Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory

Mark Gluzman[1], Jacob G. Scott[2], and Alexander Vladimirsky*[,3]

[1]Center for Applied Mathematics, Cornell University, Ithaca, NY
[2]Department of Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH
[3]Department of Mathematics and Center for Applied Mathematics, Cornell University, Ithaca, NY

Abstract

BACKGROUND: Recent clinical trials have shown that the adaptive drug therapy can be more efficient than a standard MTD-based policy in treatment of cancer patients. The adaptive therapy paradigm is not based on a preset schedule; instead, the doses are administered based on the current state of tumor. But the adaptive treatment policies examined so far have been largely *ad hoc*. In this paper we propose a method for systematically optimizing the rules of adaptive policies based on an Evolutionary Game Theory model of cancer dynamics.

METHODS: Given a set of treatment objectives, we use the framework of dynamic programming to find the optimal treatment strategies. In particular, we optimize the total drug usage and time to recovery by solving a Hamilton-Jacobi-Bellman equation based on a mathematical model of tumor evolution.

RESULTS: We compare adaptive/optimal treatment strategy with MTD-based treatment policy. We show that optimal treatment strategies can dramatically decrease the total amount of drugs prescribed as well as increase the fraction of initial tumour states from which the recovery is possible. We also examine the optimization trade-offs between the total administered drugs and recovery time.

CONCLUSIONS: The adaptive therapy combined with optimal control theory is a promising concept in the cancer treatment and should be integrated into clinical trial design.

Keywords: adaptive therapy; optimal treatment policy; evolutionary game theory; Hamilton-Jacobi-Bellman equation; heterogeneity

1 Introduction

Intratumoral heterogeneity is being increasingly recognized as a cause of metastasis, progression and resistance to therapy[1]. While genetic instability, a hallmark of malignancy[2], can result in this heterogeneity, it is being increasingly understood that eco-evolutionary factors, like selection and clonal interference, can also drive and maintain it[3,4].

While sequencing technologies have enabled increasingly in-depth quantitative understanding of the genetic heterogeneity, relatively little experimental work has sought to directly quantify the eco-evolutionary interactions involved. As more studies come to light showing the efficacy of treatments based on eco-evolutionary trial designs, this lack of quantification is coming into focus.

In line with standard, cell-autonomous growth-based theories, conventional chemotherapy is given to patients at the *maximum tolerated doses* (MTD): the highest doses that most patients can safely tolerate. Although the MTD-based chemotherapy offers advantages in survival compared to no therapy, cures remain elusive, and side effects can be severe. In addition to the toxicity, it is known that relapse is nearly inevitable due to the emergence of therapeutic resistance: a process driven by Darwinian evolutionary dynamics in which the MTD-based chemotherapy kills off the chemotherapy-sensitive cells and chemo-refractory cells eventually dominate in

*Corresponding author: vladimirsky@cornell.edu

the tumor. While it is unknown whether these resistant cells are present before therapy, or acquire resistance mutations during therapy, it is the process of variation and selection under standard therapy that drives the inevitable failure in the patient.

Metronomic chemotherapy has been proposed as a possible alternative to the MTD strategy [5,6]. Metronomic chemotherapy is given in an on/off fashion at frequent time intervals according to a set periodic schedule. The idea behind this method is to give less overall therapy, thereby increasing tolerability, and allowing time for therapy sensitive cells to regrow, allowing for resensitization of the tumor. Frustratingly, the results of clinical trials of metronomic chemotherapy have been "variable" [7] and many of them have *not* demonstrated significant efficacy [8–10] as compared to standard therapy. Recent studies argue that metronomic chemotherapy highly depends on timing, and the right scheduling can improve the results of its usage [11].

Based on the hypothesis that disease dynamics depend on the evolution of tumor heterogeneity as modulated by competition between subtypes, the idea of using *adaptive therapy* (AT) has been proposed [12]. AT is much like metronomic therapy, with an important difference. AT administers doses of therapy according to the current *state* of tumor growth and its anticipated evolutionary changes (*trajectory*). These can be estimated using direct (e.g., taking biopsies) or indirect (e.g., antigen testing, mathematical modeling) methods. Therefore, unlike the MTD-based or metronomic protocols, AT does not have preset schedule and it adjusts the doses and timing before a tumor becomes chemotherapy-resistant, prolonging time to this event. Recently, the adaptive strategies have shown promise in pre-clinical trials of breast cancer [13] and a phase 2 clinical trial in metastatic castrate-resistant prostate cancer [14].

These two recent successes [13,14] in adaptive therapy have been based on mathematical modeling of tumor evolution under therapy using a dynamical systems approach based on Evolutionary Game Theory (EGT) [15,16]. This formalism explicitly considers interactions between sub-populations and models their fitness in frequency dependent terms. EGT has been used to theoretically consider many scenarios in cancer before, including therapy scheduling and timing in prostate cancer [14,17–19]; the use of tumor microenvironment targeting therapy in glioblastoma [20]; the trade-off between healthy tissue and cancer in multiple myeloma [21,22]; and drug resistance in general [23–25]. These theoretical studies, combined with the recent empiric realizations, suggest significant opportunities to improve therapy by using this evolutionarily enlightened approach. Nevertheless, therapeutic decisions in general practice are currently *not* based on this knowledge and continue to use the MTD paradigm.

Assuming that an oncologist has perfect information about the current state of a tumor, and a faithful mathematical model that can predict its trajectory (these, of course, are very strong assumptions), it is not clear how he/she should adjust the schedule and doses. Based on a stage of the disease and patient's needs, the therapy can have different final goals: maximization of patient's duration of life, ensuring the best possible quality of the rest of life, decreasing probability of new metastases appearing, decreasing time/cost of the treatment, etc.. Unfortunately, an oncologist can usually only focus on one or two of these goals, having some reasonable constraints on the secondary parameters. Thus, an important step toward optimizing AT is to define an *objective* of the therapy and "translate" it into mathematical language. The next step is to *quantify* how good each particular strategy is with respect to that chosen objective. Optimizing this objective is a mathematical goal which can be addressed by the tools of optimal control theory.

Optimal control theory, a branch of mathematics typically applied to problems in engineering, can be applied to a wide class of problems arising in oncology [26]. The first application of optimal control theory in cancer was done by Swan and Vincent [27] who found the optimal treatment strategy for multiple myeloma with the objective to minimize the total amount of drugs used applying the Pontryagin Minimum Principle (PMP) [28]. Since then, others have used the PMP to different optimal cancer treatment problems: a chemotherapy optimization under evolving drug resistance [29–31], optimal scheduling of a vessel disruptive agent [32], MAPK inhibitors [33] input in cancer treatment, minimization amount of drugs prescribed in tumorimmune model [34], finding compromise between drug toxicity and tumor repression for the myeloma bone disease [35], and many others.

While these approaches have offered benefits in their ability to formally optimize problems written as dynamical systems, the PMP method has several limitations. First, PMP yields only a necessary condition for an optimum, and any *locally* optimal trajectory of the control system satisfies PMP. Local optimality means that the trajectory is optimal when comparing it with its small perturbations, but there may well be a different trajectory that is even better (*globally* optimal, compared with *all* possible trajectories). Secondly, PMP provides a time-dependent (open-loop) control: given an initial state, the method provides an optimal treatment strategy as a function of time – therefore a treating oncologist has to follow it regardless of the changing state of the tumor. However, if the underlying model has been perturbed or includes some noise (like a tumor acquiring mutations, say), the control cannot adapt to these unexpected changes.

A different approach to analysis of an optimal control problem is a feedback (closed-loop) control point of view. Using the Hamilton-Jacobi-Bellman (HJB) equations, one can obtain controls that depend on the current state of the dynamical system (current distribution of sub-populations of cancer cells) rather than only the current time[36,37]. In this case, the treating oncologist's decisions can be adjusted if something unexpected has happened with the trajectory. Moreover, the HJB equations guarantees that the resulting treatment feedback strategies are *globally* optimal. Despite these advantages of the HJB over PMP method, there are few works[38,39] which use the feedback control paradigm to find an optimal treatment strategy.

Here, we apply the HJB approach to solve for optimal treatment strategies for a model of lung cancer proposed by Kaznatcheev and colleagues[40]. In that paper the authors introduce an evolutionary game (system of replicator-type equations) that models the dynamics of three sub-populations of tumor cells. The article highlights the importance of a good scheduling in the polyclonal regime, when the game has cyclic dynamics. The article has an example of two different scheduling strategies with the same set of initial parameters that lead the system to opposite outcomes: putative recovery, versus putative death of a patient. While several qualitatively different treatment schedules are presented, optimal therapy is not discussed. Given the growing interest in EGT in clinical applications[14] and recent work connecting these models using direct *in vitro* parameterization[41], we believe the optimization of therapies based on such models will become increasingly important and the HJB-based approach will be used far more often in the future.

2 Methods

In this article we focus on a model of cancer evolution that has been proposed in Kaznatcheev et al.[40] and summarized in Box 1. This model considers interactions between three different sub-populations of cancer cells playing a modified version of the public goods social dilemma: glycoltyic cells (GLY), vacscular overproducers (VOP) and cells called defectors (DEF) which use both strategies to "cheat" on the others. GLY cells are anaerobic and produce lactic acid. Both VOP and DEF cells are aerobic. In addition, VOP cells spend extra energy to produce VEGF (a protein that improves the vasculature, benefiting both VOP and DEF). Based on the replicator model from EGT[15,16] summarized in equations (1) and (3), the evolution of the tumor is described by tracking the changing *proportions* of GLY, VOP and DEF cells in the full population. The patient is viewed as recovered when the GLY proportion falls below some low threshold r_b. (Below this *recovery barrier*, the validity of replicator-based model is harder to justify and we assume that the GLY cells are essentially extinct.) Conversely, we assume that GLY cells *suppress* other tumor cells and a patient dies if the total proportion of aerobic cells (VOP and DEF sub-populations combined) falls below some low threshold (a *failure barrier*) f_b.

For a range of parameter values (4), this model predicts a heterogeneous regime* in cancer evolution with coexistent and oscillating proportions of GLY, VOP, and DEF. Without any treatment, these sub-populations follow cyclic dynamics and a patient never recovers; see Figure 1(a).

Following Section 4.1. in Kaznatcheev et al.[40] we consider a cell-type-targeting therapy that preferentially penalizes the fitness of GLY cells; see formula (7) in Box 2. During the treatment, a doctor defines the timing of therapy and its intensity. This time-dependent intensity $d(t)$ can vary between 0 (no therapy) and $d_{max} > 0$ (the maximum tolerated dose, MTD). Two extreme case ($d(t) = 0$ versus $d(t) = d_{max}$ for all times t) are illustrated in Figures 1(a) and 1(b) respectively. In the latter case, GLY cells become extinct and the patient recovers quickly, however it is natural to ask whether this treatment strategy is optimal in some sense (i.e. could the recovery be much delayed if the patient received therapy less often or at a lower intensity?). In the following section we will show that the MTD-based treatment can lead to an avoidably high cumulative amount of therapy (see Figure 2(c) or might even fail to achieve a recovery in situations where AT-based treatment would otherwise have succeeded (see Figure 6), much like the early results from Zhang et al. in metastatic prostate cancer[14].

*Other tumor regimes (fully angiogenic and glycolyctic) also exist outside of this parameter range. They are less interesting from the point of view of treatment strategies, but we still consider them for the sake of completeness in Section 6.3 of Supplementary Materials.

Box 1: Mathematical model of the cancer sub-populations evolution

Subpopulation Proportions:
(x_G, x_D, x_V)
for GLY, DEF, and VOP respectively.

Note: $x_G + x_D + x_V = 1$.

Relative Subpopulation Fitness: (ψ_G, ψ_D, ψ_V)
defined in Materials and Methods of Kaznatcheev et al.[40]).

Full Population Averaged Fitness:
$< \psi >:= x_G\psi_G + x_D\psi_D + x_V\psi_V$.

Replicator equations to model the evolution of sub-populations:

$$\begin{cases} \dot{x_D} = x_D(\psi_D - < \psi >) \\ \dot{x_G} = x_G(\psi_G - < \psi >) \\ \dot{x_V} = x_V(\psi_V - < \psi >) \end{cases} \tag{1}$$

Transformation/Reduction to a 2D system:

$$\begin{cases} q = \frac{x_V}{x_V + x_D} \\ p = x_G \end{cases} \quad \text{or} \quad \begin{cases} x_D = (1-q)(1-p) \\ x_G = p \\ x_V = (1-p)q \end{cases} \tag{2}$$

Equivalent dynamics of (1) in reduced coordinates (see Appendix B in[40]):

$$\begin{cases} \dot{q} = q(1-q)\left(\frac{b_v}{n+1}\sum_{k=0}^{n} p^k - c\right) \\ \dot{p} = p(1-p)\left(\frac{b_a}{n+1} - (b_v - c)q\right) \end{cases} \tag{3}$$

Parameters:
- 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, the number of glycolytic (GLY) cells in the interaction group.

Conditions for homogeneous center equilibrium and periodic oscillations:

$$\frac{b_a}{n+1} < b_v - c < cn. \tag{4}$$

Process *terminates* as soon as either

$$\begin{cases} p(t) < r_b, & \text{if therapy succeeds;} \\ p(t) > 1 - f_b, & \text{if therapy fails.} \end{cases} \tag{5}$$

***Terminal set*:**

$$\Delta = \Big\{(q,p) \in [0,1] \times [0,1] : p < r_b \text{ or } p > 1 - f_b\Big\}. \tag{6}$$

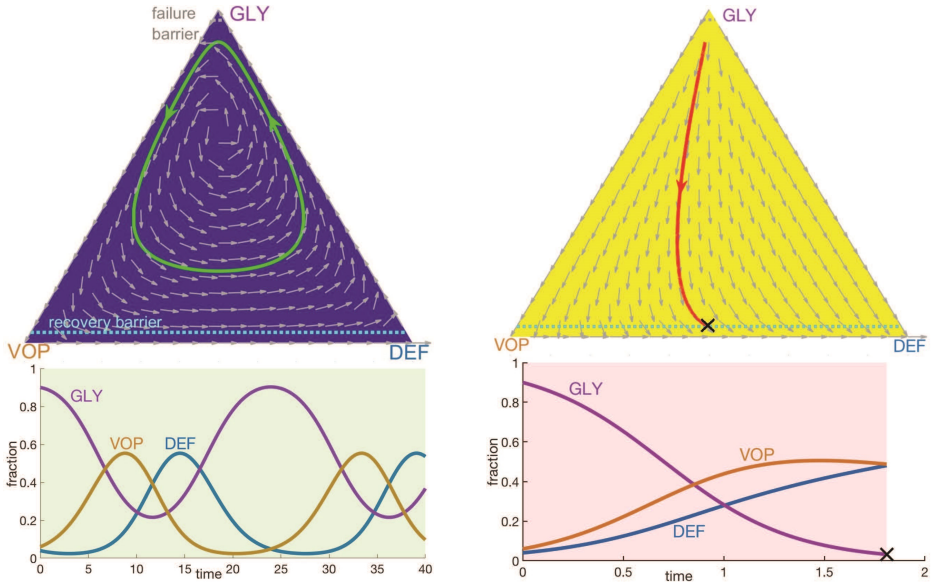


Figure 1: A comparison of two possible constant treatment scenarios starting from an initial state $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$: (a) without any therapy; (b) with the MTD-based therapy.
Top row: phase portraits of corresponding vector fields (shown by *gray arrows*) on a GLY-VOP-DEF triangle with illustrative trajectories. *Blue background* and *green reference trajectory* – no therapy at all. *Yellow background* and *red reference trajectory* – MTD-based therapy at all times. *Dash light blue* and *gray lines* separate the recovery zone (bottom) and the failure zone (top) respectively. *Black cross* – termination due to crossing the recovery barrier.
Bottom row: evolution of sub-populations with respect to time based on the reference trajectories above. *Green time range* – no therapy. *Pink time range* – MTD-based therapy. *Black cross* – termination due to crossing the failure or recovery barrier by GLY cells. Note the different scaling of the time axis.
Parameters: Following Figure 2 in[40], $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$ and $d_{max} = 3$ for (b). The recovery and failure barriers are $r_b = f_b = 10^{-1.5}$.

Box 2: Evolution dynamics with control on therapy intensity

Time-dependent intensity of GLY-targeting therapy: $d : \mathbb{R}_+ \to [0, d_{max}]$.

Evolutionary dynamics as a controlled system:

$$\begin{cases} \dot{q}(t) = q(t)\Big(1 - q(t)\Big)\Big(\frac{b_v}{n+1}\sum\limits_{k=0}^{n} p^k(t) - c\Big), \\ \dot{p}(t) = p(t)\Big(1 - p(t)\Big)\Big(\frac{b_a}{n+1} - (b_v - c)q(t) - \boldsymbol{d(t)}\Big); \\ q(0) = q_0,\ p(0) = p_0. \end{cases} \tag{7}$$

One natural objective function to minimize is the total amount of therapy administered over the course of treatment (which in this case could be a surrogate for both toxicity and cost). This can be quantified as $D = \int_0^T d(t)dt$, where the total time of treatment T is dependent on the initial cancer subpopulation fractions and on our chosen therapy policy $d(\cdot)$. However, this objective is problematic for two reasons. First, the minimum of D is clearly attained without any therapy (taking $d(t) \equiv 0$ implies $D = 0$ and $T = +\infty$), even though the dynamics become cyclic and the recovery is never achieved. Second, if we constrain our minimization to only those $d(\cdot)$ that lead to recovery, an optimal treatment policy does not exist. Instead, there is a sequence of treatment

policies that lead to successively smaller d values but with an unbounded increase in corresponding treatment times T. The idea of such policies is simple: travel along the therapy-free trajectories of Figure 1(a) for most of the time, but use short bursts of therapy only when the drugs are most effective. To approach the optimally small d, one would need to use shorter and shorter bursts, resulting in policies that are hard to implement in practice and would require unrealistically long treatment times T, but would yield a situation like a chronic disease, where while the tumor is never cured, it is always controlled.

In order to get a meaningful optimal policy we will penalize the treatment time by a *time penalty* $\sigma > 0$. The total time spent on the treatment, including time when a patient skips the doses, is an important factor by itself. Much longer time spent on a treatment causes worse quality of life and additional costs for a patient. Therefore, our objective is to minimize the sum of a *therapy cost* D and a *treatment time cost* σT, while guaranteeing the treatment results in recovery. For every choice of $\sigma > 0$, the resulting treatment policies are thus *Pareto-optimal* with respect to D and T.

In this paper we consider an *objective function* that is equal to $\int_0^T d(t)\,dt + \sigma T$, if a patient recovers, and is equal to $+\infty$ otherwise. The *value function*, u, is defined as the minimum of this objective function (over the set of treatment policies), and any policy $d(\cdot)$ that realizes this minimum is called *optimal*.

Due to the structure of this optimization problem, one can show that virtually all optimal treatment policies are *bang-bang*: at any given time t, they either administer therapy at MTD-rate ($d(t) = d_{max}$) or administer no drugs at all ($d(t) = 0$); see Section 6.1 in Supplementary Materials. For such policies, the objective function becomes a weighted sum of the total *therapy time* $\widetilde{T}$ (when $d(t) \equiv d_{max}$) and the total treatment time T, with d_{max} and σ as the corresponding weights. Moreover, this allows for a simple visual representation of any such policy: splitting the full state space into two parts (MTD dynamics vs. no-therapy dynamics), shown in yellow and blue respectively in all figures throughout this paper, and simplifies application of therapy into something familiar to clinicians: on or off.

3 Results

3.1 Quantifying the benefits of optimal treatment strategies

In this section we apply the optimal control theory to an example considered in Kaznatcheev et al.[40]. The details of the optimal control problem formulation are summarized in Box 3. We take the same model parameters as in Figure 2 of Kaznatcheev et al.[40]: $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$, strength of MTD $d_{max} = 3$, and initial state $(q_0, p_0) = (0.6, 0.9)$, which by formula (2) is equivalent to $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$. We also use $\sigma = 0.01$ to incorporate a time-penalty absent in the original model. We take $r_b = f_b = 10^{-1.5}$ as recovery and failure barriers[†]

In Figure 2 we compare the treatment cost (10) and treatment time (8) of trajectories corresponding to four different treatment strategies.

The first two strategies we consider are similar to those modeled in Kaznatcheev et al.[40]: 2(a) is an example of a bad policy that may cause a failure by stopping the therapy prematurely, while 2(b) is a good policy based on ad-hoc adjustment of the start time for the therapy. We also illustrate a "MTD-based"' policy 2(c), which is analogous to the standard of care using MTD as long as possible. Even though both 2(b) and 2(c) lead to recovery, neither of these is optimal (with the MTD-based approach resulting in excessive amount of drugs, captured by the higher cost). The policy minimizing our objective function can be found by solving the HJB equation; we illustrate it in 2(d).

[†]This change in parameter values is meant to decrease the computational cost of our numerical approach (see section 6.2 in Supplementary Materials.) The original $r_b = f_b = 10^{-4}$ from Kaznatcheev et al.[40] would require computations on a finer mesh.

Box 3: Objective function

Terminal time (or the total treatment time):,

$$T(q_0, p_0, d(\cdot)) = \min\Big\{t \in \mathbb{R}_+ \mid (q(t), p(t)) \in \Delta,\ q(0) = q_0,\ p(0) = p_0\Big\}. \tag{8}$$

If the system never gets to the terminal set, we assume that $T(q_0, p_0, d(\cdot)) = +\infty$.
Terminal cost function is $g(q, p) : \Delta \to \{0, +\infty\}$ s.t.

$$g(q,p) = \begin{cases} +\infty, & \text{if } p > 1 - f_b, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Let $T := T(q_0, p_0, d(\cdot))$ be the terminal time.
Treatment cost (objective) function:

$$J(q_0, p_0, d(\cdot)) = \int_0^T (d(s) + \sigma)ds + g\Big(q\big(T\big), p\big(T\big)\Big). \tag{10}$$

J is finite if the system (7) terminates at the recovery barrier, and is infinite otherwise (i.e., if the system terminates at the failure barrier or does not terminate at all).
Value function:

$$u(q_0, p_0) = \inf_{d(\cdot)} J(q_0, p_0, d(\cdot)) \tag{11}$$

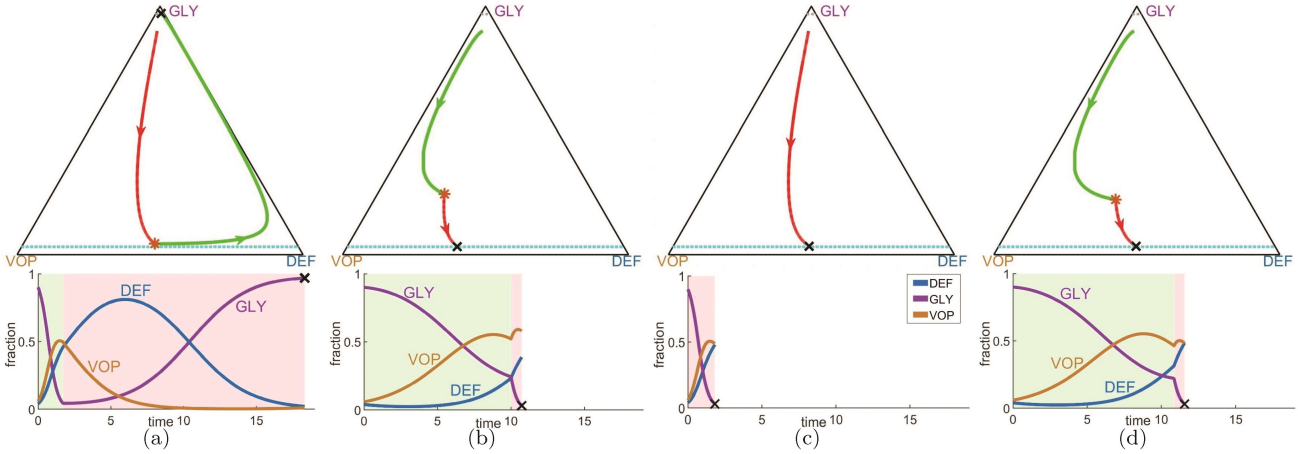
can be found by solving Hamilton-Jacobi-Bellman **(HJB)** PDE:

$$\min_{d\in[0,d_{max}]} \left\{\nabla u(q,p) \cdot \begin{pmatrix} \dot{q}(q,p,d) \\ \dot{p}(q,p,d) \end{pmatrix} + d + \sigma\right\} = 0, \qquad (q,p) \in ([0,1] \times [0,1])/\Delta. \tag{12}$$

The boundary conditions of HJB equation:

$$\begin{cases} u(q,p) = 0, & \text{if } p < r_b; \\ u(q,p) = +\infty, & \text{if } p > 1 - f_b. \end{cases} \tag{13}$$

Once u and its gradient are found through a numerical approximation, they can be used to obtain the optimal control *in feedback form:* $d^* = d(q,p)$.



Subfigure	Policy	Total time	Time till switching (denoted by *)	Overall cost
(a)	"treat immediately but not long enough"	18.32	1.7	$+\infty$
(b)	"start treating after 10 time steps"	10.73	10	2.29
(c)	"MTD-based"	1.81	—	5.45
(d)	"optimal"	11.56	10.86	2.25

Figure 2: The importance of optimal scheduling for drug therapy: ensuring recovery and decreasing the cost of treatment.
Tumor evolution under four different treatment strategies for the same initial state $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$. A seemingly reasonable treatment strategy may not lead to a recovery; see subfigure (d). Even if a patient eventually recovers (as in subfigures (b) and (c), the overall cost of treatment can be reduced by pursuing a provably optimal policy (d).
Top row: tumor evolutionary trajectories under different strategies on a GLY-VOP-DEF triangle. *Green part of a trajectory* – no therapy is used. *Red part of a trajectory* – MTD-based (standard) therapy. The moment of switching from one regime to another is denoted by (red *). *Dash light blue* and *gray lines* separate the recovery zone and the failure zone respectively. *Black cross* – termination due to crossing the failure or recovery barrier.
Bottom row: evolution of sub-populations with respect to time based on the reference trajectories above. *Green time range* – no therapy. *Pink time range* – MTD-based therapy. *Black cross* – termination due to crossing the failure or recovery barrier by GLY cells.
Parameters: Following Figure 2 in [40] and Figure 1, we use the initial state $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$; game parameters: $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$; the MTD rate: $d_{max} = 3$; the recovery and failure barriers: $r_b = f_b = 10^{-1.5}$; and the time-penalty: $\sigma = 0.01$.

The corresponding therapy on/off regions and the resulting vector field are shown in Figure 3(a). The zoomed version shows that trajectories can be prevented from crossing the failure barrier by using the MTDs just before crossing. In fact, a *chattering control* (with intermittent and sufficiently frequent use of MTDs) would be sufficient to guarantee this.

The level sets of u in Figure 3(b) show that value functions need not be smooth. Since the gradient of u is used to determine the optimal course of action (therapy on or off), there can actually be more than one optimal policy for initial states on a *shockline* (where that gradient is undefined). We show an example of such trajectories for an initial point $(x_D, x_G, x_V) \approx (0.417, 0.311, 0.272)$ (solid green and red lines in Figure 3(b)). Both trajectories yield the same cost of 2.764. Moreover, non-smoothness of the value function often poses a challenge for method's based on Pontryagin Maximum Principle (PMP) [28] even if the initial state *is not* on a shockline. For example, perturbing the initial state to a nearby $(x_D, x_G, x_V) = (0.35, 0.3, 0.35)$, denoted by a cross in Figure 3(c), one sees two *locally optimal* trajectories and PMP might yield either of these depending on the initial guess. The green one is however inferior to the *globally* optimal red trajectory, which is always recovered by solving the HJB equation.

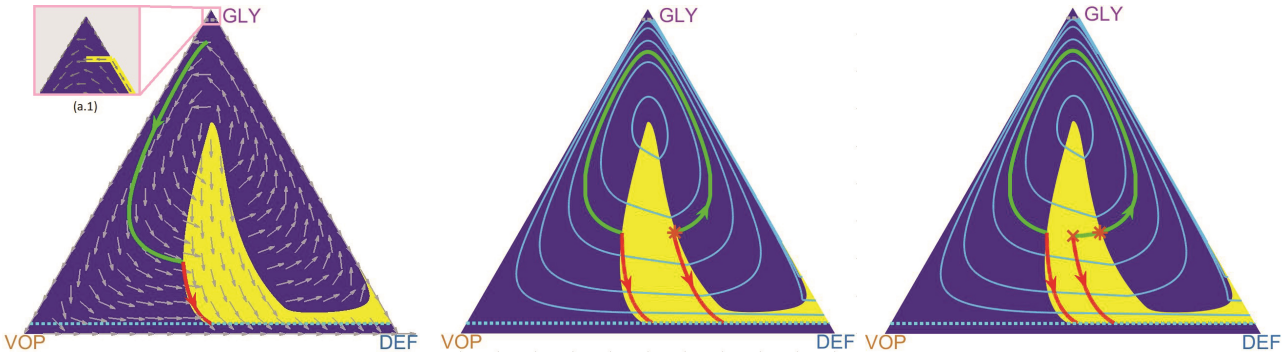
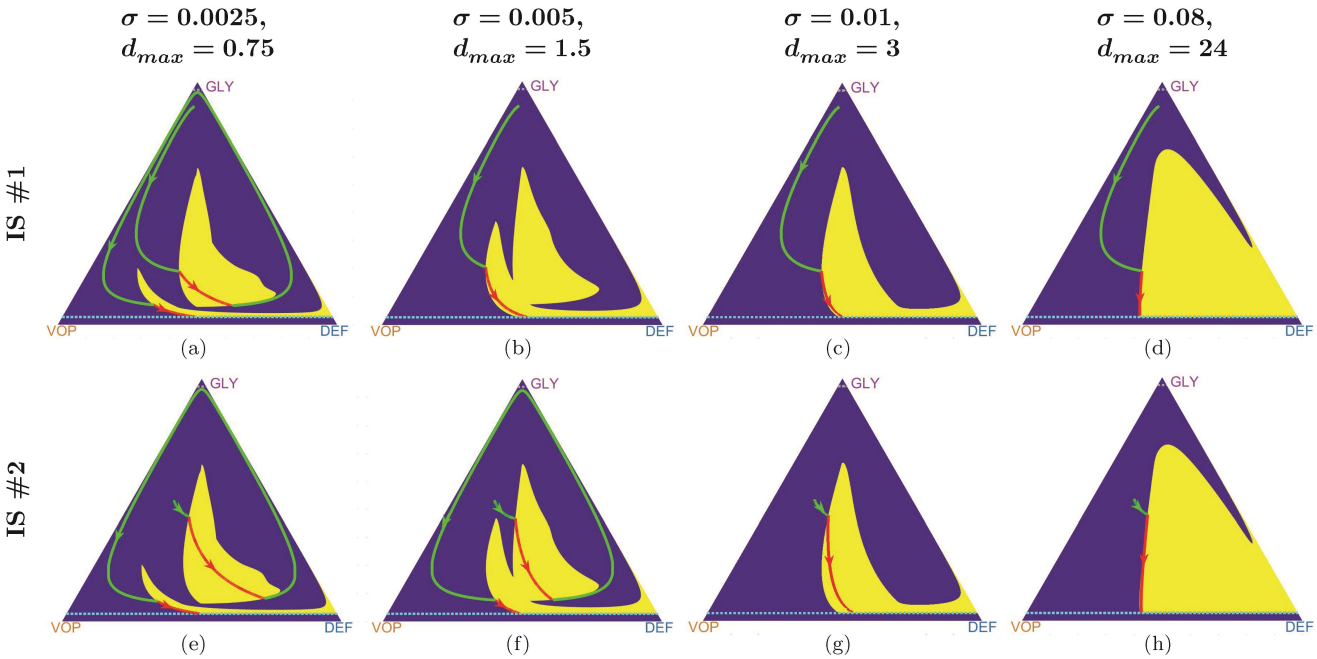


Figure 3: Optimal control in feedback form, the value function, and the pitfalls of PMP.
(a): A phase portrait of the optimal system dynamics. The vector field is shown by *gray arrows* over the optimal drugs-off (*blue background*) and drugs-on (*yellow background*) regions. A sample optimal trajectory (in green and red) corresponds to the initial state from Figure 2.
(b): Computation of the value function u (whose level curves are shown by *light blue lines*) is used to determine the optimal drugs-on and drugs-off regions (shown in yellow and dark blue respectively). Optimal trajectories are not unique for initial states on the shockline (where the level curves of u are not smooth). Two such optimal trajectories are shown starting from an asterisk (*).
The green-red trajectory takes longer to reach the recovery, but uses less drugs than the red (start-drugs-right-away) trajectory. The cumulative cost is the same for both of them.
(c): For initial conditions off the shocklines of u, there can still be multiple *locally optimal* trajectories. We show an example of two such trajectories starting from a cross marker ($\times$). The risk of applying the PMP method is that it might yield either of them, but only the red (start-drugs-right-away) is globally optimal.
Parameters: $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$, $d_{max} = 3$, $\sigma = 0.01$, $r_b = f_b = 10^{-1.5}$. Initial states of trajectories: (a) $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$; (b) denoted by (*) $(x_D, x_G, x_V) = (0.417, 0.311, 0.272)$; (c) denoted by ($\times$) $(x_D, x_G, x_V) = (0.35, 0.3, 0.35)$.

3.2 Optimization trade-offs: total administered drugs versus time to recovery

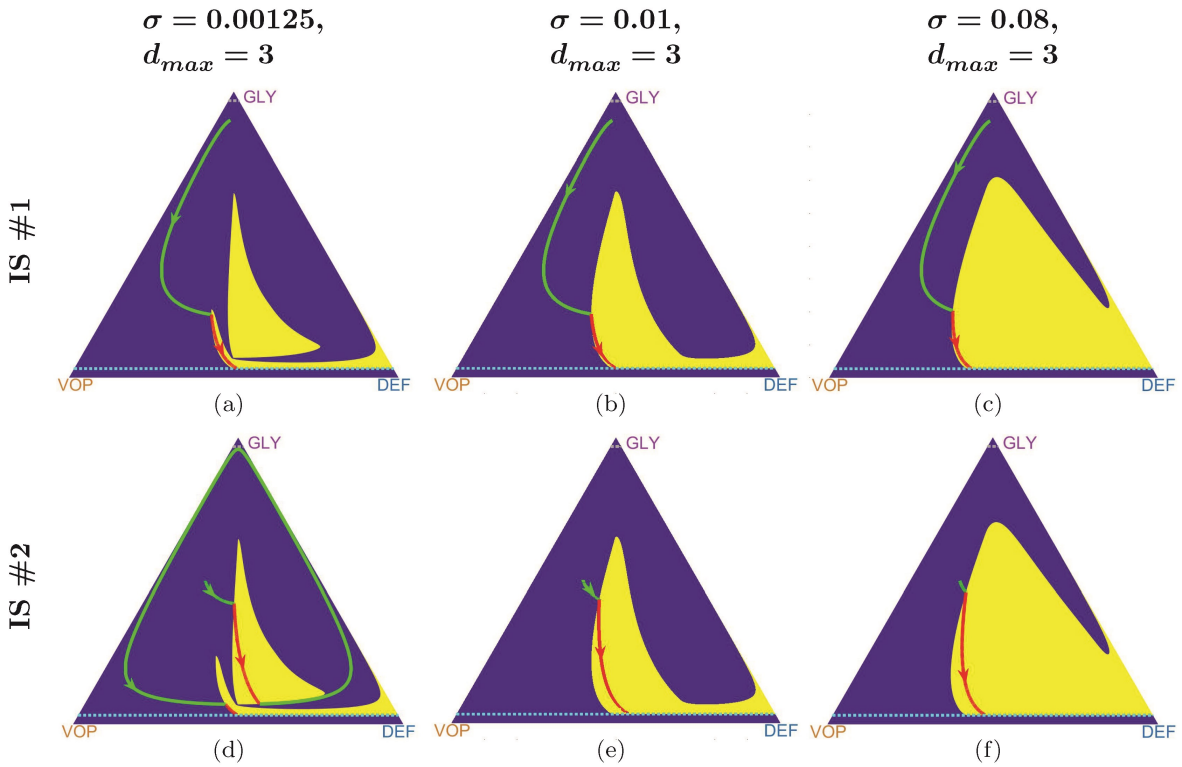
The optimal therapy-on regions are clearly dependent on specific values of all model parameters. Here we explore their dependence on σ and d_{max}. Recall that, for every policy leading to recovery, the *overall cost* of treatment is a sum of the "therapy cost" (i.e., the total amount of drugs administered, $D = \int_0^T d(t)dt$) and the treatment-time cost σT. Since the optimal control is bang-bang, this can be re-written as a weighted sum of the time-till-recovery T, and the total drug therapy time $\tilde{T} \leq T$. That is, for controls based on repeated therapy-off/MTD-level-therapy switches, we can re-write the overall cost as $J = d_{max}\tilde{T} + \sigma T$, with the ratio between the weights (σ/d_{max}) representing the "relative importance" of T and $\tilde{T}$ for the optimization. But the functional role of these weights is quite different: while σ can be chosen to reflect our preferences, the MTD-rate d_{max} is dictated by the medical reality, which will be patient and drug specific. By varying d_{max} while keeping (σ/d_{max}) constant, we can study the role played by the MTD-level in determining optimal policies under a fixed relative preference between the objectives. In Figure 4 we conduct this experiment for two different initial states and the same set of game parameters ($b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$), demonstrating that the value of d_{max} strongly influences the optimal policies and the shape of the "therapy-on" yellow regions.



Parameters		Initial State (IS) #1 : $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$				Initial State #2 : $(x_D, x_G, x_V) = (0.15, 0.5, 0.35)$			
σ	d_{max}	subfigure	total time	total drugs	overall cost	subfigure	total time	total drugs	overall cost
0.0025	0.75	(a)	42.7519	2.0808	2.1877	(e)	33.9242	3.0750	3.1598
0.005	1.5	(b)	11.5998	2.1376	2.1956	(f)	33.1649	3.0254	3.1912
0.01	3	(c)	11.5649	2.1306	2.2463	(g)	2.8230	3.1241	3.1524
0.08	24	(d)	10.9820	2.1520	3.0306	(h)	1.8920	3.1490	3.3004

Figure 4: Varying the MTD level affects all optimal trajectories.
Here we illustrate a fixed σ/d_{max} ratio (with d_{max} increasing from left to right), which is equivalent to preserving the relevant importance (trade-off) between the total therapy time $\tilde{T}$ and the total treatment time T. Nevertheless, the optimal drugs-on regions (in yellow) vary since any changes in d_{max} also affect the dynamics of the system (7).
Parameters: $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$; $r_b = f_b = 10^{-1.5}$. Two initial states and the (σ, d_{max}) values are specified in the table above.

On the other hand, for any fixed/biological d_{max} value, we can vary σ to study how the trade-off between $\tilde{T}$ and T affects the optimization. This experiment is conducted for the same two initial states in Figure 5. As we can see, smaller σ entails larger total time. Intuitively, this happens since it becomes "cheaper" to pause the therapy until we reach a "better" state to administer the drugs. Larger σ leads to a shorter time-to-recovery T, but also an increase in the total amount of administered drugs $D = d_{max}\tilde{T}$ and a larger therapy-on region (shown in yellow) in the state space.



Parameters		Initial State (IS) #1 : $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$				Initial State #2 : $(x_D, x_G, x_V) = (0.15, 0.5, 0.35)$			
σ	d_{max}	subfigure	total time	total drugs	overall cost	subfigure	total time	total drugs	overall cost
0.00125	3	(a)	11.6363	2.1004	2.1447	(d)	34.4945	3.0713	3.1144
0.01	3	(b)	11.5649	2.1306	2.2463	(e)	2.8230	3.1241	3.1524
0.08	3	(c)	10.9897	2.1574	3.0366	(f)	1.9470	3.1648	3.3206

Figure 5: Different trade-offs (time to recovery vs total drugs) yield different optimal trajectories. The MTD-rate d_{max} is fixed while σ increasing from left to right. The ratio (σ/d_{max}) defines the relevant importance of the total amount of drugs (therapy cost) d versus the total treatment time T. An increase in σ results in smaller T and larger d along the optimal trajectories.
Parameters: $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$; $r_b = f_b = 10^{-1.5}$.

3.3 "Incurable" states and periodic trajectories under MTD treatment.

One might think that, despite being sub-optimal, an aggressive MTD-based strategy is at least always fully reliable and the resulting trajectories are guaranteed to reach the recovery zone from every initial configuration, as shown in Figure 2(c). Indeed, if $\frac{b_a}{n+1} \leq d_{max}$, the MDT-based policy $(d(t) \equiv d_{max})$ guarantees that $\dot{p}$ is always negative; see equation (7). But with $\frac{b_a}{n+1} > d_{max}$ the recovery might not be attained with the constant use of MTDs (even if some other treatment policies are successful!).

Consider, for example, the following set of parameters: $b_a = 4$, $b_v = 2$, $c = 1$, $n = 4$; $r_b = f_b = 10^{-1.5}$; $d_{max} = 0.3$, $\sigma = 0.03$ and an initial state $(x_D, x_G, x_V) = (0.02, 0.8, 0.18)$. Under these parameters the MTD-based therapy has a periodic trajectory[‡]; see Figure 6(b). Since the treatment time is infinite, the cost (10) of such a policy is $+\infty$. (In reality this would lead to the emergence of drug resistance and eventual failure, but this

[‡]This is easy to prove by redefining the parameter $b_a := b_a - (n+1)d_{max} > 0$ and reducing the MTD-based case to the periodic behavior of the original uncontrolled system (3) in the parameter regime (4).

biological situation is not modeled in Kaznatcheev et al.[40].)

We can see that neither of two extreme strategies ("no-drugs-at-all" in Figure 6(a) and the MTD-based "drugs-all-the-time" in Figure 6(b)) can bring the trajectory to the recovery zone. However, their adaptive combination can still achieve the objective. We show a trajectory corresponding to the optimal policy in Figure 6(c). With a larger failure zone (e.g., $r_b = f_b = 10^{-1}$), a previously successful MTD-based treatment might even result in death (Figure 6(d)), while the adaptive strategy still leads to recovery (Figure 6(e)).

For a fixed treatment policy, we define its corresponding *"incurable" area* to be a set of states starting from which it is impossible to cross the recovery barrier. For example in Figure 6, the incurable area of the MTD-based policy includes the state $(x_D, x_G, x_V) = (0.02, 0.8, 0.18)$ (when $r_b = f_b = 10^{-1.5}$ or $r_b = f_b = 10^{-1}$). However, this state is *not* in the (dramatically smaller) incurable area of the adaptive/optimal policy; see Figure 7.

Starting from any incurable configuration, one could similarly pose a different control problem of *maximizing* the time until crossing the failure barrier. While we do not address it here, we note that the HJB approach would be quite suitable to find optimal treatment policies for this problem as well.

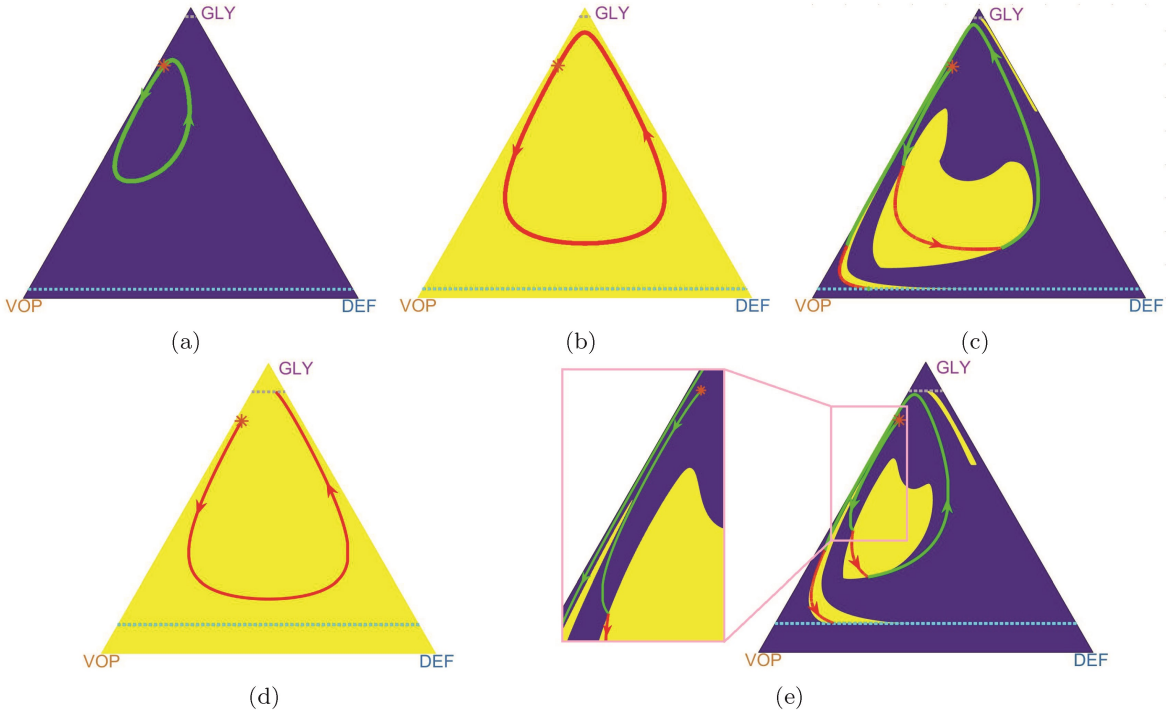


Figure 6: MTD-based policy versus the optimal (adaptive) policy when the MTD rate is low ($d_{max} < \frac{b_a}{n+1}$). **Top row:** trajectories under both (a) "no therapy" policy and (b) the MTD-based policy are cyclic and cannot cross either the recovery nor the failure barrier from the initial state denoted by (*). Nevertheless, the adaptive/optimal switching leads to a full recovery (c).
Bottom row: With a larger failure zone, an MTD-based policy leads to patient's putative death (d) even though it is still possible to cross the recovery barrier under an adaptive/optimal policy (e).
Parameters: $b_a = 4$, $b_v = 2$, $c = 1$, $n = 4$; $d_{max} = 0.3$, $\sigma = 0.03$ and the initial state (*) $(x_D, x_G, x_V) = (0.02, 0.8, 0.18)$. Top row $r_b = f_b = 10^{-1.5}$, bottom row $r_b = f_b = 10^{-1}$.

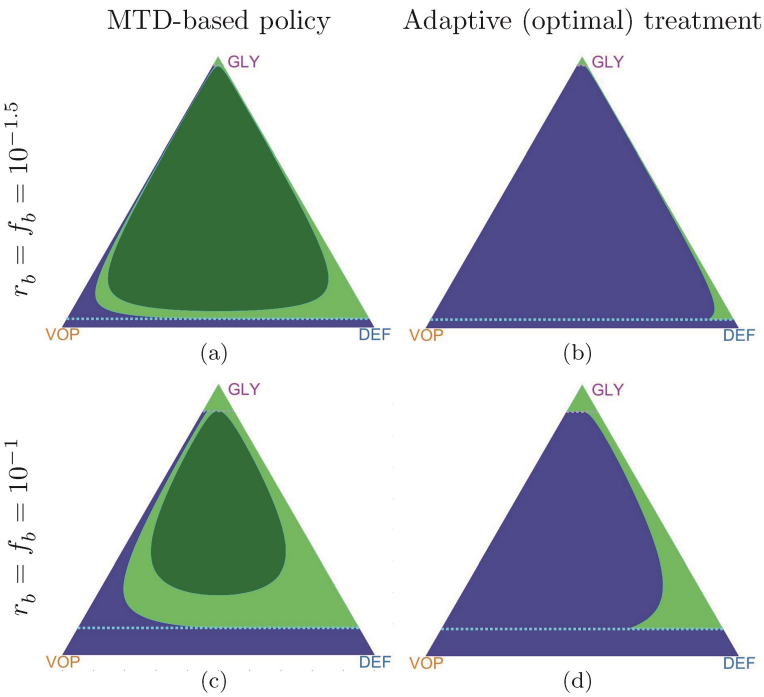


Figure 7: Comparison of the "incurable" area for the MTD-based policy versus the adaptive policy. Using adaptive strategies even with small MTD the patient can recover from most of the states that is impossible using MTD-based policy. *The blue color* indicates initial states from which the corresponding trajectories are able to achieve the recovery zone; *the green color* represents states from which recovery is impossible ("incurable" area): *the light green color* – trajectories eventually get into the failure zone; *the dark green color* – trajectories are cyclic and cannot cross either the recovery or failure zones.
Parameters: $b_a = 4$, $b_v = 2$, $c = 1$, $n = 4$; $d_{max} = 0.3$.

Of course, the "incurable areas" are highly dependent on model parameters. In section 6.4 of Supplementary Materials, we show that they can grow due to an increase in the MTD rate d_{max} or a decrease in vascularization benefits b_v.

4 Discussion

By now it is widely accepted that cancer is an evolutionary process, and that variation and selection drive the emergence of drug resistance. While this new knowledge is driving cancer research forward, it has largely not yet affected clinical practice, with the majority of clinical protocols relying on MTD-based approaches, which invariably fail in the setting of most metastatic disease. This is changing with the advent of adaptive therapy – therapeutic strategies specifically designed with a changing regimen prescribed: one that adapts to an evolving tumor [42].

To optimally design AT protocols, the underling dynamics of the tumor growth and treatment response must be known. While the methods of learning these dynamics are still in their infancy, the last decade has seen a flurry of activity using evolutionary game theory and other evolutionary models for a range of prescribed dynamics. These works have shown qualitative changes in tumor behavior in a range of treatment scenarios, and importantly, demonstrated that the order [43], sequence [40] and timing [17] of therapy can drastically change the outcomes. As we come closer to the reality of evolutionarily designed therapeutic trials in the mainstream, it is important then that we develop methods not just to make our outcomes better, but also to formally optimize them.

However, before using the mathematical tools for optimization, one also needs to choose a specific quantifiable criterion for comparing the outcomes. Once that criterion is selected and the underlying mathematical model is

sufficiently accurate, the best treatment strategy can be found by the techniques of optimal control theory. In this paper, we show how this can be done for one particular heterogeneous cancer model previously described in Kaznatcheev et al.[40]. Given a set of model parameters ($b_v,\ b_a,\ c,\ n$) and treatment/recovery parameters (d_{max}, r_b, f_b), we can find an optimal treatment policy for any initial distribution of cells ($q_0,\ p_0$). We show that the optimal treatment policy can have multiple regimes: always on $d^*(\cdot) \equiv d_{max}$, always off $d^*(\cdot) \equiv 0$, and involving several contiguous treatment periods. For the latter, the challenge is to accurately approximate the on/off "switching curves" in the state space. We show that the definition of optimal treatment policy is heavily dependent on a parameter σ, a "cost" term, describing the relative importance of minimizing the total amount of drugs administered versus the total time to recovery. We further show that, for some parameter regimes, there are "incurable regions" in the state space – the starting configurations that will not lead to a recovery regardless of the chosen treatment strategy – suggesting an alternative therapy (or goal) should be considered. Moreover, for some other starting configurations, the "always on" strategy might not lead to a recovery even if the recovery is actually achievable with some on/off hybrid strategies.

Just as any other model, the approach in Kaznatcheev et al.[40] is based on simplifying assumptions (e.g., only the subpopulation fractions are important, and no novel types can arise), which limit its practical applicability. But our message is broader, and we use this specific model primarily to illustrate the general optimization approach. With adaptive therapy trials now under way showing early promise[14], we propose that such methods will become increasingly integrated into trial design discussions, with increasingly detailed cancer evolution equations or even in data-driven/equation-free framework.

Of the two main approaches of optimal control theory, the Pontryagin Maximum Principle (PMP) has been much more widely used in cancer treatment research up till now. In contrast, our approach here is based on dynamic programming and the numerical methods for Hamilton-Jacobi-Bellman (HJB) equations. The higher computational cost of these methods is balanced by several important practical considerations. First, they yield a policy in feedback-form and are thus more robust to modeling/measurement errors. Second, they always return the globally optimal treatment strategies and avoid some of the pitfalls well-know for the PMP-based methods (e.g., see Figure 3(b)). With the advent of efficient numerical methods, we posit that the HJB equations will be soon playing a larger role in treatment optimization.

There are several obvious directions for extending our approach. First, the ability to optimize outcomes for a *range* of criteria will open new avenues to quantify physician/patient discussions concerning trade-offs and personalized therapy that have previously been only qualitative. In the current paper, computing optimal policies for different values of σ can be viewed as a small step in this direction. But there are many other possible optimization criteria of practical interest. Methods for approximating *all* Pareto-optimal policies are available[44] but are usually more computationally challenging. For probabilistic cancer evolution models, one can also choose between optimizing different characteristics of the same random quantity. (E.g., minimize the average time-to-recovery versus maximizing the probability of recovery in the next year). Finally, one can also use the choice of criterion to promote *robustness* by systematically treating possible measurement/modeling errors as perturbations chosen by some adversarial player. Such "games-against-nature", as described recently by Stankova and colleagues[42], can be similarly treated by solving Hamilton-Jacobi-Isaacs equations.

Another important limitation of the above techniques is our assumed full knowledge of the system state. E.g., we assumed that the exact subpopulation fractions are known at every point in time and can be used to decide whether to administer the drugs. In practice, one can periodically obtain an approximation of these quantities (e.g., based on a repeat biopsy), but most of the time the decisions must be made based on some less invasive measurements (e.g., based on PSA-levels in castrate-resistant prostate cancer model[14] or in cell free circulating DNA as recently shown by Khan and colleagues[45]). A rigorous treatment of such optimization challenges will require the framework of *partially-observable controlled processes*; see e.g., Davis and Varaiya[46].

With cancer evolution playing a larger and larger role in our thinking about therapy, and adaptive/evolutionary therapy coming to the fore, methods for optimizing these approaches will grow in importance. We suggest that the HJB method be used in these scenarios, and that clinicians and modelers begin discussions of further method development along these lines.

Acknowledgement

MG acknowledges the support from The Institute for Data and Decision Analytics at The Chinese University of Hong Kong, Shenzhen during his visit when this research was completed.

JGS would like to thank the NIH Case Comprehensive Cancer Center support grant P30CA043703 and the Calabresi Clinical Oncology Research Program, National Cancer Institute of Award Number K12CA076917.

AV would like to thank the Simons Foundation for its fellowship support and the National Science Foundation (award DMS-1738010) for supporting development of numerical methods for Hamilton Jacobi equations. A part of this work was performed during a sabbatical leave at Princeton/ORFE, and AV is grateful to ORFE Department for its hospitality.

5 Conflict of Interest

The authors declare no conflict of interest.

6 Supplementary Materials

6.1 Deriving a Hamilton-Jacobi-Bellman equation

We start by explaining the logic of dynamic programming that yields the HJB PDE (12) (Box 3, Section 3.1), and the "bang-bang" property of optimal treatment policies.

Recall that the evolving composition of cancer sub-populations can be fully defined by $(q(t), p(t))$; see formulas (2) and (7). The process is tracked until we cross either a recovery or failure barrier; i.e., until the trajectory leaves $\Omega = \Big([0,1]\times[0,1]\Big)\backslash\Delta$, with the terminal set Δ defined in formula (6). For an arbitrary initial state $(q_0, p_0) \in \Omega$, the goal is to choose our treatment policy to minimize the integral of an instantaneous cost $K(d(t)) = d(t) + \sigma$ up to the terminal time $T = T(q_0, p_0, d(\cdot))$. I.e., the total cost of starting at (q_0, p_0) and using a policy $d(\cdot)$ is

$$J(q_0, p_0, d(\cdot)) \;=\; \int_0^T K\left(d(s)\right)\, ds \;+\; g\Big(q\big(T\big), p\big(T\big)\Big),$$

where g is the terminal cost specified on Δ in formula (9). The *value function* $u(q_0, p_0)$ is the result of minimizing J over all available treatment policies, and we say that the policy $d^*(\cdot)$ is optimal if $u(q_0, p_0) = J(q_0, p_0, d^*(\cdot))$.

Bellman's Optimality Principle [47] is the key idea of dynamic programming. It states that, if we move along any optimal trajectory, a remaining (yet to be traversed) part of that trajectory is in itself optimal from our *current* configuration/state. In terms of the above model,

$$u(q_0, p_0) \;=\; \int_0^\tau K\left(d^*(t)\right)\, ds \;+\; u\Big(q(\tau), p(\tau)\Big) \tag{14}$$

should hold for every sufficiently small $\tau > 0$. Assuming that the value function $u(q, p)$ and $d^*(t)$ are smooth, one can use Taylor series and take the limit $\tau \to 0$ to obtain

$$\nabla u(q_0, p_0) \cdot \begin{pmatrix} \dot{q}(q_0, p_0, d_0^*) \\ \dot{p}(q_0, p_0, d_0^*) \end{pmatrix} + d_0^* + \sigma = 0. \tag{15}$$

Here $d_0^* = d^*(0)$ is the optimal *initial* rate of therapy starting from (q_0, p_0) and $(\dot{q}, \dot{p})$ are specified by the right hand side of the ODEs in (7). Since (15) does not involve $d^*(t)$ for any $t > 0$, it is now natural to switch to a *feedback control* perspective based on a state-dependent (rather than explicitly time-dependent) optimal control $d_0^*(q, p)$. Since the latter is a priori unknown, a Hamilton-Jacobi PDE (12) is obtained by minimizing over all available control values $d \in [0, d_{max}]$ and demanding that (15) should hold at every $(q, p) \in \Omega$. Additional boundary conditions $u = g$ are specified on Δ by (13).

The above derivation is merely formal since the value function u is typically non-smooth. Indeed, (12) rarely has classical solutions, and if one considers Lipschitz-continuous weak solutions (by demanding that the PDE should hold wherever ∇u is defined), one immediately loses the uniqueness. Additional test conditions introduced by Crandall and Lions [48] are employed to pick out a *viscosity solution* – the unique weak solution coinciding with the value function of the original control problem [36]. Convergence to this viscosity solution is also a requirement for all numerical methods for HJB equations used in control-theoretic applications.

Using the dynamics (7) specific to our model, the HJB equation (12) can be re-written as follows:

$$\min_{d\in[0,d_{max}]}\left\{\Big[1-u_p p(1-p)\Big]d \;+\; u_q q(1-q)\Big(\frac{b_v}{n+1}\sum_{k=0}^{n}p^k - c\Big) + u_p p(1-p)\Big(\frac{b_a}{n+1} - q(b_v-c)\Big) + \sigma\right\} \;=\; 0. \tag{16}$$

The linear d-dependence of the minimized expression allows us to find the minimizer in closed form:

$$d^* = \begin{cases} d_{max}, & \text{if } \Big(1-u_p p(1-p)\Big) < 0; \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Therefore, an optimal treatment policy takes only extreme values – either 0 or d_{max}. This is usually called the *bang-bang* property.

Using (17) in practice would require knowing u_p at every point (q,p). In principle u_p, can be computed along an optimal trajectory (backwards, from the recovery barrier to our initial state (q_0,p_0) without solving the PDE on the entire Ω. This is in a sense the main idea of Pontryagin Maximum Principle (PMP)[28]. This method will work (and will be much more computationally attractive) as long as u remains smooth along the optimal trajectory. Unfortunately, the PMP has no way of identifying whether a backward-traced trajectory passes through a shockline (where ∇u is undefined). In practice, this would result in obtaining locally (rather than globally) optimal treatment policies; see the example in Figure 3c. We thus focus on solving the full HJB equation, yielding a variational formula for $d^*(q,p)$.

6.2 Numerical methods for the Hamilton-Jacobi-Bellman equation

We obtain an approximate solution to HJB equations on a regular triangulated mesh over the (x_D, x_G, x_V) space; see Figure 8(b). Since at any moment of time $x_V(t) \equiv 1 - x_G(t) - x_D(t)$, it is enough to consider an ODE system for two sub-populations $x_D(t)$ and $x_G(t)$. To simplify the notation, we will write

$$\begin{cases} \dot{y}(t) = f\Big(y(t), d(t)\Big) \\ y(0) = x, \end{cases} \tag{18}$$

where $y(t) = (x_D(t), x_G(t))$, $x = (x_D(0), x_G(0))$ and $f(\cdot)$ denotes the right-hand side of (7) in (x_D, x_G) coordinates using the transformation (2).

Our approximation scheme is based on a first-order accurate semi-Lagrangian discretization[49]. Starting at a meshpoint x and using control d, we assume that the rate of change is constant for a small time τ, yielding a new state

$$\tilde{x}_d = x + \tau f(x,d). \tag{19}$$

Assuming that the running cost is also constant over that small time interval, one can rewrite Bellman's optimality principle as

$$u(x) \;=\; \min_d \left\{\tau K(d) \;+\; u\left(\tilde{x}_d\right)\right\} \;+\; o(\tau). \tag{20}$$

Since $\tilde{x}_d$ is usually not a meshpoint, $u\left(\tilde{x}_d\right)$ is approximated by interpolation using the neighboring meshpoint values. (This is the key idea of all *semi-Lagrangian* techniques.) While there are many ways to choose τ, we select it for each d value individually to guarantee that $\tilde{x}_d$ lies on a mesh line and only two neighboring values are needed for the interpolation; a similar approach has been used in[50,51]. More specifically, suppose that a vector $f(x,d)$ anchored at x lies within a triangle $xx_{1d}x_{2d}$; see Figure 8(a). Then $\tilde{x}_d$ lies on a segment $x_{1d}x_{2d}$ with

$$\tilde{x}_d \;=\; \frac{||x_{2d}-\tilde{x}_d||}{||x_{1d}-x_{2d}||}x_{1d} \;+\; \frac{||x_{1d}-\tilde{x}_d||}{||x_{1d}-x_{2d}||}x_{2d} \quad \text{and} \quad \tau_d \;=\; \frac{||x-\tilde{x}_d||}{||f(x,d)||}.$$

Recalling that only extreme rates d can be optimal due to the bang-bang property, we obtain a coupled system of discretized equations:

$$U(x) \;=\; \min_{d\in\{0,\,d_{max}\}}\left\{\frac{||x-\tilde{x}_d||}{||f(x,d)||}K(d) \;+\; \frac{||x_{2d}-\tilde{x}_d||}{||x_{1d}-x_{2d}||}U(x_{1d}) \;+\; \frac{||x_{1d}-\tilde{x}_d||}{||x_{1d}-x_{2d}||}U(x_{2d})\right\}, \tag{21}$$

which must hold for each meshpoint $x \in \Omega$. The boundary conditions are handled by setting $U = 0$ when $x_G < r_b$ and $U = +\infty$ when $x_G > 1 - f_b$.

In our implementation, the above coupled system of discretized equations is handled by Gauss-Seidel iterations, with an additional speed up through alternating meshpoint orderings (in a "Fast Sweeping" fashion)[52–54]. Another alternative would be to decouple the system dynamically – by selecting larger τ_d adaptively so that "already known" mesh values would be sufficient for updating the still-tentatively-known U values. The latter "Ordered Upwind" approach has been primarily used in problems with geometric dynamics[51,55,56] and offers advantages when optimal trajectories frequently change directions. In the future, it would be interesting to extend it (as well as its two-scale hybrids with sweeping[57]) to therapy optimization problems, particularly for the case of small σ.

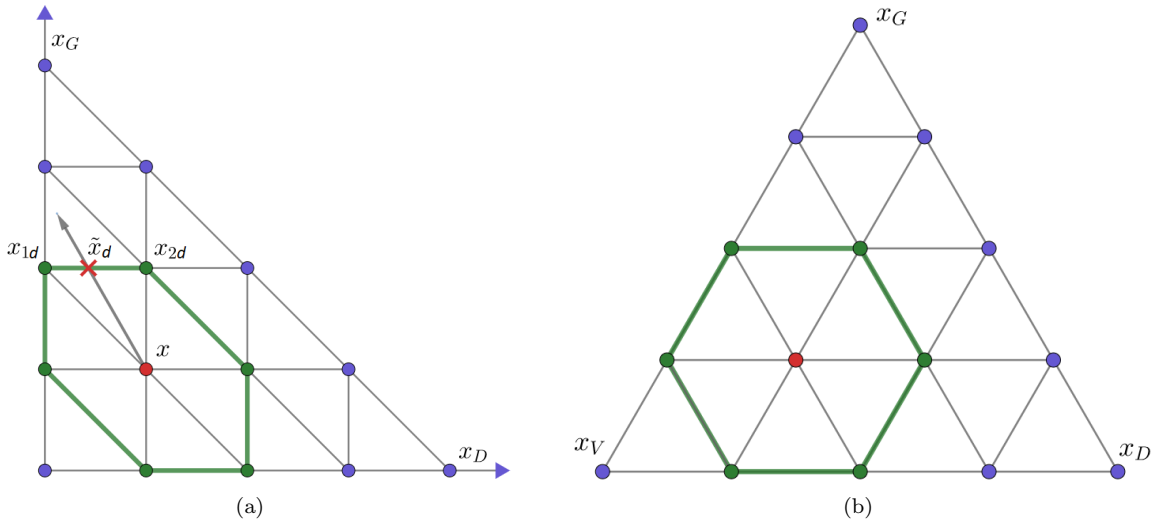


Figure 8: A Semi-Lagrangian scheme on a triangular mesh.
(a): A semi-Lagrangian discretization in (x_D, x_G).
(b): Linear transformation yields a regular triangular mesh.

From implementation purposes, it is easier to conduct computations in a Cartesian coordinate system (Figure 8(a)), which is equivalent to a regular triangular mesh (Figure 8(b)) by a linear transformation.

To ensure the accuracy of the value function in Figure 3(a), we have used $n = 9000$ meshpoints along one side of the GLY-VOP-DEF triangle, yielding $N = 37\ 988\ 686$ meshpoints in Ω with $r_b = f_b = 10^{-1.5}$. The algorithm terminates when the difference between value functions in sequential iterations falls below 10^{-5}, which required a total of 62 iterations (sweeps) for this example.

In the future, we hope to reduce the computational cost by using a higher-order accurate semi-Lagrangian discretization[49] on a coarser mesh, employing "Ordered Upwind" techniques to reduce or eliminate the coupling in the discretized system.

6.3 Fully angiogenic and glycolyctic tumors

In section 3 we have focused on optimizing treatment policies for polyclonal tumors. Under "therapy-off" policy any trajectory of a polyclonal tumour has periodic dynamics and model parameters satisfy (4). Two other types of tumours are also possible under the model[40]: fully angiogenic and fully glycolyctic.

A tumor has a fully angiogenic regime if $\max\left(\frac{b_a}{n+1}, cn\right) < b_v - c$. If the model (3) satisfies this condition all cells tend to switch to VOP type and the trajectory converges to the recovery zone[40]. In some sense, the fully angiogenic regime is less interesting case for our analysis because even without any therapy a patient will recover. However, the optimal control analysis still might be useful when, for example, time-penalty is high (a patient wants to recover as soon as possible) and some amount of drugs can be applied to accelerate the recovery, see Figure 9(a).

A tumor has a fully glycolyctic regime if $\frac{b_a}{n+1} > b_v - c$, all cells tend to be GLY type cells and trajectories converge to the failure zone from any initial state. Even if a treatment policy gives some short-term results, the trajectory will turn towards the failure zone once the therapy is stopped. Nevertheless, crossing the recovery

barrier means full recovery under assumptions of the model[40]. We consider an example of optimal policy for fully glycolyctic tumour in Figure 9(b).

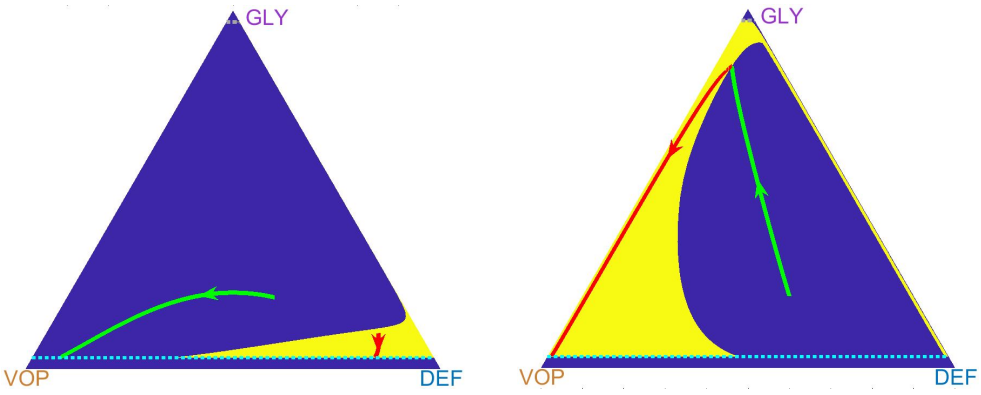
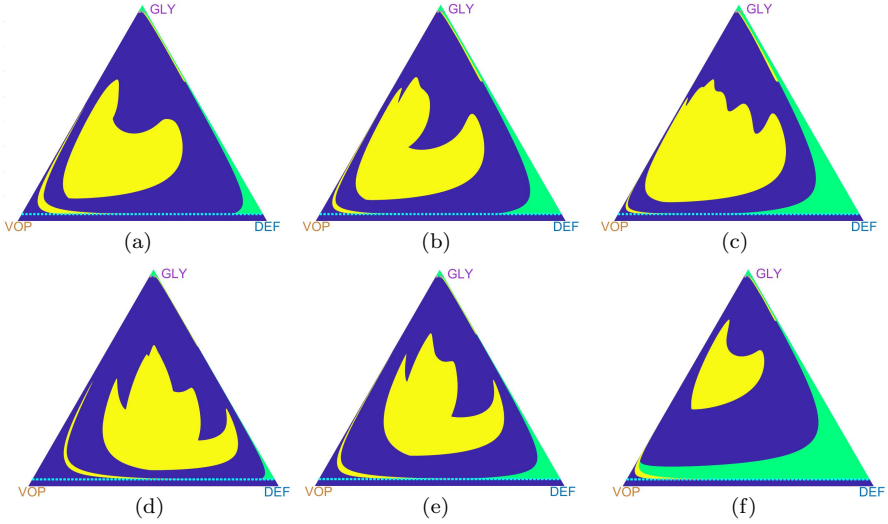


Figure 9: The benefits of optimization in fully angiogenic and glycolyctic cases.
(a): two trajectories under AT strategy in a *fully angiogenic* tumour. Parameters: $b_a = 2$, $b_v = 3$, $c = 1$, $n = 1$; $d_{max} = 3$, $\sigma = 0.3$, $r_b = f_b = 10^{-1.5}$.
(b): a trajectory under AT strategies in a *fully glycolyctic* tumour. Parameters: $b_a = 30$, $b_v = 6$, $c = 1$, $n = 4$; $d_{max} = 3$, $\sigma = 0.01$, $r_b = f_b = 10^{-1.5}$.

6.4 Incurable areas changing under parameter variation

In Figure 10 we examine the changes in the optimal/minimal incurable area due to variations in the MTD rate and model parameters.



Subfigure	Parameters			
	d_{max}	σ	b_a	b_v
(a)	**0.3**	0.03	4	2
(b)	**0.2**	0.03	4	2
(c)	**0.1**	0.03	4	2
(d)	0.2	0.01	2.5	**2**
(e)	0.2	0.01	2.5	**1.8**
(f)	0.2	0.01	2.5	**1.6**

Figure 10: Optimal drugs-on regions (in yellow) and incurable areas (in green) changing under parameter variation. "Incurable" areas can grow due to a decrease in the MTD rate d_{max} (top row) or a decrease in vascularization benefits b_v (bottom row). Common parameters for the figures: $c = 1$, $n = 4$, $r_b = f_b = 10^{-1.5}$.

References

[1] Andriy Marusyk and Kornelia Polyak. Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1805(1):105–117, 2010.

[2] D Hanahan and R A Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

[3] Jacob Scott and Andriy Marusyk. Somatic clonal evolution: A selection-centric perspective. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1867(2):139–150, 2017. doi: 10.1016/J.BBCAN.2017.01.006.

[4] Andriy Marusyk, Doris P. Tabassum, Philipp M. Altrock, Vanessa Almendro, Franziska Michor, and Kornelia Polyak. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, 514(7520): 54–58, 2014. doi: 10.1038/nature13556.

[5] Douglas Hanahan, Gabriele Bergers, and Emily Bergsland. Less is more, regularly: metronomic dosing of cytotoxic drugs can target tumor angiogenesis in mice. *Journal of Clinical Investigation*, 105(8):1045–1047, 2000.

[6] K. Lien, S. Georgsdottir, L. Sivanathan, K. Chan, and U. Emmenegger. Low-dose metronomic chemotherapy: A systematic literature analysis. *European Journal of Cancer*, 49(16):3387–3395, 2013. doi: 10.1016/J.EJCA.2013.06.038.

[7] Eddy Pasquier, Maria Kavallaris, and Nicolas André. Metronomic chemotherapy: new rationale for new directions. *Nature Reviews Clinical Oncology*, 7(8):455–465, 2010. doi: 10.1038/nrclinonc.2010.82.

[8] Santosh Kesari, David Schiff, Lisa Doherty, Debra C Gigas, Tracy T Batchelor, and Alona Muzikansky et al. Phase II study of metronomic chemotherapy for recurrent malignant gliomas in adults. *Neuro-oncology*, 9(3):354–363, 2007.

[9] Simone Steinbild, Jann Arends, Michael Medinger, Brigitte Häring, Annette Frost, and Joachim Drevs et al. Metronomic Antiangiogenic Therapy with Capecitabine and Celecoxib in Advanced Tumor Patients - Results of a Phase II Study. *Oncology Research and Treatment*, 30(12):629–635, 2007.

[10] Andreza A. Senerchia, Carla Renata Macedo, Sima Ferman, Marcelo Scopinaro, Walter Cacciavillano, and Erica Boldrini et al. Results of a randomized, prospective clinical trial evaluating metronomic chemotherapy in nonmetastatic patients with high-grade, operable osteosarcomas of the extremities: A report from the Latin American Group of Osteosarcoma Treatment. *Cancer*, 123(6):1003–1010, 2017.

[11] Chong-Sheng Chen, Joshua C. Doloff, and David J. Waxman. Intermittent Metronomic Drug Schedule Is Essential for Activating Antitumor Innate Immunity and Tumor Xenograft Regression. *Neoplasia*, 16(1): 84–96, 2014. doi: 10.1593/NEO.131910.

[12] R. A. Gatenby, A. S. Silva, R. J. Gillies, and B. R. Frieden. Adaptive Therapy. *Cancer Research*, 69(11): 4894–4903, 2009.

[13] Pedro M Enriquez-Navas, Yoonseok Kam, Tuhin Das, Sabrina Hassan, Ariosto Silva, and Parastou Foroutan et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Science translational medicine*, 8(327):327ra24, 2016.

[14] Jingsong Zhang, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications*, 8(1):1816, 2017. doi: 10.1038/s41467-017-01968-5.

[15] John Maynard Smith. *Evolution and the theory of games*. Cambridge University Press, Cambridge, 1982. ISBN 9780511806292. doi: 10.1017/CBO9780511806292.

[16] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998. ISBN 9780521625708.

[17] D Basanta, J G Scott, M N Fishman, G Ayala, S W Hayward, and A R A Anderson. Investigating prostate cancer tumour-stroma interactions: clinical and biological insights from an evolutionary game. *British Journal of Cancer*, 106(1):174–181, 2012. doi: 10.1038/bjc.2011.517.

[18] Li You, Joel S. Brown, Frank Thuijsman, Jessica J. Cunningham, Robert A. Gatenby, and Jingsong Zhang et al. Spatial vs. non-spatial eco-evolutionary dynamics in a tumor growth model. *Journal of Theoretical Biology*, 435:78–97, 2017. doi: 10.1016/j.jtbi.2017.08.022.

[19] Jessica J. Cunningham, Joel S. Brown, Robert A. Gatenby, and Kateina Staková. Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. *Journal of Theoretical Biology*, 459: 67–78, 2018.

[20] David Basanta, Jacob G Scott, Russ Rockne, Kristin R Swanson, and Alexander R A Anderson. The role of IDH1 mutated tumour cells in secondary glioblastomas: an evolutionary game theoretical view. *Physical Biology*, 8(1):015016, 2011. doi: 10.1088/1478-3975/8/1/015016.

[21] D Dingli, C Offord, R Myers, K-W Peng, T W Carr, and K Josic et al. Dynamics of multiple myeloma tumor therapy with a recombinant measles virus. *Cancer gene therapy*, 16(12):873–82, 2009. doi: 10.1038/cgt.2009.40.

[22] A. Wu, D. Liao, T. D. Tlsty, J. C. Sturm, and R. H. Austin. Game theory in the death galaxy: interaction of cancer and stromal cells in tumour microenvironment. *Interface Focus*, 4(4):20140028, 2014. doi: 10.1098/rsfs.2014.0028.

[23] N. L. Komarova and D. Wodarz. Drug resistance in cancer: Principles of emergence and prevention. *Proceedings of the National Academy of Sciences*, 102(27):9714–9719, 2005. doi: 10.1073/pnas.0501870102.

[24] Paul A Orlando, Robert A Gatenby, and Joel S Brown. Cancer treatment as a game: integrating evolutionary game theory into the optimal control of chemotherapy. *Physical biology*, 9(6):065007, 2012.

[25] Jeffrey West, Zaki Hasnain, Jeremy Mason, and Paul K Newton. The prisoner's dilemma as a cancer model. *Convergent Science Physical Oncology*, 2(3):035002, 2016. doi: 10.1088/2057-1739/2/3/035002.

[26] Heinz Schättler and Urszula Ledzewicz. *Optimal Control for Mathematical Models of Cancer Therapies*, volume 42 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY, 2015. ISBN 978-1-4939-2971-9. doi: 10.1007/978-1-4939-2972-6.

[27] George W. Swan and Thomas L. Vincent. Optimal control analysis in the chemotherapy of IgG multiple myeloma. *Bulletin of Mathematical Biology*, 39(3):317–337, 1977.

[28] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko. *The mathematical theory of optimal processes*. John Wiley & Sons, Inc., New York, 1962.

[29] Heinz Schattler and Urszula Ledzewicz. Drug resistance in cancer chemotherapy as an optimal control problem. *Discrete and Continuous Dynamical Systems - Series B*, 6(1):129–150, 2006. doi: 10.3934/dcdsb.2006.6.129.

[30] Shuo Wang and Heinz Schättler. Optimal control of a mathematical model for cancer chemotherapy under tumor heterogeneity. *Mathematical Biosciences and Engineering*, 13(6):1223–1240, 2016. doi: 10.3934/mbe.2016040.

[31] Cécile Carrère. Optimization of an in vitro chemotherapy to avoid resistant tumours. *Journal of Theoretical Biology*, 413:24–33, 2017. doi: 10.1016/J.JTBI.2016.11.009.

[32] Alberto D'Onofrio, Urszula Ledzewicz, Helmut Maurer, and Heinz Schättler. On optimal delivery of combination therapy for tumors. *Mathematical Biosciences*, 222(1):13–26, 2009. doi: 10.1016/J.MBS.2009.08.004.

[33] Yongmei Su, Chen Jia, and Ying Chen. Optimal Control Model of Tumor Treatment with Oncolytic Virus and MEK Inhibitor. *BioMed research international*, 2016:5621313, 2016.

[34] Urszula Ledzewicz, Mohammad Naghnaeian, and Heinz Schättler. Optimal response to chemotherapy for a mathematical model of tumor-immune dynamics. *Journal of Mathematical Biology*, 64(3):557–577, 2012.

[35] João M. Lemos, Daniela V. Caiado, Rui Coelho, and Susana Vinga. Optimal and receding horizon control of tumor growth in myeloma bone disease. *Biomedical Signal Processing and Control*, 24:128–134, 2016. doi: 10.1016/J.BSPC.2015.10.004.

[36] Martino Bardi and Italo Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhäuser, Boston, MA, 1997.

[37] Daniel Liberzon. *Calculus of variations and optimal control theory : a concise introduction*. Princeton University Press, Princeton, Oxford, 2012. ISBN 0691151873.

[38] A. Nowakowski and A. Popa. A Dynamic Programming Approach for Approximate Optimal Control for Cancer Therapy. *Journal of Optimization Theory and Applications*, 156(2):365–379, 2013. doi: 10.1007/s10957-012-0137-z.

[39] Alexander Lorz, Tommaso Lorenzi, Michael E. Hochberg, Jean Clairambault, and Benoît Perthame. Populational adaptive evolution, chemotherapeutic resistance and multiple anti-cancer therapies. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(2):377–399, 2013.

[40] Artem Kaznatcheev, Robert Vander Velde, Jacob G Scott, and David Basanta. Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *British Journal of Cancer*, 116(6):785–792, 2017.

[41] Artem Kaznatcheev, Jeffrey Peacock, David Basanta, Andriy Marusyk, and Jacob G. Scott. Fibroblasts and alectinib switch the evolutionary games that non-small cell lung cancer plays. *bioRxiv*, 2017. doi: 10.1101/179259. URL `https://www.biorxiv.org/content/early/2017/09/20/179259`.

[42] Katerina Stanková, Joel S. Brown, William S. Dalton, and Robert A. Gatenby. Optimizing Cancer Treatment Using Game Theory. *JAMA Oncology*, 2018. doi: 10.1001/jamaoncol.2018.3395. URL `http://oncology.jamanetwork.com/article.aspx?doi=10.1001/jamaoncol.2018.3395`.

[43] Daniel Nichol, Peter Jeavons, Alexander G. Fletcher, Robert A. Bonomo, Philip K. Maini, and Jerome L. Paul et al. Steering Evolution with Sequential Therapy to Prevent the Emergence of Bacterial Antibiotic Resistance. *PLOS Computational Biology*, 11(9):e1004493, 2015. doi: 10.1371/journal.pcbi.1004493.

[44] Ajeet Kumar and Alexander Vladimirsky. An efficient method for multiobjective optimal control and optimal control subject to integral constraints. *Journal of Computational Mathematics*, 28:517–551, 2010. doi: 10.2307/43693920.

[45] Khurum H Khan, David Cunningham, Benjamin Werner, Georgios Vlachogiannis, Inmaculada Spiteri, and Timon Heide et al. Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer discovery*, 8(10):1270–1285, 2018. doi: 10.1158/2159-8290.CD-17-0891.

[46] M. H. A. Davis and P. Varaiya. Dynamic Programming Conditions for Partially Observable Stochastic Systems. *SIAM Journal on Control*, 11(2):226–261, 1973.

[47] Richard Bellman. *Dynamic programming*. Princeton University Press, 1957. ISBN 0486428095.

[48] Michael G. Crandall and Pierre-Louis Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–1, 1983. doi: 10.1090/S0002-9947-1983-0690039-8.

[49] Maurizio Falcone and Roberto Ferretti. *Semi-Lagrangian Approximation Schemes for Linear and Hamilton-Jacobi Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013. ISBN 978-1-61197-304-4. doi: 10.1137/1.9781611973051.

[50] R. Gonzalez and E. Rofman. On Deterministic Control Problems: An Approximation Procedure for the Optimal Cost I. The Stationary Problem. *SIAM Journal on Control and Optimization*, 23(2):242–266, 1985. doi: 10.1137/0323018.

[51] James A Sethian and Alexander Vladimirsky. Ordered upwind methods for static Hamilton–Jacobi equations: Theory and algorithms. *SIAM Journal on Numerical Analysis*, 41(1):325–363, 2003.

[52] Michelle Boué and Paul Dupuis. Markov Chain Approximations for Deterministic Control Problems with Affine Dynamics and Quadratic Cost in the Control. *SIAM Journal on Numerical Analysis*, 36(3):667–695, 1999. doi: 10.1137/S0036142997323521.

[53] Hongkai Zhao. A fast sweeping method for Eikonal equations. *Mathematics of Computation*, 74(250):603–628, 2004.

[54] Jianliang Qian, YongTao Zhang, and HongKai Zhao. Fast Sweeping Methods for Eikonal Equations on Triangular Meshes. *SIAM Journal on Numerical Analysis*, 45(1):83–107, 2007.

[55] Ken Alton and Ian M Mitchell. An ordered upwind method with precomputed stencil and monotone node acceptance for solving static convex Hamilton-Jacobi equations. *Journal of Scientific Computing*, 51(2): 313–348, 2012.

[56] Jean-Marie Mirebeau. Efficient fast marching with Finsler metrics. *Numerische mathematik*, 126(3):515–557, 2014.

[57] Adam Chacon and Alexander Vladimirsky. Fast Two-scale Methods for Eikonal Equations. *SIAM Journal on Scientific Computing*, 34(2):A547–A578, 2012.