

Analysis of a Mathematical Model for Tumor Anti-Angiogenesis*

Urszula Ledzewicz[†]

Dept. of Mathematics and Statistics,
Southern Illinois University at Edwardsville,
Edwardsville, Illinois, 62026-1653,
`uledzew@siue.edu`

Heinz Schättler

Dept. of Electrical and Systems Engineering,
Washington University,
St. Louis, Missouri, 63130-4899,
`hms@wustl.edu`

July 12, 2006

Abstract

Anti-angiogenic therapy is a novel treatment approach for cancer that aims at preventing a tumor from developing its own blood supply system that it needs for growth. In this paper we consider a mathematical model where the stimulation term in the dynamics is proportional to the number of endothelial cells. This model is an example from a class of mathematical models for anti-angiogenic treatment that were developed and medically validated by Hahnfeldt, Panigrahy, Folkman and Hlatky [8]. The problem how to schedule a given amount of angiogenic inhibitors to achieve a maximum reduction in the primary cancer volume is considered as an optimal control problem and it is shown that optimal controls are bang-bang of the type $0a0$ with 0 denoting a trajectory corresponding to no treatment and a a trajectory with treatment at maximum dose along which all inhibitors are being exhausted.

1 Introduction

The reason for the failure of most cancer chemotherapy treatments lies in both intrinsic and acquired drug resistance. Malignant cancer cell populations are highly heterogeneous - the number of genetic errors present within one cancer cell can lie in the thousands [14] - and fast duplications combined with genetic instabilities provide just one of several mechanisms which allow for quickly developing acquired resistance to anti-cancer drugs. In addition, intrinsic resistance (i.e. the specific drug's activation mechanism simply doesn't work) makes some cancer cells not susceptible to many cytotoxic agents. "... the truly surprising thing is that some malignancies can be cured even with current approaches" [7, pp. 65]. Healthy cells (e.g. bone

*This material is based upon research supported by the National Science Foundation under grants No. DMS 0305965 and collaborative research grants DMS 0405827/0405848.

[†]U. Ledzewicz's research is also partially supported by SIUE 2005 Summer Research Fellowship.

marrow cells), on the other hand are genetically very stable and do not develop similar features [9]. So, while the cancer population becomes increasingly more resistant, the drugs keep on killing the healthy cells eventually leading to a failure of the therapy. Thus naturally the search for cancer treatment methods that would circumvent the problem of drug resistance is of tantamount importance. One such approach is tumor anti-angiogenesis.

A growing tumor, after it reaches just a few millimeters in size, no longer can rely on blood vessels of the host for its supply of nutrients, but it needs to develop its own vascular system for blood supply. In this process, called *angiogenesis*, there is a bi-directional signaling between endothelial cells, which provide the lining for the newly forming blood vessels of the tumor, and tumor cells stipulating growth. Endothelial cells produce growth factors that stimulate the proliferation of the tumor cell population and the major targets of pharmacologic therapies are vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF). Overall, angiogenesis can be viewed as a complex balance of tightly regulated stimulatory and inhibitory mechanisms balanced by microenvironmental factors. Angiogenic inhibitors, like endostatin, target the endothelial cells aiming to prevent the tumor from developing its own blood vessel system and thus block its growth. The tumor, deprived of necessary nutrition, regresses. Since the treatment targets normal cells, no occurrence of drug resistance has been reported in lab studies. (These treatments, however, still are only in an experimental stage.) For this reason tumor anti-angiogenesis has been called a therapy resistant to resistance which provides a new hope in treatment of tumor type cancers [9].

By now several mathematical models for tumor anti-angiogenesis have been formulated in the literature (e.g. [8, 5, 4, 1, 6]). One of the earliest models as a dynamical system is the one by Hahnfeldt, Panigrahy, Folkman and Hlatky in [8]. This model was clinically validated in lab experiments and became the basis for several modifications and simplifications undertaken in an effort to both better understand the dynamical properties of the underlying mechanisms and to make the mathematical model easier and more tractable for analysis. The models considered by d'Onofrio and Gandolfi in [4] and Ergun, Camphausen and Wein in [5], respectively, are all variations of the underlying dynamics from [8]. A dynamical systems analysis of various treatment schedules (e.g. stability properties of equilibria) of different versions of the underlying model is performed in [4] and in [5] the scheduling of anti-angiogenic inhibitors is considered as an optimal control problem both for a stand alone monotherapy and in combination with radiotherapy. While the models considered in these papers are variations on the specific dynamics proposed by Hahnfeldt et al. in [8], in the papers by Agur, Arakelyan, Daugulis and Ginosar [1] and Forys, Kheifetz and Kogan [6] more generally dynamical properties of models for angiogenesis are investigated under minimal assumptions on the form of the growth functions describing the dynamics.

Starting with the paper by Ergun, Camphausen and Wein [5], several versions of the mathematical model by Hahnfeldt et al. [8] have been analyzed as an optimal control problem. In these formulations the objective is to minimize the size of the tumor at the end of therapy with a given amount of drug as constraint. A modified problem where the overall amount of inhibitors is not restricted a priori, but is included in the objective functional was considered by us in [12]. In our paper [11] we developed a synthesis of optimal solutions for the model considered in [5] bringing the analysis of that paper for the monotherapy case to a conclusion. However, the dynamical system considered in this paper was a mathematical simplification of the original model and the question of how close optimal solutions of the various models are to each other comes up naturally. In [13] we therefore analyzed also the corresponding optimal

control problem for the dynamics as it was originally formulated in [8] and it turned out that both models indeed led to qualitatively equivalent structures: optimal controls are of the form “**Oasa0**” where **a** and **0** denote trajectories with *full*, respectively *no* anti-angiogenic therapy and **s** stands for a segment along an optimal *singular* arc. However, depending on the initial condition not all of these pieces need be present. Our theoretical analysis reduces the type of optimal controls to this structure, but possibly allows for a one-parameter family of extremals of this form. The optimal solution then is easily computed numerically based on our analysis. For the medically most typical and relevant scenarios optimal protocols take the form “**bs0**” with **b** standing for either **0** or **a**.

From a practical point of view, however, optimal singular arcs with their time-varying dosages are not very realistic and the question comes up how close the much simpler controls of the type “**Oa0**” come to the optimal ones. In this paper we consider a third variation of the underlying model which indeed leads to this class as optimal controls. The specific model considered here is also formulated in [8] and described there as a valid alternative to the model pursued further in that paper. This dynamics for this model was analyzed in [4] and here we consider an optimal control formulation. It will be shown that no singular arcs exist for this formulation and the analysis of bang-bang trajectories can be extended from our earlier papers to limit the possible concatenations to this simple form. Together with the conclusions for the other models [5, 11, 13], our results provide a complete classification of optimal controls for a class of mathematical models for tumor anti-angiogenesis based on the underlying dynamics formulated in [8]. Our analysis and conclusions are independent of the specific parameter values and lead to robust implications about the structure of optimal controls for this model.

2 Mathematical Model [8]

Tumor cells and endothelial cells both stimulate and inhibit each others growth. In the model developed by Hahnfeldt, Panigrahy, Folkman and Hlatky in [8] these effects are summarized in a two-dimensional dynamical system with the numbers of primary tumor cells, p , and of vascular endothelial cells, q , as variables. A growth function describes the size of the tumor dependent on the volume of endothelial cells and is chosen as Gompertzian with a variable carrying capacity defined by q in the original model. Other models are equally realistic and are considered, for instance in [4] or [6], but here we stay with the original choice. Thus the rate of change in the volume of primary tumor cells is modelled as

$$\dot{p} = -\xi p \ln \left(\frac{p}{q} \right) \quad (1)$$

where ξ denotes a tumor growth parameter. Endothelial cells produce growth factors that stimulate the proliferation of the tumor cell population, but also have receptors which make them sensitive to inhibitors of inducers of angiogenesis like, for example, endostatin. The overall dynamics is a balance between stimulation and inhibition and its basic structure is of the form

$$\dot{q} = -\mu q + S(p, q) - I(p, q) - Guq \quad (2)$$

where μq describes the loss of endothelial cells due to natural causes (death etc.), I and S denote inhibition and stimulation terms, respectively, and Guq represents a loss of endothelial cells due to additional outside inhibition. The variable u represents the control in the system and

corresponds to the angiogenic dose rate while G is a constant that represents the anti-angiogenic killing parameter. Generally μ is small and often this term is negligible compared to the other factors and thus in the literature often μ is set to 0 in this equation.

In [8] a spatial analysis of the underlying consumption-diffusion model was carried out that led to the following two principal conclusions:

1. The inhibitor will impact endothelial cells in a way that grows like volume of cancer cells to the power $\frac{2}{3}$. The exponent $\frac{2}{3}$ arises through the interplay of the surface of the tumor through which the inhibitor needs to be released with the volume of endothelial cells.

Thus in [8] the inhibitor term is taken in the form

$$I(p, q) = dp^{\frac{2}{3}}q \quad (3)$$

with d a constant, the death rate. The second implication of the analysis in [8] is that:

2. The inhibitor term will tend to grow at a rate of $q^\alpha p^\beta$ faster than the stimulator term with $\alpha + \beta = \frac{2}{3}$.

However, here now there are many choices for α and β and this is one of the main sources for various models considered in the literature [4, 5]. In their original work [8] Hahnfeldt et al. select $\alpha = 1$ and $\beta = -\frac{1}{3}$ resulting in the stimulation term $S(p, q) = bp$ with b a constant, the birth rate. However, other choices are possible and, for example, choosing $\alpha = 0$ and $\beta = \frac{2}{3}$ results in an equally simple form

$$S(p, q) = bq \quad (4)$$

chosen in [4]. In this case the control term can be combined with the stimulation term as $(b - Gu)q$ and thus the control can be interpreted as lowering the birth-rate of endothelial cells. In [4] the dynamics of both models from [8] is analyzed and it is shown for the uncontrolled system that there exists a unique globally asymptotically stable equilibrium (which, however, of course is not viable medically). Adding a control term, this equilibrium can be shifted to lower values, or, depending on the parameter values, even eliminated altogether. In the latter case all trajectories converge to the origin in infinite time. This, in principle, would be the desired situation.

The problem then becomes how to administer a given amount of inhibitors to achieve the “best possible” effect. Following the approach taken by Ergun, Camphausen and Wein [5], for a free terminal time T we consider the problem to minimize the value $p(T)$ over all Lebesgue measurable functions $u : [0, T] \rightarrow [0, a]$ which satisfy a constraint on the total amount of anti-angiogenic treatment administered of the form

$$\int_0^T u(t)dt \leq A. \quad (5)$$

Here a is a maximum dose at which the inhibitors can be given. The optimal solution to this problem gives the maximum tumor reduction achievable with a given amount of inhibitors. However, depending on the form of stimulation term chosen, different solutions emerge. In [13] we have considered the original formulation considered by Hahnfeldt et al. [8] and shown that optimal controls typically are of the form $\mathbf{bsa0}$ with \mathbf{b} denoting an arc corresponding to one

of the two bang controls $u = 0$ or $u = a$, \mathbf{s} a piece along an optimal singular arc and \mathbf{a} and $\mathbf{0}$, respectively, denoting arcs along $u = a$ and $u = 0$. Changing the dynamics to (4) changes the qualitative structure of solutions and in this paper we use this modified dynamics and consider the following optimal control problem:

[P] For a free terminal time T , minimize the value $p(T)$ subject to the dynamics

$$\dot{p} = -\xi p \ln\left(\frac{p}{q}\right), \quad p(0) = p_0, \quad (6)$$

$$\dot{q} = q\left(b - (\mu + dp^{\frac{2}{3}} + Gu)\right), \quad q(0) = q_0, \quad (7)$$

$$\dot{y} = u, \quad y(0) = 0, \quad (8)$$

over all measurable functions $u : [0, T] \rightarrow [0, a]$ for which the corresponding trajectory satisfies $y(T) \leq A$.

As it is customary in optimal control formulations, we adjoin the constraint as third variable. The following statement about the dynamical behavior of the system is an easy corollary of the results proven in [4].

Proposition 2.1 *For any admissible control u and arbitrary positive initial conditions p_0 and q_0 the corresponding solution (p, q) exists for all times $t \geq 0$ and both p and q remain positive.*
□

3 The Dynamical Systems for Constant Controls

For the analysis of the optimal control problem it is useful to understand the dynamic properties of the systems for a constant control $u \equiv v$ with v some value in the control set $[0, a]$. Our statements here are only minor extensions of the analysis given in the paper by d’Onofrio and Gandolfi [4] and we refer the reader to this paper for the proofs about our claims of stability properties of the equilibria. All statements are for the natural domain $\mathbb{R}_+^2 = \{(p, q) : p > 0, q > 0\}$ of the system. The uncontrolled system ($u = 0$) has a unique globally asymptotically stable focus at (\bar{p}, \bar{q}) given by $\bar{p} = \bar{q} = \left(\frac{b-\mu}{d}\right)^{\frac{3}{2}}$ [4]. This value naturally is far too high to be acceptable and it does not make sense to consider trajectories that would increase beyond \bar{p} . In order to exclude irrelevant discussions about the structure of optimal controls in regions where the model does not represent the underlying medical problem to begin with, we henceforth restrict our discussions to the following domain \mathcal{D} ,

$$\mathcal{D} = \{(p, q) : 0 < p \leq \bar{p}, 0 < q \leq \bar{q}\}, \quad (9)$$

restricted in both variables p and q by the equilibrium for the dynamics with $u = 0$. By increasing the value v of the control, the equilibrium can be shifted towards the origin along the diagonal and finally be eliminated altogether. As a function of v the equilibrium is given by

$$\bar{p}(v) = \bar{q}(v) = \left(\frac{b - \mu - Gv}{d}\right)^{\frac{3}{2}} \quad (10)$$

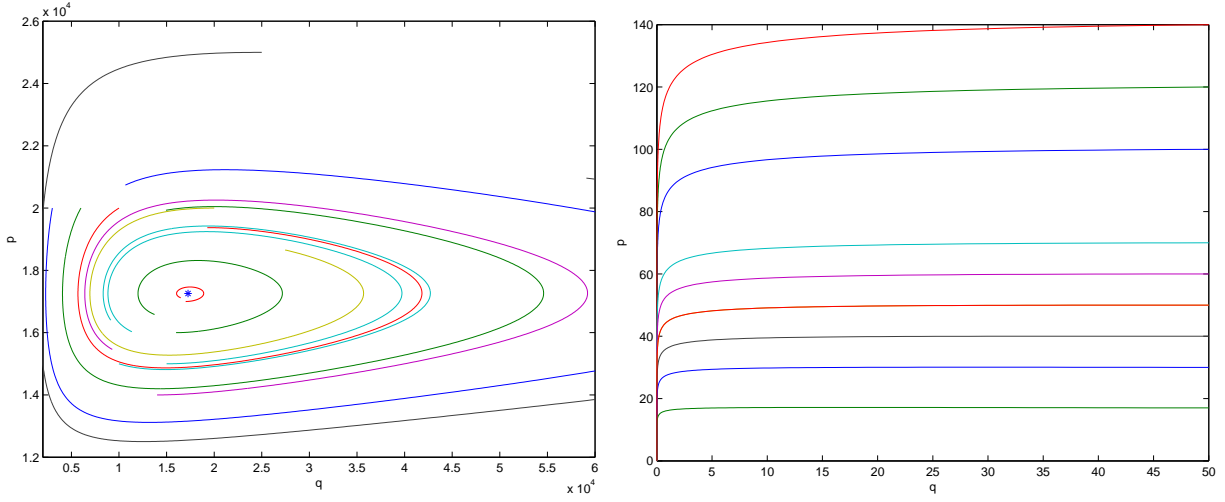


Figure 1: Phase portraits for $u = 0$ (left) and $u = a$ (right)

provided $b - \mu > Gv$, and this equilibrium $(\bar{p}(v), \bar{q}(v))$ still is globally asymptotically stable. As $b - \mu \leq Gv$, the system no longer has an equilibrium point and now all trajectories converge to the origin as $t \rightarrow \infty$ [4]. Thus, theoretically eradication of the tumor were possible in this case under the unrealistic scenario of constant treatment with unlimited supply of inhibitors. Since this is the most desirable situation, for our analysis of the optimal control problem we also **assume** that

$$(A) \quad \mathbf{Ga} > \mathbf{b} - \mu > \mathbf{0}. \quad (11)$$

Fig. 1 shows the phase portraits of the uncontrolled system on the left and for $u \equiv a$ on the right. In the figures q is along the horizontal axis and p along the vertical axis.

However, the domain \mathcal{D} still contains initial conditions that give rise to degenerate cases that we want to exclude. Let $\mathcal{D}_+ = \{(p, q) \in \mathcal{D} : p > q\}$, $\mathcal{D}_0 = \{(p, q) \in \mathcal{D} : p = q\}$ and $\mathcal{D}_- = \{(p, q) \in \mathcal{D} : p < q\}$. Both the trajectories for the constant controls $u = 0$ and $u = a$ cross the diagonal portion \mathcal{D}_0 transversally: for $u = 0$ trajectories cross from \mathcal{D}_+ into \mathcal{D}_- while they cross in opposite direction from \mathcal{D}_- into \mathcal{D}_+ for $u = a$. Trajectories for $u = 0$ eventually leave the region \mathcal{D} through the boundary segment $\{(p, q) : 0 < p < \bar{p}, q = \bar{q}\}$ only to return through the segment $\{(p, q) : p = \bar{p}, 0 < q < \bar{q}\}$. Such a scenario is unrealistic for the problem and does not arise for optimal solutions. Henceforth we do not consider this aspect of the dynamics. Trajectories for $u = a$ converge to the origin as $t \rightarrow \infty$ in the region \mathcal{D}_+ . However, it follows from the dynamics for p , (6), that the p -value of all trajectories is decreasing in \mathcal{D}_+ and increasing in \mathcal{D}_- . As a result, for some initial conditions (p_0, q_0, y_0) with $(p_0, q_0) \in \mathcal{D}_-$ it is possible that the optimal time T is $T = 0$. This situation arises when the amount of available inhibitors simply is not sufficient to reach a point in the region \mathcal{D}_+ that would have a lower p -value than p_0 . This situation is illustrated qualitatively below in Figure 2.

In such a case it is not possible to decrease the tumor volume with the available amount of inhibitors. It is only possible to slow down the tumor's growth. Indeed it is correct that the best way of doing this is to give the full dose $u = a$ until all inhibitors run out - this follows from the

structure of optimal controls to be shown here - but this is not the mathematically “optimal” solution for problem [P]. This one is simply to do nothing and take $T = 0$. Since this introduces a number of degeneracies into the analysis, we make the following definition:

Definition 3.1 *We say an initial condition (p_0, q_0, y_0) with $(p_0, q_0) \in \mathcal{D}_-$ is ill-posed if for any admissible control it is not possible to reach a point (p, q) with $p < p_0$. In this case the optimal solution for the problem [P] is given by $T = 0$. Otherwise (p_0, q_0, y_0) is well-posed and the optimal time T will be positive.*

It is clear that all initial conditions with $(p_0, q_0) \in \mathcal{D}_+ \cup \mathcal{D}_0$ are well-posed (since p decreases in \mathcal{D}_+ and trajectories with $u = a$ enter \mathcal{D}_+ from \mathcal{D}_0). It is easily determined whether an initial condition (p_0, q_0, y_0) with $(p_0, q_0) \in \mathcal{D}_-$ is ill-posed once the structure of optimal controls has been determined. For our analysis of optimal controls, however, we *only consider well-posed initial conditions*.

4 Analysis of Optimal Controls

It follows from classical results that there exists an optimal solution to our problem [3]. First-order necessary conditions for optimality of a control u are given by the *Pontryagin Maximum Principle* [15, 2, 3]: If u_* is an optimal control defined over an interval $[0, T]$ with corresponding trajectory (p_*, q_*, y_*) , then there exist a constant $\lambda_0 \geq 0$ and an absolutely continuous co-vector, $\lambda : [0, T] \rightarrow (\mathbb{R}^3)^*$, (which we write as row-vector) such that **(a)** $(\lambda_0, \lambda(t)) \neq (0, 0)$ for all $t \in [0, T]$, **(b)** the adjoint equations hold with transversality conditions,

$$\dot{\lambda}_1 = \xi \lambda_1 \left(\ln \left(\frac{p_*(t)}{q_*(t)} \right) + 1 \right) + \frac{2}{3} \lambda_2 d \frac{q_*(t)}{p_*^{\frac{1}{3}}(t)}, \quad \lambda_1(T) = \lambda_0, \quad (12)$$

$$\dot{\lambda}_2 = -\xi \lambda_1 \frac{p_*(t)}{q_*(t)} + \lambda_2 \left(b - \mu - d p_*^{\frac{2}{3}}(t) - Gu \right), \quad \lambda_2(T) = 0, \quad (13)$$

$$\dot{\lambda}_3 = 0, \quad \lambda_3(T) = \begin{cases} 0 & \text{if } y(T) < A \\ \text{free} & \text{if } y(T) = A \end{cases}, \quad (14)$$

and **(c)** the optimal control u_* minimizes the Hamiltonian H ,

$$H = -\lambda_1 \xi p \ln \left(\frac{p}{q} \right) + \lambda_2 q \left(b - \mu - d p^{\frac{2}{3}} - Gu \right) + \lambda_3 u, \quad (15)$$

along $(\lambda(t), p_*(t), q_*(t))$ over the control set $[0, a]$ with minimum value given by 0.

We call a pair $((p, q, y), u)$ consisting of an admissible control u with corresponding trajectory (p, q, y) for which there exist multipliers (λ_0, λ) such that the conditions of the Maximum Principle are satisfied an *extremal* (pair) and the triple $((p, q, y), u, (\lambda_0, \lambda))$ is an extremal lift (to the cotangent bundle). Extremals with $\lambda_0 = 0$ are called abnormal while those with a positive multiplier λ_0 are called normal. In this case it is possible to normalize $\lambda_0 = 1$. The following Lemmas summarize some elementary properties of optimal controls and extremals for well-posed initial conditions.

Lemma 4.1 *If u_* is an optimal control with corresponding trajectory (p_*, q_*, y_*) , then at the final time $p_*(T) = q_*(T)$ and $y_*(T) = A$, i.e. all available inhibitors have been used up.*

Proof. Since the p -dynamics is Gompertzian, (6), the number of cancer cells is growing for $p < q$ and is shrinking for $p > q$. This implies that optimal trajectories can only terminate at times where $p_*(T) = q_*(T)$. For, if $p_*(T) < q_*(T)$, then it would simply have been better to stop earlier since p was increasing over some interval $(T - \varepsilon, T]$. (Recall that we are assuming that the initial condition is well-posed so that the optimal final time T is positive.) On the other hand, if $p_*(T) > q_*(T)$, then we can always add another small interval $(T, T + \varepsilon]$ with the control $u = 0$ without violating any of the constraints and p will decrease along this interval if ε is small enough. Thus at the final time necessarily $p_*(T) = q_*(T)$. If now $y(T) < A$, then we can still add a small piece of a trajectory for $u = a$ over some interval $[0, \varepsilon]$. Since $\dot{q} < 0$ on the diagonal \mathcal{D}_0 the corresponding trajectory lies in \mathcal{D}_+ and thus the value of p is decreasing along this trajectory contradicting the optimality of T . \square

Lemma 4.2 *Extremals are normal. The multipliers λ_1 and λ_2 have simple zeros and cannot vanish simultaneously; λ_3 is constant and non-negative.*

Proof. The multipliers λ_1 and λ_2 satisfy the homogeneous linear system (12) and (13) and thus they vanish identically if they vanish at some time t . If $\lambda_0 = 0$, then the nontriviality of $(\lambda_0, \lambda(t))$ implies that the multiplier λ_3 , which is constant, is not zero. The condition $H \equiv 0$ on the Hamiltonian therefore gives $u \equiv 0$, i.e. the initial condition is ill-posed. Thus, without loss of generality we may assume that $\lambda_0 = 1$ and hence λ_1 and λ_2 cannot vanish simultaneously. Furthermore, whenever $\lambda_1(t) = 0$, then $\dot{\lambda}_1(t) = \frac{2}{3}\lambda_2(t)dq_*(t)/p_*^{\frac{1}{3}}(t) \neq 0$ and whenever $\lambda_2(t) = 0$, then $\dot{\lambda}_2(t) = -\xi\lambda_1(t)\frac{p_*(t)}{q_*(t)} \neq 0$ and thus both λ_1 and λ_2 have simple zeroes. At the final time T it follows from $p_*(T) = q_*(T)$, the transversality condition $\lambda_2(T) = 0$, and the condition $H(T) \equiv 0$ that $\lambda_3 u_*(T) = 0$. If $\lambda_3 < 0$, then the function $\Phi(t) = \lambda_3 - \lambda_2(t)Gq_*(t)$ will be negative on some interval $(T - \varepsilon, T]$ and thus by the minimization condition (c) on the Hamiltonian the control must be given by $u_*(t) = a$ on this interval. Contradiction. Hence $\lambda_3 \geq 0$. \square

Lemma 4.3 *If $\lambda_3 = 0$, then the corresponding optimal control is constant over the interval $[0, T]$ and given by the control $u \equiv a$.*

Proof. In this case the Hamiltonian function reduces to

$$H = -\lambda_1 \xi p \ln \left(\frac{p}{q} \right) + \lambda_2 q \left(b - \mu - dp^{\frac{2}{3}} - Gu \right) \quad (16)$$

and thus the minimization condition (c) implies that

$$u_*(t) = \begin{cases} 0 & \text{if } \lambda_2(t) < 0 \\ a & \text{if } \lambda_2(t) > 0 \end{cases} \quad (17)$$

Since $\lambda_2(T) = 0$ and $\dot{\lambda}_2(T) = -\xi\lambda_1(T)\frac{p_*(T)}{q_*(T)} = -\xi < 0$, λ_2 is positive on some interval $(\tau, T]$ and here the control is given by $u_*(t) = a$. Since $p_*(T) = q_*(T)$, it follows that the trajectory entirely lies in \mathcal{D}_- as long as the control is $u \equiv a$. But then λ_2 cannot have another zero τ since otherwise $H(\tau) = -\lambda_1(\tau)\xi p(\tau) \ln \left(\frac{p(\tau)}{q(\tau)} \right) \neq 0$. Hence λ_2 will be positive and thus the control must be constant $u \equiv a$. \square

Except for this extremely degenerate case (the initial condition is such that with giving the full dose we reach the diagonal exactly when all inhibitors have been exhausted) we can, as we henceforth do, without loss of generality therefore assume that λ_3 is positive.

Lemma 4.4 *If $\lambda_3 > 0$, then optimal controls end with an interval $(\tau, T]$ where $u_* \equiv 0$.*

Proof. In this case the function $\Phi(t) = \lambda_3 - \lambda_2(t)Gq_*(t)$ is positive on some interval $(T - \varepsilon, T]$.
□

The function

$$\Phi(t) = \lambda_3 - \lambda_2(t)Gq_*(t), \quad (18)$$

which determines the structure of the optimal control u_* through the minimization property (c) on the Hamiltonian H is called *switching function* of the problem and optimal controls satisfy

$$u_*(t) = \begin{cases} 0 & \text{if } \Phi(t) > 0 \\ a & \text{if } \Phi(t) < 0 \end{cases}. \quad (19)$$

A priori the control is not determined by the minimum condition at times when $\Phi(t) = 0$. If $\Phi(\tau) = 0$, but $\dot{\Phi}(\tau) \neq 0$, then the control switches between $u = 0$ and $u = a$ depending on the sign of $\dot{\Phi}(\tau)$. On the other hand, if $\Phi(t)$ vanishes identically on an open interval, then the minimization property in itself gives no information about the control. However, in this case also all derivatives of $\Phi(t)$ must vanish and this in fact may and typically does determine the control. Controls of this kind are called *singular* while we refer to the constant controls as *bang* controls. Optimal controls then need to be synthesized from these candidates through an analysis of the switching function. It is therefore clear that one needs to analyze the derivatives of the switching function.

The required computations can be expressed concisely within the framework of geometric optimal control theory and we therefore now write the state as $z = (p, q, y)^T$ and express the dynamics in the form

$$\dot{z} = f(z) + ug(z) \quad (20)$$

where

$$f(z) = \begin{pmatrix} -\xi p \ln\left(\frac{p}{q}\right) \\ \left(b - \mu - dp^{\frac{2}{3}}\right)q \\ 0 \end{pmatrix} \quad (21)$$

and

$$g(z) = \begin{pmatrix} 0 \\ -Gq \\ 1 \end{pmatrix}. \quad (22)$$

Using this notation, the adjoint equation can simply be expressed as

$$\dot{\lambda}(t) = -\lambda(t) (Df(z(t)) + u_*(t)Dg(z(t))) \quad (23)$$

where Df and Dg denote the matrices of the partial derivatives of the vector fields which are evaluated along $z(t)$. The derivatives of the switching function can easily be computed using the following well-known result that can be verified by an elementary direct calculation.

Proposition 4.1 *Let h be a continuously differentiable vector field and define*

$$\Psi(t) = \langle \lambda(t), h(z(t)) \rangle = \lambda(t)h(z(t)) \quad (24)$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on \mathbb{R} . Then the derivative of Ψ along a solution to the system equation (20) for control u and a solution λ to the corresponding adjoint equation (23) is given by

$$\dot{\Psi}(t) = \langle \lambda(t), [f + ug, h]z(t) \rangle, \quad (25)$$

where

$$[f, h](z) = Dh(z)f(z) - Df(z)h(z) \quad (26)$$

denotes the Lie bracket of the vector fields f and h . \square

Proposition 4.2 *The switching function Φ is three times continuously differentiable and optimal controls are bang-bang.*

Proof. For the switching function $\Phi(t) = \lambda_3 - \lambda_2(t)Gq_*(t) = \langle \lambda(t), g(z(t)) \rangle$ we have that

$$\dot{\Phi}(t) = \langle \lambda(t), [f, g]z(t) \rangle, \quad (27)$$

and

$$\ddot{\Phi}(t) = \langle \lambda(t), [f + ug, [f, g]]z(t) \rangle. \quad (28)$$

Direct calculations verify that

$$[f, g](z) = \xi Gp \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad [f, [f, g]](z) = \xi G \begin{pmatrix} \xi p \\ \frac{2}{3}dqp^{\frac{2}{3}} \\ 0 \end{pmatrix}, \quad (29)$$

and

$$[g, [f, g]](z) \equiv 0. \quad (30)$$

In particular therefore

$$\ddot{\Phi}(t) = \langle \lambda(t), [f, [f, g]]z(t) \rangle = \langle \lambda(t), ad^2 f(g)(z(t)) \rangle \quad (31)$$

where $ad(f)g = [f, g]$ and inductively $ad^n(f)g = [f, ad^{n-1}(f)g]$. Hence

$$\Phi^{(3)}(t) = \langle \lambda(t), [f + ug, ad^2 f(g)]z(t) \rangle. \quad (32)$$

But it follows from the Jacobi identity that also

$$[g, ad^2 f(g)] = -[f, [g, [f, g]]] = 0 \quad (33)$$

and thus for $i = 1, 2, 3$ we have

$$\Phi^{(i)}(t) = \langle \lambda(t), ad^i f(g)(z(t)) \rangle. \quad (34)$$

Hence Φ is three times continuously differentiable, regardless of the control u that is being used.

Suppose $\Phi(\tau) = 0$ for some time τ . If $\dot{\Phi}(\tau) \neq 0$, then the switching function changes sign at time τ and thus the corresponding control has a bang-bang switch. The derivative $\dot{\Phi}(t) = \xi G\lambda_1(t)p(t)$ vanishes at τ if and only if $\lambda_1(\tau) = 0$. But in this case we then have $\lambda_2(\tau) \neq 0$ and therefore

$$\ddot{\Phi}(\tau) = \frac{2}{3}\lambda_2(\tau)\xi Gd(\tau)p(\tau)^{\frac{2}{3}} \neq 0. \quad (35)$$

Hence, if $\dot{\Phi}(\tau) = 0$, then the switching function has a second-order contact point with 0 and no switching occurs. In particular, $\dot{\Phi}$ and $\ddot{\Phi}$ can never vanish simultaneously and therefore no singular controls exist for this model. Optimal controls are bang-bang with switchings at the simple zeros of the switching function. \square

We now analyze the possible switchings between $u = 0$ and $u = a$. By only considering trajectories that are relevant for the underlying medical problem we can restrict the class of optimal controls significantly. Henceforth we only consider trajectories that lie in the region \mathcal{D} . This region certainly contains all medically viable points and there is no point to analyze trajectories outside of \mathcal{D} since the model simply does not apply any longer.

Lemma 4.5 *Suppose the optimal control is given by $u \equiv 0$ over some maximal interval (α, β) with corresponding trajectory $z_* = (p_*, q_*, y_*)$. If the projection (p_*, q_*) lies in the region \mathcal{D} , then α and β cannot both be switching times.*

Proof. Suppose the optimal control is $u_* \equiv 0$ on an interval (α, β) and both α and β are switching times. Then the switching function Φ is positive over this interval and has a maximum at some time $\tau \in (\alpha, \beta)$ where $\dot{\Phi}(\tau) = 0$ and $\ddot{\Phi}(\tau) \leq 0$. Thus we have $\lambda_1(\tau) = 0$ and $\lambda_2(\tau) \leq 0$. But by Lemma 4.2 $\lambda_2(\tau)$ cannot vanish and thus it is negative. Furthermore, along $u \equiv 0$ the Hamiltonian (15) reduces to

$$H = -\lambda_1 \xi p \ln \left(\frac{p}{q} \right) + \lambda_2 q \left(b - \mu - dp^{\frac{2}{3}} \right) \equiv 0 \quad (36)$$

and therefore we must have $p(\tau) = \bar{p}$, the equilibrium value for the dynamics for $u = 0$. But then the trajectory either before or after time τ has values that exceed \bar{p} and thus does not lie in \mathcal{D} . Contradiction. \square

Lemma 4.6 *Optimal trajectories that lie in \mathcal{D} do not cross over from \mathcal{D}_+ into \mathcal{D}_- . They only leave the region \mathcal{D}_+ when they terminate on \mathcal{D}_0 at the final time T .*

Proof. Trajectories can only cross the diagonal from \mathcal{D}_+ into \mathcal{D}_- with control $u = 0$. By the last Lemma we only need to consider the cases of an initial or a final segment. But since p is strictly increasing in \mathcal{D}_- , it is clear the optimal way to terminate the trajectory is on \mathcal{D}_0 (c.f. Lemma 4.1). Thus consider an initial segment and suppose the trajectory for $u = 0$ crosses from \mathcal{D}_+ into \mathcal{D}_- at some time κ . Since the control $u \equiv 0$ is clearly not optimal, there has to be another time $\rho > \tau$ when the trajectory crosses back from \mathcal{D}_- into \mathcal{D}_+ . Since \dot{p} is increasing in \mathcal{D}_- , the point $(p(\rho), q(\rho))$ lies higher on the diagonal than the point $(p(\kappa), q(\kappa))$ and at time ρ less inhibitors are available than there were at time κ since some needed to be used up to get back to the diagonal. It follows from the geometry of the integral curves of trajectories for the control $u \equiv a$ (see Fig. 1) that the trajectory starting at point $(p(\rho), q(\rho))$ cannot do better than the one starting at the lower point $(p(\kappa), q(\kappa))$. Thus it would have been better not to cross over. \square

Stronger statements about possible concatenations can be made by making some explicit variations. The required explicit Lie-algebraic computations are best carried out using exponential notation. We therefore briefly introduce this notation and review one of its basic formulas [16]. Given a point z_0 and a differentiable vector field Z , there exists a unique integral curve $z(\cdot)$ of the vector field Z that passes through z_0 at time $t = 0$. This curve is defined over a

maximal interval I and simply is the solution to the differential equation $\dot{z}(t) = Z(z(t))$ with initial condition $z(0) = z_0$. We henceforth denote the point $z(t)$ at time t on this curve by

$$z(t) = z_0 \exp(tZ). \quad (37)$$

Note that we let the flows of the vector fields act on the *right*. This agrees with standard Lie algebraic conventions and does streamline notation and computations. Thus, for example, $z_0 \exp(t(f + ag))$ denotes the point at time t on the trajectory corresponding to control $u \equiv a$ that starts at z_0 at time 0. The same notation will also be used for moving vectors along the flow of Z . Thus, if X is some tangent vector at z_0 , then

$$z_0 X \exp(tZ) \quad (38)$$

denotes the vector obtained by transporting X from z_0 to the point $z(t)$ along the integral curve $z(s) = z_0 \exp(sZ)$, $0 \leq s \leq t$. Computationally, $z_0 X \exp(tZ)$ is the vector obtained by integrating the variational equation of the vector field Z with initial condition given by X at time 0 along the integral curve $z(s)$ from 0 to t . In our computations the following identity (related to the adjoint representation on a Lie group) is crucial [16]: the curve $t \mapsto z_0 \exp(tZ) X \exp(-tZ)$ has a Taylor series expansion of the form

$$z_0 \exp(tZ) X \exp(-tZ) \asymp \left(\sum_{n=0}^{\infty} \frac{t^n}{n!} ad^n Z(X) \right) (z_0) = \exp(t \, adZ(X))(z_0) \quad (39)$$

where $adZ(X) = [Z, X]$ denotes the Lie bracket and $ad^n Z(X) = [Z, ad^{n-1} Z(X)]$ for $n > 1$. On the right hand side of (39) we use the more common notation with the point z_0 at which the Lie brackets are being evaluated on the right, but we will only use this notation for asymptotic computations whenever convenient.

Proposition 4.3 *Along optimal trajectories there are no switchings from $u = 0$ to $u = a$ at points $(\tilde{p}, \tilde{q}) \in \mathcal{D}_+$.*

Proof. Consider a reference trajectory ς that switches from $u = 0$ to $u = a$ at a point $\tilde{z} = (\tilde{p}, \tilde{q}, \tilde{y})$. An important aspect of the construction is to set up the variations so that they all take the same amount of inhibitors. For t and ε small, non-negative parameters, define variations $\gamma(\varepsilon, t)$ as

$$\gamma(\varepsilon, t) = \tilde{z} \exp(t(f + ag)) \exp(-\varepsilon f) \exp(-t(f + ag)). \quad (40)$$

Here, starting at \tilde{z} we follow the trajectory for some small time t and compare it with another trajectory that switches earlier by backtracking along the trajectory for $u = 0$ through this point for some time ε and then backtracking along the $u = a$ trajectory for time t (see Fig. 3).

Note that all trajectories in the family use up exactly the same amount of inhibitors, i.e. for all t and ε we have $y(\gamma(\varepsilon, t)) = \tilde{y}$. We can determine which of the two strategies is better by determining the location of the point $\gamma(\varepsilon, t)$ relative to the earlier part on the reference trajectory, $\varsigma(s) = \tilde{z} \exp(sf)$ for $s < 0$: if the curve

$$r \mapsto \tilde{z} \exp(t(f + ag)) \exp(-\varepsilon f) \exp(-r(f + ag)), \quad r > 0, \quad (41)$$

intersects the reference trajectory for a time $r < t$, then it follows that if we switch at this intersection point from the control $u = 0$ to $u = a$ and then follow the variation forward in time,

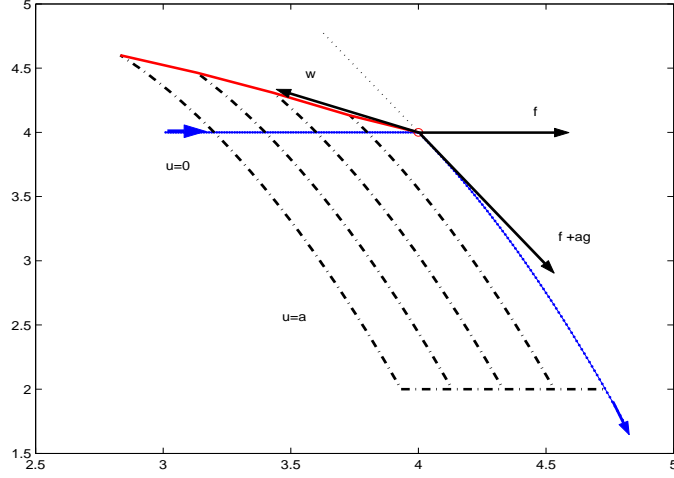


Figure 2: Variation $\gamma(\varepsilon, t)$

then we can reach the point $\tilde{z} \exp(t(f+ag))$ on the reference trajectory by using up less inhibitors (namely only ra for the variation compared with ta for the reference trajectory). But then in this case the reference trajectory cannot be optimal: from the point $\tilde{z} \exp(t(f+ag))$ onward, we can simply follow the reference trajectory until its final time. By Lemma 4.1 the endpoint lies on the diagonal. But since we saved some inhibitors during the variation, we can now still use them to lower the value for p further. Note that, although in our construction we backtrack curves, all of them give rise to admissible trajectories forward in time. The backtracking becomes necessary since we need to arrive on a point on the reference trajectory. Only this then allows to use the final step in the argument.

In order to determine whether the variation does better, we need to compare the geometric locations of the tangent vector of the reference trajectory at \tilde{z} , $f(\tilde{z}) + ag(\tilde{z})$, with the tangent vector of the variation γ at \tilde{z} ,

$$\begin{aligned} w &= \frac{\partial \gamma}{\partial \varepsilon}(0, t) = \tilde{z} \exp(t(f+ag)) (-f) \exp(-t(f+ag)) \\ &= -\exp(t \operatorname{ad}(f+ag)(f))(\tilde{z}) \end{aligned} \quad (42)$$

$$\begin{aligned} &= -f(\tilde{z}) - t[f+ag, f](\tilde{z}) + o(t) \\ &= -f(\tilde{z}) + at[f, g](\tilde{z}) + o(t). \end{aligned} \quad (43)$$

Since we keep track of the amount of inhibitors used, for this argument we can again ignore the third coordinate and consider g as a vector field on (p, q) -space only. The variations γ are better if its tangent vector w and the tangent vector of the reference trajectory at \tilde{z} , $f(\tilde{z}) + ag(\tilde{z})$, point to opposite sides of the vector $f(\tilde{z})$. For in this case there exists the desired intersection for $r < t$ (see Fig. 3). This is equivalent to $g(\tilde{z})$ and $[f, g](\tilde{z})$ pointing to opposite sides of $f(\tilde{z})$. This is easily computed. We have

$$f(p, q) \wedge g(p, q) = \begin{vmatrix} -\xi p \ln\left(\frac{p}{q}\right) & 0 \\ \left(b - \mu - dp^{\frac{2}{3}}\right)q & -Gq \end{vmatrix} = \xi G p q \ln\left(\frac{p}{q}\right) \quad (44)$$

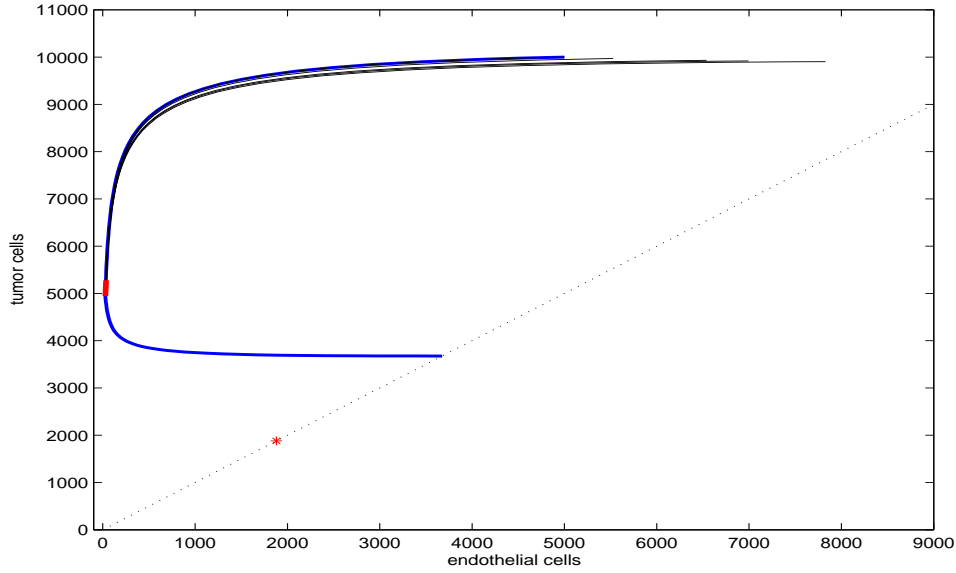


Figure 3: A family of $0a0$ -trajectories starting in \mathcal{D}_+

and

$$f(p, q) \wedge [f, g](p, q) = \begin{vmatrix} -\xi p \ln\left(\frac{p}{q}\right) & \xi G p \\ (b - \mu - dp^{\frac{2}{3}}) q & 0 \end{vmatrix} = -\xi G p q \left(b - \mu - dp^{\frac{2}{3}}\right). \quad (45)$$

For $p < \bar{p}$, we always have $b - \mu - dp^{\frac{2}{3}} > 0$ and thus $f(p, q) \wedge [f, g](p, q)$ is always negative for these points. On the other hand, $f(p, q) \wedge g(p, q)$ is positive in \mathcal{D}_+ and negative in \mathcal{D}_- . Thus g and $[f, g]$ point to opposite sides of f in \mathcal{D}_+ and $0a$ -concatenations are not optimal there. \square

Figs. 3 and 4 illustrate Proposition 4.3 for points in \mathcal{D}_+ . For our numerical illustrations we use the following parameter values which are taken from [8]: $\xi = \frac{0.192}{\ln 10} = 0.084$ per day (adjusted to the natural logarithm), $b = 5.85$ mm per day, $d = 0.00873$ per mm per day, $G = 0.15$ kg per mg of dose, and for illustrative purposes we chose a small positive value for μ , $\mu = 0.02$. The initial condition was chosen as $p_0 = 10,000$ and $q_0 = 5,000$; the maximum limit on the inhibitors was $a = 30$ with maximum total amount $A = 75$. In Fig. 3 we show a parameterized family of $0a0$ -trajectories starting at (p_0, q_0) . The various initial conditions depicted here all lie on the 0 -trajectory starting at (p_0, q_0) . The optimal trajectory, whose final segment is shown in blue, is the one corresponding to the $a0$ -trajectory in accordance with the proposition above. The red curve is the curve of points at the time when all inhibitors have been used up and a blown-up picture of this curve is shown in Fig. 4. We summarize our results in the Theorem below.

Theorem 4.1 *Optimal controls are bang-bang with at most two switchings of the form $0a0$. All inhibitors are being used up along the a -trajectory. If the initial condition (p_0, q_0) lies in \mathcal{D}_+ ,*

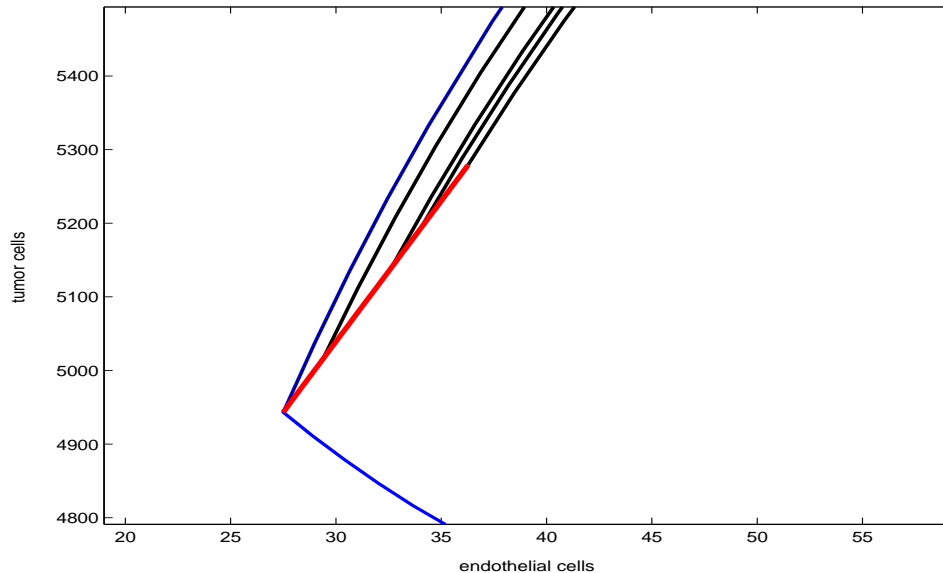


Figure 4: Detail of the family

optimal controls apply the full dose $u = a$ until all inhibitors have been exhausted and then follow a $u = 0$ trajectory to the diagonal \mathcal{D}_0 where the minimum value for p is attained.

5 Conclusion

In this paper we have considered a mathematical model for tumor anti-angiogenesis that was originally developed in [8] as an optimal control problem. For the problem of scheduling a given amount of angiogenic inhibitors in a way to minimize the tumor volume we have shown that optimal controls are bang-bang with at most two switchings in the order “**0a0**”. Except for some extreme initial conditions in \mathcal{D}_- in fact the optimal control is of the form “**a0**” immediately given all possible inhibitors. This indeed would be the practical choice to be pursued in therapy. The maximum tumor reduction is then realized as the trajectory crosses the diagonal \mathcal{D}_0 . This structure compliments the optimal strategies of the form “**0asa0**” that are found for the models considered in [5, 11] in the sense that even for these models the optimal strategies reduce to the form “**0a0**” in regions where the singular arc present in these models is no longer admissible. Therefore the model considered here, that does not have a singular arc, is close to the other models in these regions.

References

- [1] Z. Agur, L. Arakelyan, P. Daugulis and Y. Ginosar, Hopf point analysis for angiogenesis models, *Discrete and Continuous Dynamical Systems, Series B*, **4**, No. 1, (2004), pp. 29-38
- [2] A.E. Bryson and Y.C. Ho, *Applied Optimal Control*, Hemisphere Publishing, 1975
- [3] L. Cesari, *Optimization - Theory and Applications*, Springer Verlag, New York, 1983

- [4] A. d’Onofrio and A. Gandolfi, Tumour eradication by antiangiogenic therapy: analysis and extensions of the model by Hahnfeldt et al. (1999), *Math. Biosci.*, **191**, (2004), pp. 159-184
- [5] A. Ergun, K. Camphausen and L.M. Wein, Optimal scheduling of radiotherapy and angiogenic inhibitors, *Bull. of Math. Biology*, **65**, (2003), pp. 407-424
- [6] U. Forays, Y. Keifetz and Y. Kogan, Critical-point analysis for three-variable cancer angiogenesis models, *Mathematical Biosciences and Engineering*, **2**, no. 3, (2005), pp. 511-525
- [7] J.H. Goldie, Drug resistance in cancer: a perspective, *Cancer and Metastasis Review*, **20**, (2001), pp. 63-68
- [8] P. Hahnfeldt, D. Panigrahy, J. Folkman and L. Hlatky, Tumor development under angiogenic signaling: a dynamical theory of tumor growth, treatment response, and postvascular dormancy, *Cancer Research*, **59**, (1999), pp. 4770-4775
- [9] R.S. Kerbel, A cancer therapy resistant to resistance, *Nature*, **390**, (1997), pp. 335-336
- [10] A. Krener, The high-order maximal principle and its application to singular controls, *SIAM J. Control and Optimization*, **15**, (1977), pp. 256-293
- [11] U. Ledzewicz and H. Schättler, A synthesis of optimal controls for a model of tumor growth under angiogenic inhibitors, Proceedings of the 44th IEEE Conference on Decision and Control, Sevilla, Spain, December 2005, pp. 934-939
- [12] U. Ledzewicz and H. Schättler, Optimal control for a system modelling tumor anti-angiogenesis, *ICGST-ACSE Journal*, **6**, (2006), pp. 33-39
- [13] U. Ledzewicz and H. Schättler, Anti-Angiogenic Therapy in Cancer treatment as an Optimal Control Problem, *SIAM J. on Control and Optimization*, submitted
- [14] L.A. Loeb, A mutator phenotype in cancer, *Cancer Research*, **61**, (2001), pp. 3230-3239
- [15] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mishchenko, *The Mathematical Theory of Optimal Processes*, MacMillan, New York, (1964)
- [16] F.W. Warner, *Foundations of Differentiable Manifolds and Lie groups*, Springer Verlag, New York, 1983