*dashed blue* line) is 4.95 while its mean conditioning on success is 4.91. The *solid green* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.50$; and the *solid red* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.95$.

A common 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 [12], is suitable for modeling heterogeneous tumors, in which subpopulations not only interact,[27] but can also vary in their growth rates over time.[32] Adopting the usual probabilistic notation of using capital letters for random variables, we can again start with the subpopulation sizes $(Z_{\mathrm{G}}, Z_{\mathrm{D}}, Z_{\mathrm{V}})$ and derive the *Stochastic Differential Equations* modeling the evolution of $(X_{\mathrm{G}}, X_{\mathrm{D}}, X_{\mathrm{V}})$. The details of the derivation are in Supplementary Materials section 2S, but the summary is reported in Box 2. Following [15], we also assume the drug therapy targets GLY cells only; the resulting effect is included in equation (2.9) for the reduced coordinates (Q, P).

As before, the process terminates as soon as $P(t)$ crosses a recovery barrier (GLY's are low, the patient is "cured") 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 $d(\cdot)$. It might feel natural to select a policy minimizing the expected $\mathcal{J}$, but since there is a non-zero probability of failure under every policy, this expectation is actually infinite. So, our alternative approach is to select a policy maximizing the probability of desirable outcomes; i.e., the probability $\mathbb{P}(\mathcal{J}(q_0, p_0, d(\cdot)) \leq \bar{s})$, where $\bar{s} > 0$ is a desired cost threshold. Our goal is to compute such threshold-aware policies efficiently for all starting tumor configurations (q_0, p_0) and all threshold levels $\bar{s}$ simultaneously[III]. It is easy to see that here feedback policies will have to determine the drug administration rate $d(\cdot)$ based on the current tumor configuration

[III] A similar approach has been recently developed in [4] for controlling "piecewise-deterministic" processes, where perturbations happen at discrete points in time and amount to abrupt switches in system dynamics.

and the cost accumulated so far. This makes it natural to treat our chosen $\bar{s}$ as an *initial cost budget*, tracking the remaining budget $s(t)$ by solving Eq. (2.17) in Box 3.

To solve this optimization problem, we define a *value function* $v(q,p,s) = \max_{d(\cdot)} \mathbb{P}(\mathcal{J}(q,p,d(\cdot)) \leq s)$, encoding the maximized chance of starting with (q,p) and curing the patient without exceeding s. It can be shown that v satisfies Eq. (2.14) in Box 3, which can be solved numerically; see Supplementary Materials section 4S.1 for detailed derivations and section 5S.1 for a description of our numerical method. The value function is useful due to the fundamental *tail optimality* property underlying most dynamic programming approaches: the optimal strategy from the current state is the same as it would have been if we started here. In our context, this means that v can be used to produce the optimal $d_*(q,p,s)$, where (q,p) is the *current* tumor and s is the *current* remaining budget. So, starting with $Q(0) = q_0, P(0) = p_0$, the threshold-aware optimal policy for a target threshold $\bar{s}$ can be obtained as $d_*^{\bar{s}} = d_*\Big(Q(t), P(t), s(t)\Big)$, where $s(0) = \bar{s}$. As the time increases, we use up some of the budget (decreasing $s(t)$) and slide into the lower s-slices of v to determine the new optimal rate of drug administration. The process continues until the patient is either cured or dies or our budget is exhausted ($s(t) = 0$).

As in the deterministic dynamics case [15], our threshold-aware optimal policy is still *bang-bang*; i.e., d_* always prescribes either 0 (no drugs) or $d_{\max}$ (the MTD level); see Supplementary Materials section 4S.1. Nevertheless, the drugs-on region in the (q,p) space becomes s-dependent. A subtle consequence of this is that a policy optimal for a chosen threshold $\bar{s}_1$ is typically sub-optimal for any other threshold $\bar{s}_2 \neq \bar{s}_1$; see section 3.2.

3 Results

The results in this section are obtained by solving the optimal control problem described in Box 3. 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. [22] and Gluzman et al. [15]. However, we use $r_b = f_b = 10^{-2}$ and $\delta = 0.05$ as opposed to $r_b = f_b = 10^{-1.5}$ and $\delta = 0.01$ in [15]. Additionally, here 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. Example with different volatilities can be found in section 6S of Supplementary Materials. To be consistent with [15], 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.

3.1 Threshold-aware treatment policies and optimal probability of success

In Figure 3, we present some representative s-slices of threshold-aware optimal policies and their corresponding optimal probability of success for these specific threshold values[IV]. We observe that the drugs-on region is strongly s-dependent and completely different from the one in the deterministic-optimal case shown in Figure 2(A)&(B). Note that since now the cancer evolution follows a stochastic dynamics given in (2.9), one would have different optimal sample paths for different realizations of random perturbations even if the starting configuration stays the same. We show 3 representative random paths in Figure 4. In particular, Figure 4(A) shows a sample path that leads to recovery incurring the total cost of 4.70, which is below our initial budget $\bar{s} = 5.0$. We will use this

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

Box 1: Cancer evolution model from Kaznatcheev et al. [22] and optimal control problem from Gluzman et al. [15]

$(x_{\mathrm{G}}, x_{\mathrm{D}}, x_{\mathrm{V}})$ for GLY, DEF, and VOP respectively.

Note: $x_{\mathrm{G}} + x_{\mathrm{D}} + x_{\mathrm{V}} = 1$.

$$\begin{cases} q = \dfrac{x_{\mathrm{V}}}{x_{\mathrm{V}} + x_{\mathrm{D}}}, \\ p = x_{\mathrm{G}}, \end{cases} \quad \text{or} \quad \begin{cases} x_{\mathrm{G}} = p, \\ x_{\mathrm{D}} = (1-q)(1-p), \\ x_{\mathrm{V}} = (1-p)q. \end{cases} \tag{2.1}$$

Evolution dynamics in reduced coordinates with control on therapy intensity:

$$\begin{cases} \dot{q}(t) = q(t)\Big(1 - q(t)\Big)\Big(\frac{b_v}{n+1}\sum\limits_{k=0}^{n} p^k(t) - c\Big) =: \beta(q,p), \\ \dot{p}(t) = p(t)\Big(1 - p(t)\Big)\Big(\frac{b_a}{n+1} - (b_v - c)q(t) - \boldsymbol{d(t)}\Big) =: \alpha(q,p,d); \\ q(0) = q_0,\ p(0) = p_0. \end{cases} \tag{2.2}$$

Definitions and Parameters:

- $d : \mathbb{R}_+ \to [0, d_{\max}]$, time-dependent intensity of GLY-targeting therapy
- b_a, the benefit per unit of acidification;
- b_v, the benefit from the oxygen per unit of vascularization;
- c, the cost of production VEGF;
- $(n+1)$, the number of cells in the interaction group.

Conditions for the heterogeneous regime: $\dfrac{b_a}{n+1} < b_v - c < cn.$ (2.3)

Process *terminates* as soon as either

$$\begin{cases} p(t) < r_b, & \text{if therapy succeeds;} \\ p(t) > 1 - f_b, & \text{if therapy fails.} \end{cases}$$

Terminal set:

$$\Delta = \Big\{(q,p) \in [0,1] \times [0,1] \ \mid \ p < r_b \text{ or } p > 1 - f_b\Big\}. \tag{2.4}$$

Total treatment time: $T(q_0, p_0, d(\cdot)) = \inf\Big\{t \in \mathbb{R}_+ \ \mid \ (q(t), p(t)) \in \Delta,\ q(0) = q_0,\ p(0) = p_0\Big\}.$ (2.5)

Treatment cost function: $\mathcal{J}(q_0, p_0, d(\cdot)) = \displaystyle\int_0^T (d(\tau) + \delta)\,\mathrm{d}\tau + g\Big(q(T), p(T)\Big),$ (2.6)

where $T := T(q_0, p_0, d(\cdot))$ is the terminal time and $g(q,p)$ is the terminal cost function where $g(q,p) = +\infty$ if $p > 1 - f_b$ and $g(q,p) = 0$ otherwise.

Value function: $u(q_0, p_0) = \inf\limits_{d(\cdot)} \mathcal{J}(q_0, p_0, d(\cdot))$

is found by numerically solving a first-order HJB PDE. The deterministic-optimal policy is found in feedback form $d_* = d(q,p)$. See details in Gluzman et al. [15].

example to illustrate how threshold-aware policies work.

Starting from the initial budget $\bar{s} = 5.0$, the optimal decision on whether to use drugs right away is based on the first diagram (the "$s = 5.0$" case) in Figure 3(A). For the initial tumor state $(q_0, p_0) = (0.26, 0.665)$,

Box 2: Stochastic model of cancer evolution

Geometric Brownian motion as the sub-population growth model:

$(Z_{\mathrm{G}},\, Z_{\mathrm{D}},\, Z_{\mathrm{V}})$ for actual sub-population sizes of GLY, DEF, and VOP cells.

$$\begin{cases} \mathrm{d}Z_{\mathrm{G}} = (\psi_{\mathrm{G}}\,\mathrm{d}t + \sigma_{\mathrm{G}}\,\mathrm{d}W_{\mathrm{G}})Z_{\mathrm{G}}, \\ \mathrm{d}Z_{\mathrm{D}} = (\psi_{\mathrm{D}}\,\mathrm{d}t + \sigma_{\mathrm{D}}\,\mathrm{d}W_{\mathrm{D}})Z_{\mathrm{D}}, \\ \mathrm{d}Z_{\mathrm{V}} = (\psi_{\mathrm{V}}\,\mathrm{d}t + \sigma_{\mathrm{V}}\,\mathrm{d}W_{\mathrm{V}})Z_{\mathrm{V}}. \end{cases} \tag{2.7}$$

Definitions and Parameters:

- $W_{\mathrm{G}}, W_{\mathrm{D}}, W_{\mathrm{V}}$, independent standard Brownian motions for GLY, DEF, and VOP cells respectively;
- $\sigma_{\mathrm{G}}, \sigma_{\mathrm{D}}, \sigma_{\mathrm{V}} \geq 0$, volatilities for Z_{G}, Z_{D}, and Z_{V} respectively.

Transformations of proportions of cancer cells:

$(X_{\mathrm{G}},\, X_{\mathrm{D}},\, X_{\mathrm{V}})$ for proportions of GLY, DEF, and VOP cells.

Note we still have $X_{\mathrm{G}} + X_{\mathrm{D}} + X_{\mathrm{V}} = 1$.

$$\begin{cases} Q = \dfrac{X_{\mathrm{V}}}{X_{\mathrm{V}} + X_{\mathrm{D}}}, \\ P = X_{\mathrm{G}}, \end{cases} \quad \text{or} \quad \begin{cases} X_{\mathrm{G}} = P, \\ X_{\mathrm{D}} = (1-Q)(1-P), \\ X_{\mathrm{V}} = (1-P)Q. \end{cases} \tag{2.8}$$

Stochastic evolution dynamics in reduced coordinates with control on therapy intensity:

$$\begin{cases} \mathrm{d}Q = \beta(Q,P)\,\mathrm{d}t + \zeta(Q;\sigma_{\mathrm{D}},\sigma_{\mathrm{V}})\,\mathrm{d}t + h_1(Q;\sigma_{\mathrm{V}})\,\mathrm{d}W_{\mathrm{V}} + h_2(Q;\sigma_{\mathrm{D}})\,\mathrm{d}W_{\mathrm{D}}, \\ \mathrm{d}P = \alpha(Q,P,d)\,\mathrm{d}t + \eta(Q,P;\sigma_{\mathrm{G}},\sigma_{\mathrm{D}},\sigma_{\mathrm{V}})\,\mathrm{d}t + m_1(P;\sigma_{\mathrm{G}})\,\mathrm{d}W_{\mathrm{G}} + m_2(Q,P;\sigma_{\mathrm{D}})\,\mathrm{d}W_{\mathrm{D}} + m_3(Q,P;\sigma_{\mathrm{V}})\,\mathrm{d}W_{\mathrm{V}}; \\ Q(0) = q_0,\ \ P(0) = p_0, \end{cases} \tag{2.9}$$

where α and β are defined in Box 1 while the exact definitions of functions ζ, η, h_1, h_2, m_1, m_2, and m_3 can be found in section 3S of Supplementary Materials.

Process *terminates* as soon as either

$$\begin{cases} P(t) < r_b, & \text{if therapy succeeds;} \\ P(t) > 1 - f_b, & \text{if therapy fails.} \end{cases}$$

Terminal set:

$$\Delta = \Big\{(q,p) \in [0,1] \times [0,1] \ \mid\ p < r_b \text{ or } p > 1 - f_b\Big\}. \tag{2.10}$$

this indicates that $d_*(q_0, p_0, 5) = 0$ (not prescribing drugs initially) would maximize the probability of curing the patient 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. Of course, in reality we constantly reevaluate the decision on d_* (as s changes continuously while Figure 3(A) presents just a few representative slices) taking into account the changing tumor configuration $(Q(t), P(t))$. For example, the path in Figure 4(A) shows the drug use starting shortly after $s = 4.3$, from which point $(Q(t), P(t))$ stays in the yellow region until recovery.

However, due to random perturbations, the above process can also stop when 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 it.

Box 3: Objective function

Total treatment time: $$T(q_0, p_0, d(\cdot)) = \inf\Big\{t \in \mathbb{R}_+ \;\mid\; (Q(t), P(t)) \in \Delta,\; Q(0) = q_0,\; P(0) = p_0\Big\}. \tag{2.11}$$

Treatment cost function: $$\mathcal{J}(q_0, p_0, d(\cdot)) = \int_0^T (d(\tau) + \delta)\,\mathrm{d}\tau + g\Big(Q(T), P(T)\Big), \tag{2.12}$$

where $T := T(q_0, p_0, d(\cdot))$ is the terminal time and $g(q, p)$ is the terminal cost function where $g(q, p) = +\infty$ if $p > 1 - f_b$ and $g(q, p) = 0$ otherwise.

Value function: $$v(q_0, p_0, \bar{s}) = \sup_{d(\cdot)} \mathbb{P}\Big(\mathcal{J}(q_0, p_0, d(\cdot)) \le \bar{s}\Big) \tag{2.13}$$

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

$$\max_{d\in[0,d_{\max}]} \left\{ \left[-p(1-p)\frac{\partial v}{\partial p} - \frac{\partial v}{\partial s}\right] d \right\} - \delta\frac{\partial v}{\partial s} + f_1(q,p)\frac{\partial v}{\partial q} + f_2(q,p)\frac{\partial v}{\partial p} + f_3(q,p)\frac{\partial^2 v}{\partial q^2} + f_4(q,p)\frac{\partial^2 v}{\partial p^2} + f_5(q,p)\frac{\partial^2 v}{\partial q \partial p} = 0,$$

$$(q, p) \in ([0, 1] \times [0, 1])/\Delta, \tag{2.14}$$

where $s \ge 0$ and the exact definitions of functions f_1, f_2, f_3, f_4, and f_5 can be found in section 4S.1 of Supplementary Materials.

The boundary conditions of HJB equation: $$\begin{cases} v(q,p,s) = 1, & \text{if } p < r_b; \\ v(q,p,s) = 0, & \text{if } p > 1 - f_b; \\ v(q,p,0) = 0, & \text{if } p > r_b. \end{cases} \tag{2.15}$$

The optimal policy can be found in feedback form: $$d_*(q,p,s) = \begin{cases} d_{\max}, & \text{if } \Big(v_p p(1-p) + v_s\Big) < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.16}$$

The ODE that budget s satisfies: $$\dot{s}(t) = -\Big(d\Big(Q(t), P(t), s(t)\Big) + \delta\Big), \quad s(0) = \bar{s}. \tag{2.17}$$

We also fail in keeping the cost below the threshold when the budget is exhausted before the patient is cured, as in Figure 4(C). Threshold-aware policies provide no guidance once $s = 0$, but we need to continue the treatment since the patient is still alive. To estimate the patient's chances of recovery and the probability distribution for the resulting cost beyond the originally specified threshold $\bar{s}$, we then switch to the deterministic-optimal policy described in section 2.1 and Figure 2.

Figure 3(B) shows the optimal probability of success $v(q, p, s)$ for different threshold values. One observes that v has particularly large gradient near the level curves of the deterministic-optimal value function u shown in Figure 2(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 be cured under 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.

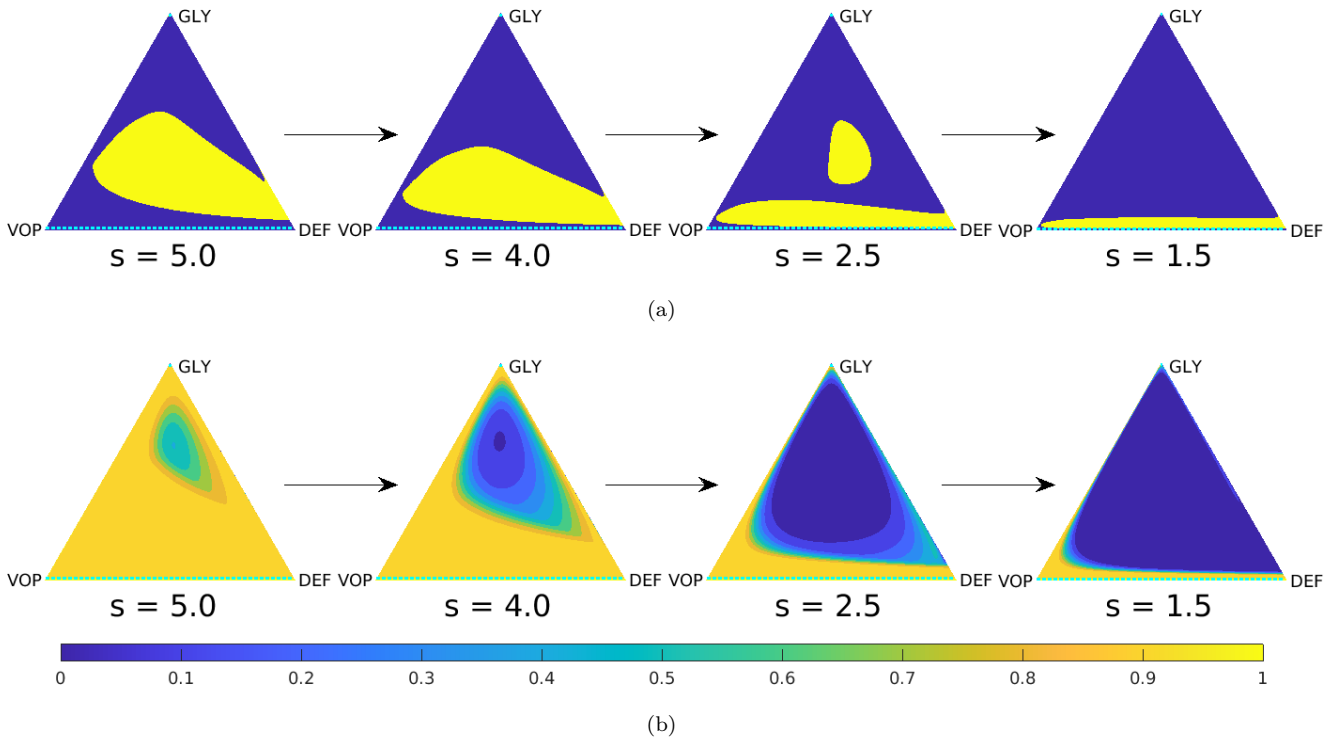


Figure 3: Representative slices of threshold-aware policy and their corresponding optimal probability of success: (a) optimal policy; (b) optimal probability of success. Each triangle represents the optimal policy/probability of success corresponding to a specific threshold value s, which is shown below the the triangle. The arrows indicate one should go from higher threshold to lower threshold s values.

3.2 Different thresholds and different initial configurations

In Figure 2(C), we have already provided a preliminary comparison of outcomes under the deterministic-optimal policy developed in [15]. Here, we include additional simulations that highlight the increased probability of success achieved under our threshold-aware policies described in section 2.2. In section 6S of Supplementary Materials, we show that the advantages of threshold-aware policies can be even more significant when the volatilities $(\sigma_{\mathrm{G}}, \sigma_{\mathrm{D}}, \sigma_{\mathrm{V}})$ are higher.

We consider two initial tumor configurations at the same GLY-level ($p_0 = 0.4$). The first one, based on $q_0 = 0.27$ as shown in Figure 5, is chosen to ensure that the initial configuration is inside the yellow (drugs-on) region prescribed by the deterministic-optimal policy. The second, based on $q_0 = 0.8$ and shown in Figure 6, starts in the blue (drugs-off) region. In both Figures 5(C) and 6(C), the CDFs generated with the deterministic-optimal policy are shown in *blue* and the CDFs corresponding to our threshold-aware policies are shown in *red* and *green*. The details of our random simulations used to built these can be found in section 5S.3 of Supplementary Materials.

In Figure 5, under the deterministic-optimal policy only a half of the simulations yield the cost not exceeding 4.71. Our threshold-aware policy (red), maximizes this $\mathbb{P}(\mathcal{J} \leq 4.71)$ to achieve 64%. The potential for improvement is even more significant with lower threshold values. For instance, the deterministic-optimal policy CDF yields the probability of recovery under $\bar{s} = 4.35$ as less than 10%, while our threshold-aware policy (green) ensures that $\mathbb{P}(\mathcal{J} \leq 4.35) \approx 46\%$. The improvement can be also translated to simple medical terms: starting from this initial tumor configuration, the deterministic-optimal policy will likely keep using the drugs at the maximum

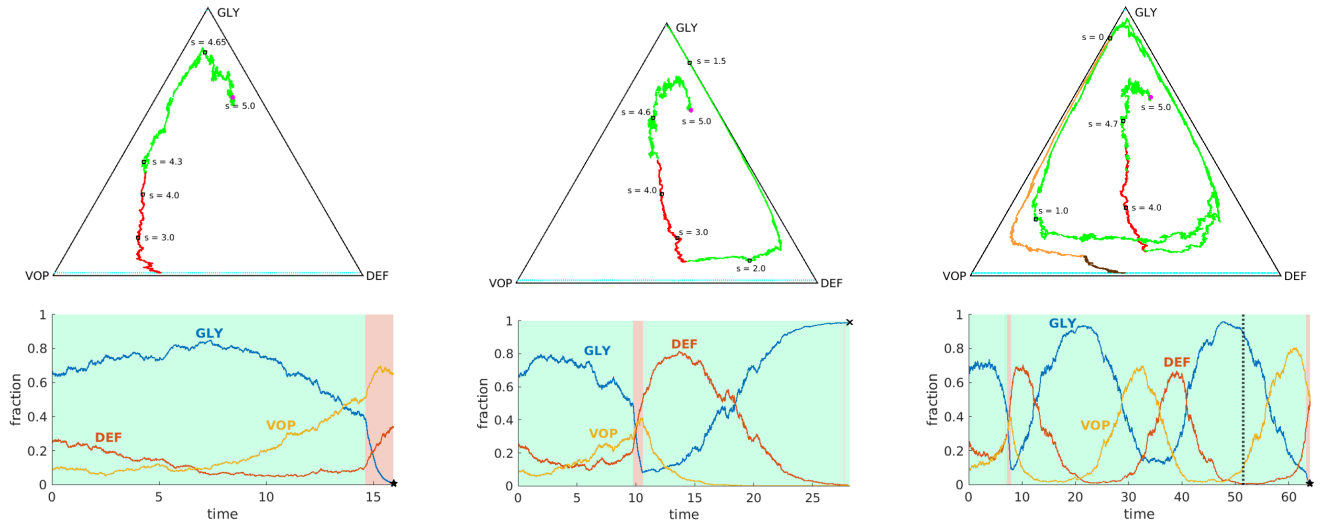
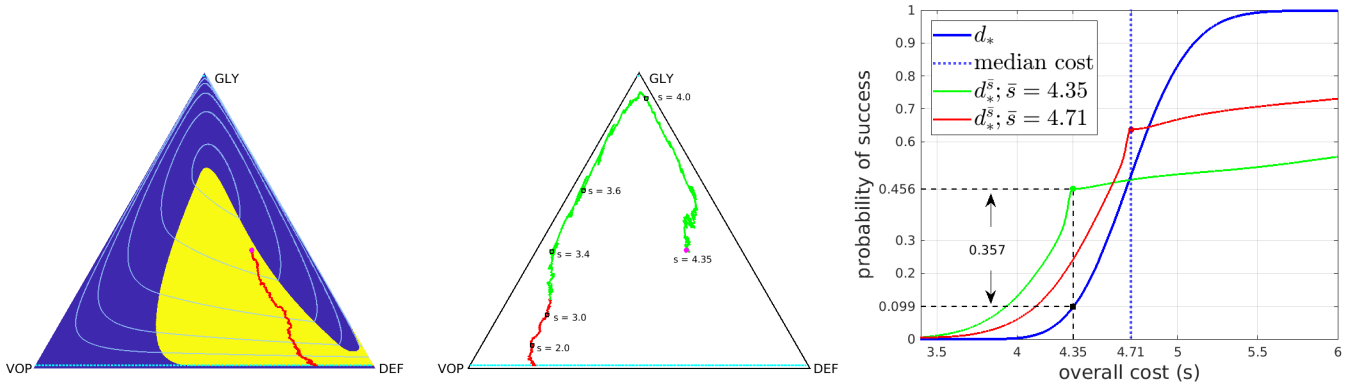


Figure 4: Representative sample paths starting from the same initial state $(q_0, p_0) = (0.26, 0.665)$ (magenta dot) and the same initial budget $\bar{s} = 5$. **Top row:** sample paths on a GLY-DEF-VOP triangle. (a) eventual recovery with a cost of 4.70 (within the budget); (b) eventual death; (c) failure by running out of budget (eventual recovery 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 success and death, respectively. In (c), we use a *dashed black* line to indicate the time of exhaustion of our budget.

rate $d_{\max}$ all the way to recovery; 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. For our second example (Figure 6), the threshold-aware policy (red) also 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 the deterministic-optimal policy. When starting from a lower initial budget $\bar{s} = 3.45$, the deterministic-optimal policy can only yield a 15.6% chance of recovery, while our threshold-aware policy (green) doubles this $\mathbb{P}(\mathcal{J} \leq 3.45)$ to 31.5%. We can see from Figure 6(A) that the deterministic-optimal policy basically prescribes drugs till recovery once the (random) tumor state enters the yellow region. In contrast, Figure 6(B) shows that our threshold-aware optimal policy opens the possibility of prescribing drugs for separate time intervals over the entire treatment period. This again yields a reduction in the total amount of drugs used until recovery.

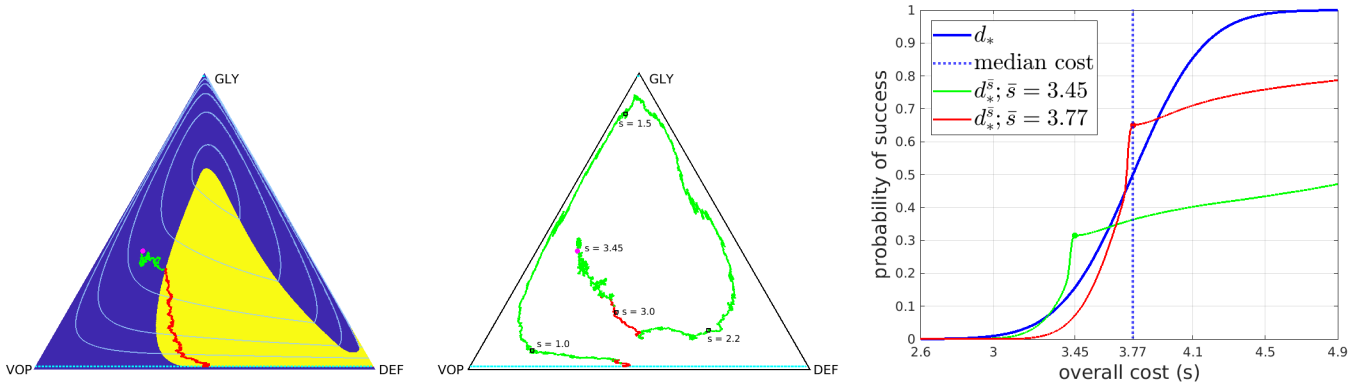
It is worth noting that each threshold-aware policy maximizes the probability of success for a single/specific threshold value only. E.g., for all the red/green CDFs we have provided, the probability of success is only maximized at those red/green dots. Moreover, we clearly see from Figure 5(C) that the probability of not exceeding *any* $s \leq 4.35$ is lower on the red CDF than on the green CDF (computed for $\bar{s} = 4.35$). Intuitively, this is not too surprising. In the early stages of treatment, a (red) 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 (green) policy starting from a lower initial budget $\bar{s} = 4.35$; see also the budget-dependent size of drugs-on regions in Figure 3(A). So, starting with a higher budget often makes the probability of a low-cost success much lower.

Additionally, we observe that all red and green CDFs become much flatter for all $s \geq \bar{s}$. This evidently shows that switching to the deterministic-optimal policy after the budget runs out lowers the effectiveness of the treatment under stochastic tumor dynamics.



(a) (b) (c)

Figure 5: Comparison between threshold-aware policies and deterministic-optimal policy Case I: starting from an initial state $(q_0, p_0) = (0.27, 0.4)$ (magenta dot): (a) a sample path with cost 4.75 under deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 4.35$ with a total cost of 4.02 under 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 green* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.35$; and the *solid red* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 4.71$.



(a) (b) (c)

Figure 6: Comparison between threshold-aware policies and deterministic-optimal policy 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 deterministic-optimal policy; (b) a sample path starting at $\bar{s} = 3.45$ with a total cost of 3.41 under 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 green* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 3.45$; and the *solid red* curve is the CDF generated with the threshold-aware policy with $\bar{s} = 3.77$.

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*[21] and *in vivo*[9, 27], however, have begun to provide methods for more rigorous quantification of these interactions. As these tools mature, the goal of understanding these interactions in patients begins to become feasible.

In our previous work, we showed how, given appropriate information, a mathematical model of the tumor growth and interactions could be used together with dynamic programming to calculate a formal optimal solution [15]. That work however, while a step forward in adaptive therapy, did not consider the reality of the underlying stochasticity of tumor biology and evolution: to make mathematical progress, we had to simplify the situation to one that was deterministic. In the present work, we build on our previous advance to include stochasticity in growth rates in time, and show that stochastic optimal control theory can be applied to maximize the probability of recovery under a given threshold cumulative cost (interpreted as a combination of the total drugs used and the time to recovery). We show that the optimal treatment policy becomes *threshold-aware*. Even though the optimal treatment policy is still *bang-bang*, the drugs-on/off regions vary from one threshold value to another. We further compare CDFs generated with the deterministic-optimal policy and threshold-aware policies at different threshold values. We show that for any tested starting configurations, our threshold-aware policies offer significant advantages over the deterministic-optimal policy and often result in a significant reduction of drugs used to treat the patient.

As mentioned in Gluzman et al.[15], a limitation of the underlying EGT model is that only proportions of cancer cells are considered. An immediate mathematical extension of our work is to bring the full tumor size or the actual sizes of each type of tumor cells into the model; e.g., see [28] and [22, Appendix C.2]. In principle, the same extension can be also used in our stochastic setting though the computational cost will increase accordingly. 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. We also note that maximizing the probability of success under a given threshold cost is not the only choice for stochastic models. Minimizing the probability of failure or the expected total treatment time to recovery are among other potential choices. It will be interesting to study the tradeoffs between these as well as the multiobjective control problem of optimizing threshold-aware policies for two different threshold values simultaneously.

In summary, we have presented a theoretical 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 empirical designs and are not formulated to consider the underlying stochasticities. Developing theoretical foundation for future clinical studies requires EGT models directly grounded in objectively measurable biology [21]. Therapy optimization based

on such models requires efficient computational methods, particularly in the presence of stochastic perturbations. We hope that the 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. MW and AV would like to acknowledge funding from the National Science Foundation (DMS-1645643, DMS-1738010, and DMS-2111522). JGS would like to acknowledge funding from the National Cancer Institute (R37 CA244613) and the American Cancer Society through their Research Scholar Grant (132691-RSG-20-096-01).

References

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

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

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

[4] Cartee, E., Nellis, A., Van Hook, J., Farah, A., & Vladimirsky, A. (2021). Quantifying and managing uncertainty in piecewise-deterministic Markov processes. *preprint: https://arxiv.org/abs/2008.00555*.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] Farrokhian, N., Maltas, J., Ellsworth, P., Durmaz, A., Dinh, M., Hitomi, M., Kaznatcheev, A., Marusyk, A., & Scott, J. G. (2020). Dose dependent evolutionary game dynamics modulate competitive release in cancer therapy. *bioRxiv*.

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

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

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

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

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

[27] 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.

[28] Melbinger, A., Cremer, J., & Frey, E. (2010). Evolutionary game theory in growing populations. *Physical review letters*, 105(17):178101.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

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

[36] 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.

[37] 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.