Supplementary Materials

Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory

Mark Gluzman, Jacob G. Scott, and Alexander Vladimirsky

Abstract

This document provides additional details for the study 'Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory' in Proceeding B.

Note the numbering of equations and figures continues from the main text. We continue to use reference numbers of equations and figures previously introduced in the paper for this document. Bibliographic references used in Supplementary Materials are listed and numbered separately.

1S Deriving the tumor model from Kaznatcheev et. al.

In this section we summarize the main steps of deriving system of equations (2.2) from Kaznatcheev et al. [9]. The tumor model in that paper involves glycolytic (GLY) cancer cells, which are anaerobic and produce acid; and normal cells, which are aerobic and benefit from the increased oxygen from vascularization. The acidity benefits all tumor cells, regardless of whether they have aerobic or anaerobic metabolism. Normal cells are divided into two subpopulations: VOP cells that (over)produce Vascular Endothelial Growth Factor (VEGF) improving vascularization at some cost and DEF cells that do not (over)produce VEGF. GLY cells do not produce VEGF and thus do not increase vascularization.

We assume that every cell interacts with n nearby cells and derives benefits based on its type and type of cells around it. If among those nearby cells and cell itself k cells are GLY, the benefit (per unit time) due to acidity is $A_n(k) = \frac{b_a k}{n+1}$. Aerobic cells are additionally receiving a benefit per unit time $V_{n-n_{\rm G}}(l) = \frac{b_v l}{n-n_{\rm G}+1}$ due to vascularization, where l is the number of VOP cells that (over)produce VEGF at a cost c. An individual cell, depending on its type, can have the following payoff, see details in Section A of Appendix in Kaznatcheev et al. [9]:

- A GLY cell interacting with $n_{\rm G}$ other GLY cells, has payoff $A_n(n_{\rm G}+1)$.
- A VOP cell interacting with $n_{\rm G}$ GLY cells and $n_{\rm V}$ other VOP cells has a payoff $A_n(n_{\rm G}) + V_{n-n_{\rm G}}(n_{\rm V}+1) - c$.
- A DEF cell interacting with $n_{\rm G}$ GLY cells and $n_{\rm V}$ VOP cells has a payoff $A_n(n_{\rm G}) + V_{n-n_{\rm G}}(n_{\rm V})$.

By averaging over all possible interaction group compositions, one can get the expected fitness of a population of glycolytic (GLY) cells

$$w_{\text{G}} = \langle A_n(n_{\text{G}} + 1)\rangle_{n_{\text{G}} \sim B_n(x_{\text{G}})},$$

where $B_n(x_{\text{G}})$ is the binomial distribution with n samples and x_{G} is the probability of success. By $\langle f(\xi)\rangle_{\xi\sim\Omega}$ we denote the expected value of $f(\xi)$ over a random variable ξ with distribution Ω.

Other two types of cells have the following expected fitness:

$$\begin{aligned} w_{\text{V}} &= \langle A_n(n_{\text{G}})\rangle_{n_{\text{G}} \sim B_n(x_{\text{G}})} + \langle V_{n-n_{\text{G}}}(n_{\text{V}} + 1)\rangle_{(n_{\text{G}},n_{\text{V}})\sim M_n(x_{\text{G}},x_{\text{V}})} - c, \\ w_{\text{D}} &= \langle A_n(n_{\text{G}})\rangle_{n_{\text{G}} \sim B_n(x_{\text{G}})} + \langle V_{n-n_{\text{G}}}(n_{\text{V}})\rangle_{(n_{\text{G}},n_{\text{V}})\sim M_n(x_{\text{G}},x_{\text{V}})}, \end{aligned}$$

where $M_n(x_{\text{G}}, x_{\text{V}})$ is the multinomial distribution with n samples and x_{G} is the probability of the first outcome, x_{V} of the second.

Let $m_{\text{G}}, m_{\text{V}}, m_{\text{D}}$ be a total number cells of type GLY, VOP, DRF in the tumor correspondingly. Assume that each sub-population grows exponentially with growth rate equal to their fitness:

$$\begin{cases} \dot{m}_{\text{G}} = w_{\text{G}} m_{\text{G}}, \\ \dot{m}_{\text{V}} = w_{\text{V}} m_{\text{V}}, \\ \dot{m}_{\text{D}} = w_{\text{D}} m_{\text{D}}. \end{cases}$$

The GLY-targeting therapy attacks acid-producing GLY cells killing them with time-dependent intensity $d : \mathbb{R}_+ \to [0, d_{max}]$, see Box 1 in the main text. When the therapy is applied the sub-population dynamics follows:

$$\begin{cases} \dot{m}_{\text{G}}(t) = w_{\text{G}}(t) m_{\text{G}}(t) - d(t) m_{\text{G}}(t), \\ \dot{m}_{\text{V}}(t) = w_{\text{V}}(t) m_{\text{V}}(t), \\ \dot{m}_{\text{D}}(t) = w_{\text{D}}(t) m_{\text{D}}(t). \end{cases}$$

Let $m = m_{\text{G}} + m_{\text{V}} + m_{\text{D}}$ be the total number of cells in the tumor. Now we look at the dynamics of sub-population fractions $x_{\text{G}} = m_{\text{G}}/m$, $x_{\text{V}} = m_{\text{V}}/m$, $x_{\text{D}} = m_{\text{D}}/m$. For GLY cells:

$$
\begin{aligned}
\dot{x}_{\mathrm{G}} &= \frac{\dot{m}_{\mathrm{G}}}{m} - \frac{m_{\mathrm{G}}\dot{m}}{m^2} \\
&= \frac{(w_{\mathrm{G}} - d)m_{\mathrm{G}}}{m} - \frac{m_{\mathrm{G}}}{m}\frac{(w_{\mathrm{G}} - d)m_{\mathrm{G}} + w_{\mathrm{V}}m_{\mathrm{V}} + w_{\mathrm{D}}m_{\mathrm{D}}}{m} \\
&= \frac{m_{\mathrm{G}}}{m}\left(w_{\mathrm{G}} - d - (w_{\mathrm{G}} - d)\frac{m_{\mathrm{G}}}{m} - w_{\mathrm{V}}\frac{m_{\mathrm{V}}}{m} - w_{\mathrm{D}}\frac{m_{\mathrm{D}}}{m}\right) \\
&= x_{\mathrm{G}}(w_{\mathrm{G}} - d - (w_{\mathrm{G}} - d)x_{\mathrm{G}} - w_{\mathrm{V}}x_{\mathrm{V}} - w_{\mathrm{D}}x_{\mathrm{D}}) \\
&= x_{\mathrm{G}}\left(w_{\mathrm{G}} - (w_{\mathrm{G}}x_{\mathrm{G}} + w_{\mathrm{V}}x_{\mathrm{V}} + w_{\mathrm{D}}x_{\mathrm{D}})\right) - x_{\mathrm{G}}(1 - x_{\mathrm{G}})d \\
&= x_{\mathrm{G}}\left(w_{\mathrm{G}} - \langle w\rangle\right) - x_{\mathrm{G}}(1 - x_{\mathrm{G}})d,
\end{aligned}
$$

where $\langle w\rangle = x_{\mathrm{G}}w_{\mathrm{G}} + x_{\mathrm{V}}w_{\mathrm{V}} + x_{\mathrm{D}}w_{\mathrm{D}}$ is an average fitness.

Using a similar argument to derive $\dot{x}_{\mathrm{V}}$ and $\dot{x}_{\mathrm{D}}$, we can now write down the replicator equation for *subpopulation fractions* based on the GLY-VOP-DEF interactions subject to the GLY-targeting therapy:

$$
\begin{cases}
\dot{x}_{\mathrm{G}} = x_{\mathrm{G}}(w_{\mathrm{G}} - \langle w\rangle) - x_{\mathrm{G}}(1 - x_{\mathrm{G}})d, \\
\dot{x}_{\mathrm{V}} = x_{\mathrm{V}}(w_{\mathrm{V}} - \langle w\rangle), \\
\dot{x}_{\mathrm{D}} = x_{\mathrm{D}}(w_{\mathrm{D}} - \langle w\rangle).
\end{cases}
\tag{1S.1}
$$

One can transform the system of replicator equations (1S.1) into the system with two equations using the fact that $x_{\mathrm{G}} + x_{\mathrm{V}} + x_{\mathrm{D}} = 1$. Define the proportion of GLY cells as $p = x_{\mathrm{G}}$ and the proportion of VEGF producers among the aerobic cells as $q = \frac{x_{\mathrm{V}}}{x_{\mathrm{V}} + x_{\mathrm{D}}}$.

Then

$$
\begin{aligned}
\dot{p} &= p(w_{\mathrm{G}} - \langle w\rangle) - p(1 - p)d \\
&= p(w_{\mathrm{G}} - x_{\mathrm{G}}w_{\mathrm{G}} - x_{\mathrm{V}}w_{\mathrm{V}} - x_{\mathrm{D}}w_{\mathrm{D}}) - p(1 - p)d \\
&= p(w_{\mathrm{G}} - pw_{\mathrm{G}} - x_{\mathrm{V}}w_{\mathrm{V}} - x_{\mathrm{D}}w_{\mathrm{D}}) - p(1 - p)d \\
&= p(w_{\mathrm{G}} - pw_{\mathrm{G}} - (1 - p)\langle w\rangle_{\mathrm{V,D}}) - p(1 - p)d \\
&= p(1 - p)(w_{\mathrm{G}} - \langle w\rangle_{\mathrm{V,D}} - d),
\end{aligned}
$$

where $\langle w\rangle_{\mathrm{V,D}} = qw_{\mathrm{V}} + (1 - q)w_{\mathrm{D}} = w_{\mathrm{D}} + q(w_{\mathrm{V}} - w_{\mathrm{D}})$.

Similarly,

$$\begin{aligned}
\dot{q} &= \frac{\dot{x}_{\mathrm{V}} x_{\mathrm{D}} - \dot{x}_{\mathrm{D}} x_{\mathrm{V}}}{(x_{\mathrm{V}} + x_{\mathrm{D}})^2} \\
&= \frac{1}{(x_{\mathrm{V}} + x_{\mathrm{D}})^2}(x_{\mathrm{V}} x_{\mathrm{D}}(w_{\mathrm{V}} - \langle w \rangle) - x_{\mathrm{V}} x_{\mathrm{D}}(w_{\mathrm{D}} - \langle w \rangle)) \\
&= \frac{x_{\mathrm{V}} x_{\mathrm{D}}}{(x_{\mathrm{V}} + x_{\mathrm{D}})^2}(w_{\mathrm{V}} - w_{\mathrm{D}}) \\
&= q(1-q)(w_{\mathrm{V}} - w_{\mathrm{D}}).
\end{aligned}$$

The reduced system of replicator equations is

$$\begin{cases}
\dot{q} = q(1-q)(w_{\mathrm{V}} - w_{\mathrm{D}}), \\
\dot{p} = p(1-p)(w_{\mathrm{G}} - \langle w \rangle_{\mathrm{V,D}} - d).
\end{cases} \tag{1S.2}$$

Substituting expressions of expected fitness $(w_{\mathrm{G}}, w_{\mathrm{V}}, w_{\mathrm{D}})$ into system (1S.2), one can obtain a system (2.2) used in the main text. Below we reproduce the derivation from Section D of Supplementary Material of Kaznatcheev et al. [9].

First we simplify the gain function for q:

$$\begin{aligned}
w_{\mathrm{V}} - w_{\mathrm{D}} &= \sum_{n_{\mathrm{G}}+n_{\mathrm{V}}+n_{\mathrm{D}}=n} \binom{n}{n_{\mathrm{G}}, n_{\mathrm{V}}, n_{\mathrm{D}}} x_{\mathrm{G}}^{n_{\mathrm{G}}} x_{\mathrm{V}}^{n_{\mathrm{V}}} x_{\mathrm{D}}^{n_{\mathrm{D}}} \left(\frac{b_v}{n - n_{\mathrm{G}} + 1} - c \right) \\
&= \sum_{l=0}^{n} \binom{n}{l} p^{n-l}(1-p)^l \frac{b_v}{l+1} - c && l = n_v - n_{\mathrm{D}} \\
&= \sum_{l=0}^{n} \binom{n+1}{l+1} p^{n-l}(1-p)^l \frac{b_v}{n+1} - c && \binom{n}{l} = \frac{l+1}{n+1}\binom{n+1}{l+1} \\
&= \frac{b_v}{(1-p)(n+1)} \left(\left[\sum_{l'=0}^{n'} \binom{n'}{l'} p^{n'-l'}(1-p)^{l'} \right] - \binom{n'}{0} p^{n'} \right) - c && n' = n+1;\ l' = l+1 \\
&= \frac{b_v}{(1-p)(n+1)}(1 - p^{n+1}) - c \\
&= \frac{b_v}{n+1} \sum_{k=0}^{n} p^k - c && \text{polynomial division.}
\end{aligned} \tag{1S.3}$$

The gain function for p is equal to $w_{\mathrm{G}} - \langle w \rangle_{\mathrm{V,D}} = w_{\mathrm{G}} - w_{\mathrm{D}} - q(w_{\mathrm{V}} - w_{\mathrm{D}})$ and one needs to compute the expected

payoff difference between GLY and DEF cells:

$$\begin{aligned} w_{\mathrm{G}} - w_{\mathrm{D}} &= \sum_{n_{\mathrm{G}}+n_{\mathrm{V}}+n_{\mathrm{D}}=n} \binom{n}{n_{\mathrm{G}}, n_{\mathrm{V}}, n_{\mathrm{D}}} x_{\mathrm{G}}^{n_{\mathrm{G}}} x_{\mathrm{V}}^{n_{\mathrm{V}}} x_{\mathrm{D}}^{n_{\mathrm{D}}} \left(\frac{b_a}{n+1} - \frac{b_v n_{\mathrm{V}}}{n - n_{\mathrm{G}} + 1} \right) \\ &= \frac{b_a}{n+1} - \left(\sum_{l=0}^{n} \binom{n}{l} p^{n-l} (1-p)^l \left(\sum_{k=0}^{l} \binom{l}{k} q^k (1-q)^{l-k} \frac{b_v k}{l+1} \right) \right) \\ &= \frac{b_a}{n+1} - \sum_{l=0}^{n} \binom{n}{l} p^{n-l} (1-p)^l b_v q (1 - \frac{1}{l+1}) \\ &= \frac{b_a}{n+1} - b_v q + \frac{b_v q}{(1-p)(n+1)} (1 - p^{n+1}). \end{aligned}$$

Simplifying the expression for $w_{\mathrm{G}} - \langle w \rangle_{\mathrm{V,D}}$,

$$\begin{aligned} w_{\mathrm{G}} - \langle w \rangle_{\mathrm{V,D}} &= w_{\mathrm{G}} - w_{\mathrm{D}} - q(w_{\mathrm{V}} - w_{\mathrm{D}}) \\ &= \frac{b_a}{n+1} - b_v q + \frac{b_v q}{(1-p)(n+1)} (1 - p^{n+1}) - q \left(\frac{b_v}{(1-p)(n+1)} (1 - p^{n+1}) - c \right) \\ &= \frac{b_a}{n+1} - q(b_v - c). \end{aligned} \tag{1S.4}$$

Substituting expressions (1S.3), (1S.4) into (1S.2), we finally obtain the system (2.2) from Box 1 in the main text.

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

2S Deriving a Hamilton-Jacobi-Bellman equation

We start by explaining the logic of dynamic programming that yields the HJB PDE (3.4) (Box 2, Section 3.1 in the main text), 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 (2.1) and (2.2). 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 (2.4). 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 (3.2). 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 [3] 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,

$$u(q_0, p_0) \;=\; \int_0^{\tau} K\left(d^*(t)\right)\, ds \;+\; u\Big(q(\tau), p(\tau)\Big) \tag{2S.1}$$

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

$$\nabla u(q_0, p_0) \cdot \begin{pmatrix} \dot{q}(q_0, p_0, d_0^*) \\ \dot{p}(q_0, p_0, d_0^*) \end{pmatrix} + d_0^* + \sigma = 0. \tag{2S.2}$$

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 (2.2). Since (2S.2) 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 (3.4) is obtained by minimizing over all available control values $d \in [0, d_{max}]$ and demanding that (2S.2) should hold at every $(q, p) \in \Omega$. Additional boundary conditions $u = g$ are specified on Δ by (3.5).

The above derivation is merely formal since the value function u is typically non-smooth. Indeed, (3.4) 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 [6] are employed to pick out a *viscosity solution* – the unique weak solution coinciding with the value function of the original control problem [2]. Convergence to this viscosity solution is also a requirement for all numerical methods for HJB equations used in control-theoretic applications.

Using the dynamics (2.2) specific to our model, the HJB equation (3.4) can be re-written as follows:

$$\min_{d \in [0, d_{max}]} \left[\Big(1 - u_p p(1-p)\Big) d \;+\; u_q q(1-q)\Big(\frac{b_v}{n+1} \sum_{k=0}^{n} p^k - c\Big) + u_p p(1-p)\Big(\frac{b_a}{n+1} - q(b_v - c)\Big) + \sigma \right] \;=\; 0. \tag{2S.3}$$

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

$$d^* = \begin{cases} d_{max}, & \text{if } \Big(1 - u_p p(1-p)\Big) < 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2S.4}$$

Therefore, an optimal treatment policy takes only extreme values – either 0 or d_{max}. This is usually called the *bang-bang* property. Using (2S.4) 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) [11]. 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 3(c). We thus focus on solving the full HJB equation, yielding a variational formula for $d^*(q, p)$.

3S Numerical methods for the Hamilton-Jacobi-Bellman equation

We obtain an approximate solution to HJB equations on a regular triangulated mesh over the $(x_{\text{D}}, x_{\text{G}}, x_{\text{V}})$ space; see Figure 6(b). Since at any moment of time $x_{\text{V}}(t) \equiv 1 - x_{\text{G}}(t) - x_{\text{D}}(t)$, it is enough to consider an ODE system for two sub-populations $x_{\text{D}}(t)$ and $x_{\text{G}}(t)$. To simplify the notation, we will write

$$\begin{cases} \dot{y}(t) = f\Big(y(t), d(t)\Big), \\ y(0) = x, \end{cases} \tag{3S.1}$$

where $y(t) = (x_{\text{D}}(t), x_{\text{G}}(t))$, $x = (x_{\text{D}}(0), x_{\text{G}}(0))$ and $f(\cdot)$ denotes the right-hand side of (2.2) in $(x_{\text{D}}, x_{\text{G}})$ coordinates using the transformation (2.1).

Our approximation scheme is based on a first-order accurate semi-Lagrangian discretization [7]. 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

$$\tilde{x}_d = x + \tau f(x, d). \tag{3S.2}$$

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

$$u(x) \;=\; \min_d \left\{\tau K(d) \;+\; u\left(\tilde{x}_d\right)\right\} \;+\; o(\tau). \tag{3S.3}$$

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 [8, 13]. More specifically, suppose that a vector $f(x, d)$ anchored

at x lies within a triangle $xx_{1d}x_{2d}$; see Figure 6(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:

$$U(x) \;=\; \min_{d\in\{0,\,d_{max}\}} \left\{ \frac{||x-\tilde{x}_d||}{||f(x,d)||}K(d) \;+\; \frac{||x_{2d}-\tilde{x}_d||}{||x_{1d}-x_{2d}||}U(x_{1d}) \;+\; \frac{||x_{1d}-\tilde{x}_d||}{||x_{1d}-x_{2d}||}U(x_{2d}) \right\}, \tag{3S.4}$$

which must hold for each meshpoint $x \in \Omega$. The boundary conditions are handled by setting $U = 0$ when $x_{\text{G}} < r_b$ and using any prohibitively expensive* exit cost (e.g., $U = 10^5$ in our implementation) whenever $x_{\text{G}} > 1 - f_b$.

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) [4, 12, 14]. 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 [1, 10, 13] 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 [5]) to therapy optimization problems, particularly for the case of small σ.

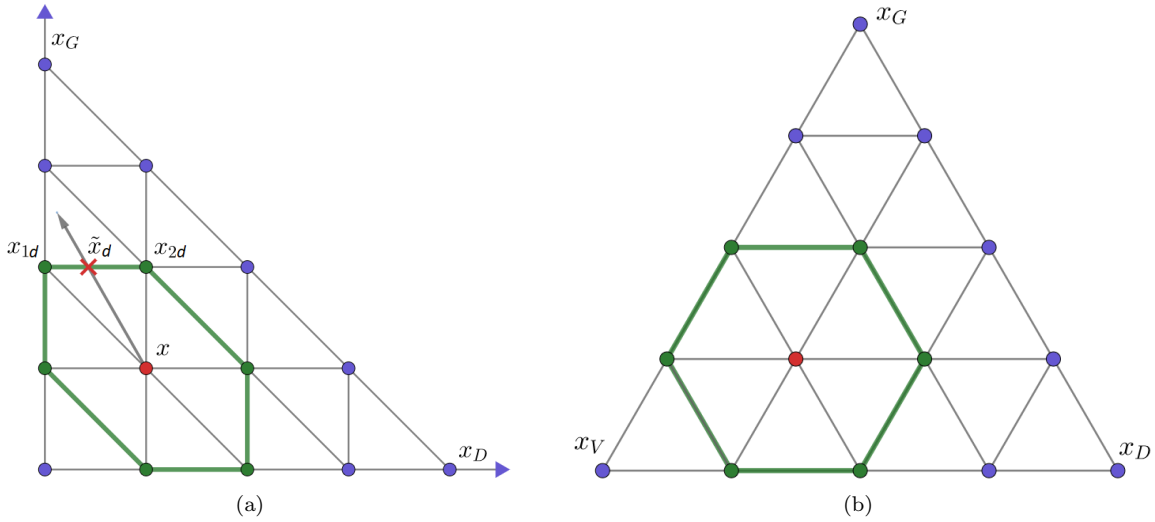


Figure 6: A Semi-Lagrangian scheme on a triangular mesh.
(a): A semi-Lagrangian discretization in $(x_{\text{D}}, x_{\text{G}})$.
(b): Linear transformation yields a regular triangular mesh.

For implementation purposes, it is easier to conduct computations in a Cartesian coordinate system (Figure 6(a)), which is equivalent to a regular triangular mesh (Figure 6(b)) by a linear transformation. To ensure the accuracy of

*We are not using $U = +\infty$ in the failure zone ($x_{\text{G}} > 1 - f_b$) because of the numerical diffusion stemming from the interpolation in semi-Lagrangian discretizations.

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 [7] on a coarser mesh, employing "Ordered Upwind" techniques to reduce or eliminate the coupling in the discretized system.

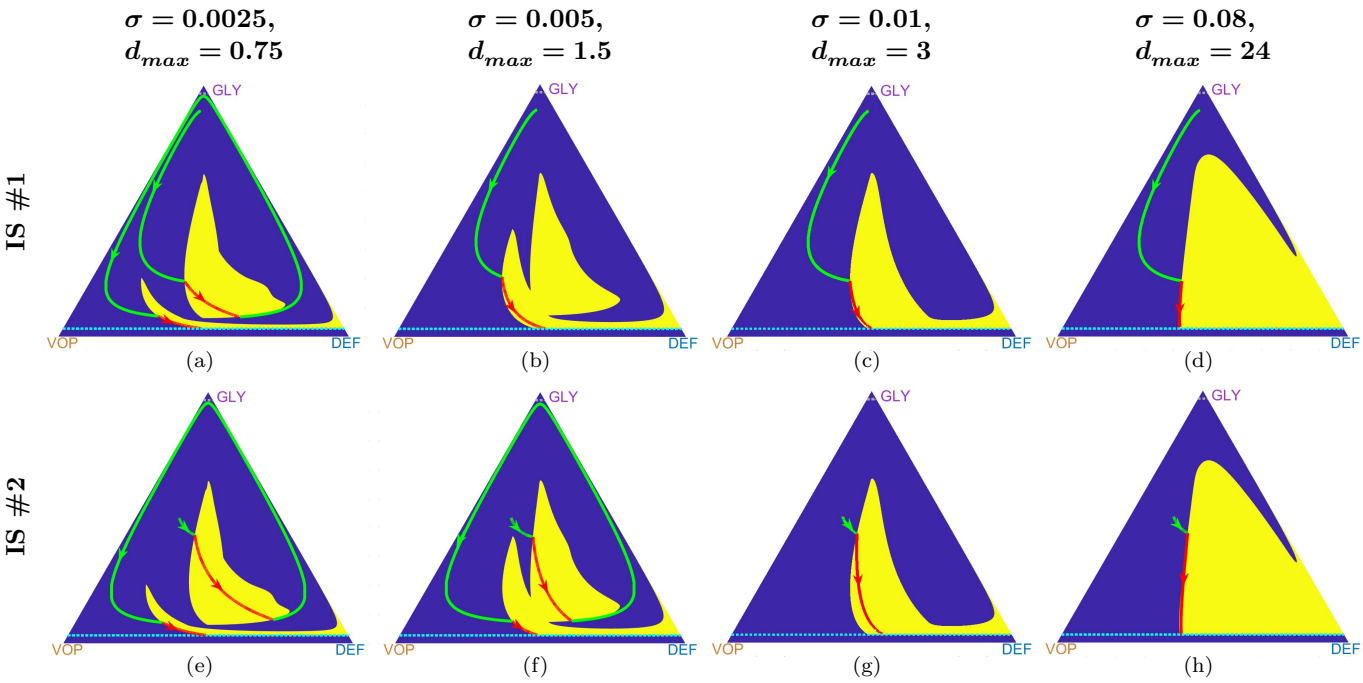
4S Model parameters used in the main text

Figure	Initial state of trajectories in $(x_{\rm D}, x_{\rm G}, x_{\rm V})$ coordinates	Parameters						
		d_{max}	σ	b_a	b_v	c	n	$r_b = f_b$
Figure 1	(0.04, 0.9, 0.06)	3 in 1(b)	—	2.5	2	1	4	$10^{-1.5}$
Figure 2	(0.04, 0.9, 0.06)	3	0.01	2.5	2	1	4	$10^{-1.5}$
Figure 3	$(0.04, 0.9, 0.06)$ in 3(a) $(0.417, 0.311, 0.272)$ in 3(b) denoted by (*) $(0.35, 0.3, 0.35)$ in 3(c) denoted by ($\times$)	3	0.01	2.5	2	1	4	$10^{-1.5}$
Figure 4	$(0.02, 0.8, 0.18)$	0.3	0.03	4	2	1	4	$10^{-1.5}$ in 4(a),4(b),4(c) 10^{-1} in 4(d), 4(e)
Figure 5	—	0.3	—	4	2	1	4	$10^{-1.5}$ in 5(a), 5(b) 10^{-1} in 5(c), 5(d)

5S Optimization trade-offs: administered drugs vs. 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 7

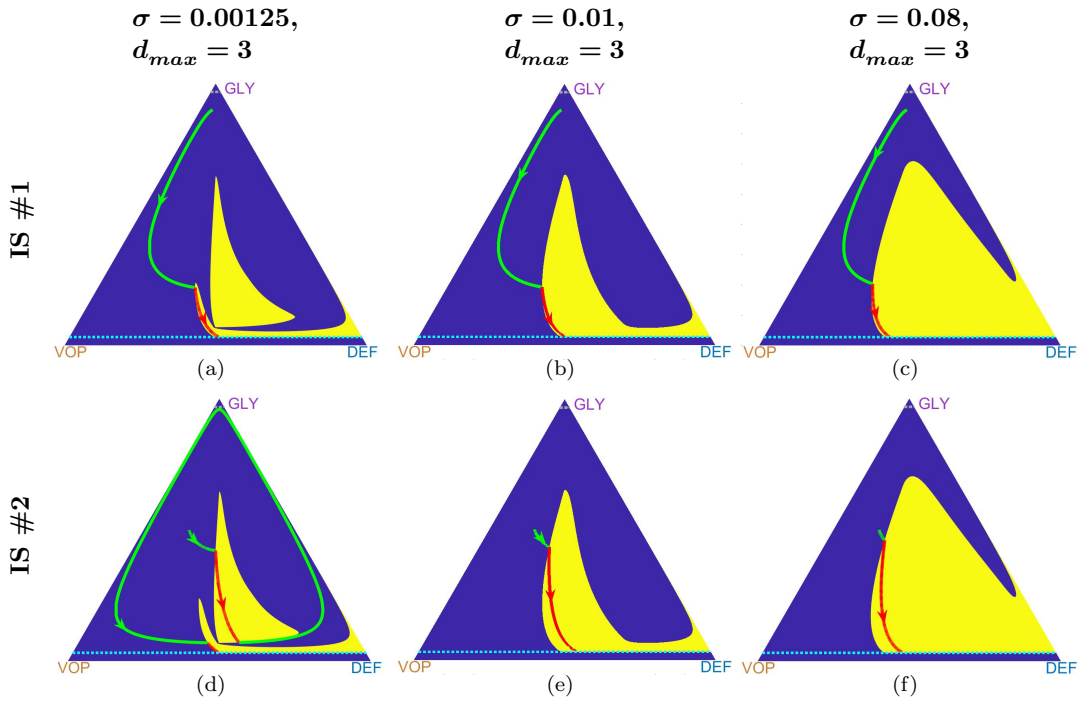
we conduct this experiment for two different initial states and the same set of game parameters ($b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$), demonstrating that the value of d_{max} strongly influences the optimal policies and the shape of the "therapy-on" yellow regions.



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 7: 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 (2.2).
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 8. 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.

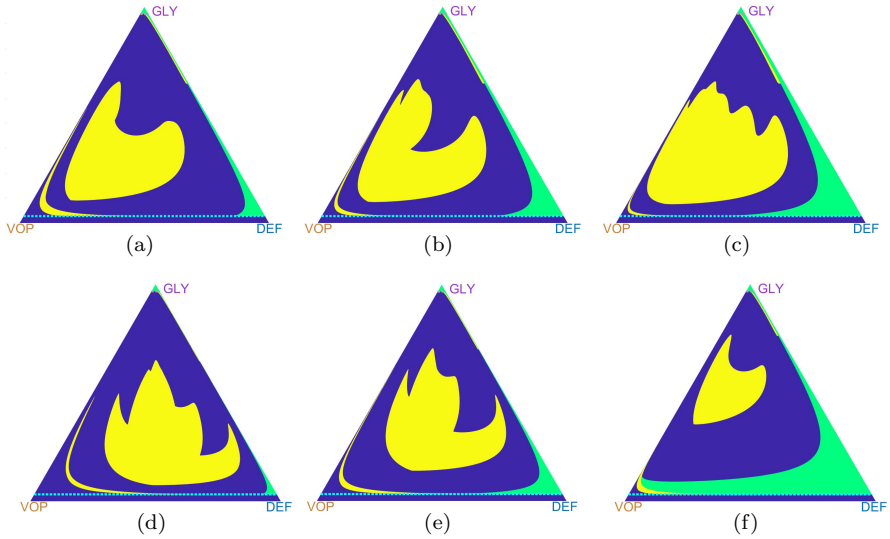


Parameters		Initial State (IS) #1 : $(x_{\mathrm{D}}, x_{\mathrm{G}}, x_{\mathrm{V}}) = (0.04, 0.9, 0.06)$				Initial State #2 : $(x_{\mathrm{D}}, x_{\mathrm{G}}, x_{\mathrm{V}}) = (0.15, 0.5, 0.35)$			
σ	d_{max}	subfigure	total time	total drugs	overall cost	subfigure	total time	total drugs	overall cost
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 8: 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:** $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$; $r_b = f_b = 10^{-1.5}$.

6S Incurable areas changing under parameter variation

In Figure 9 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 9: 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: $c = 1$, $n = 4$, $r_b = f_b = 10^{-1.5}$.

7S Fully angiogenic and glycolyctic tumors

Throughout the paper 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 (2.3). Two other types of tumours are also possible under the model [9]: 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 (2.2) satisfies this condition all cells tend to switch to VOP type and the trajectory converges to the recovery zone [9]. 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 10(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 [9]. We consider an example of optimal policy for fully glycolyctic tumour in Figure 10(b).

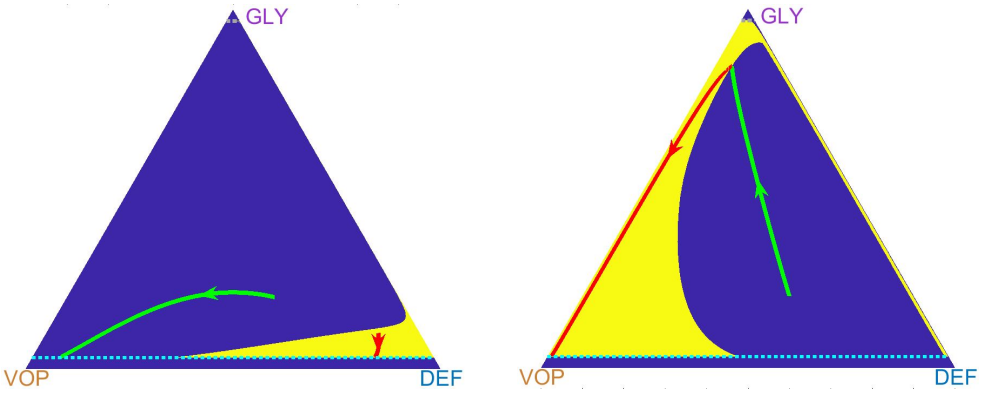


Figure 10: The benefits of optimization in fully angiogenic and glycolyctic cases.
(a): two trajectories under AT strategy in a *fully angiogenic* tumour. Parameters: $b_a = 2$, $b_v = 3$, $c = 1$, $n = 1$; $d_{max} = 3$, $\sigma = 0.3$, $r_b = f_b = 10^{-1.5}$.
(b): a trajectory under AT strategies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

References

[1] K. Alton and I. 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.

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

[3] R. Bellman. *Dynamic programming*. Princeton University Press, 1957.

[4] M. Boué and P. 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.

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

[6] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–1, 1983.

[7] M. Falcone and R. Ferretti. *Semi-Lagrangian Approximation Schemes for Linear and Hamilton-Jacobi Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[8] 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.

[9] A. Kaznatcheev, R. Vander Velde, J. G. Scott, and D. Basanta. Cancer treatment scheduling and dynamic

heterogeneity in social dilemmas of tumour acidity and vasculature. *British Journal of Cancer*, 116(6):785–792, 2017.

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

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

[12] J. Qian, Y. Zhang, and H. Zhao. Fast Sweeping Methods for Eikonal Equations on Triangular Meshes. *SIAM Journal on Numerical Analysis*, 45(1):83–107, 2007.

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

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