Mathematical Biosciences 314 (2019) 28–42

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Numerical optimal control of a size-structured PDE model for metastatic cancer treatment☆

Jun Liu[a], Xiang-Sheng Wang*[,b]

[a] Department of Mathematics and Statistics, Southern Illinois University Edwardsville, Edwardsville, IL 62026, USA
[b] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70503, USA

ARTICLE INFO

Keywords:
Optimal control
Bang-bang control
Size-structured population model
Metastatic cancer treatment
Characteristic scheme
Projection gradient descent method

ABSTRACT

In this paper, we propose a unified size-structured PDE model for the growth of metastatic tumors, which extends a well-known coupled ODE-PDE dynamical model developed and studied in the literature. A treatment model based on the proposed unified PDE model is investigated via optimal control theory, where its first-order necessary optimality system characterizing the optimal control is derived. We prove that the uniqueness of the optimal control depends on the chosen objective functional, and the optimal control is of bang-bang type when it is unique. For obtaining its efficient numerical solutions, a projection gradient descent algorithm based on the characteristic scheme is developed for solving the established optimal treatment model. Several numerical examples are provided to validate our mathematical analysis and numerical algorithm, and also illustrate the biologically interesting treatment outcomes of different models and control strategies. Our simple model reveals that: (i) only the total drug dosage matters if one just cares about the final treatment output; (ii) given the same total drug dosage, the optimal bang-bang treatment plan outperforms the others in the sense that it maximally reduces the total tumor sizes during the whole period of treatment, although their final tumor sizes are the same.

1. Introduction

Cancer is a leading cause of death worldwide and many cancers remain incurable, although the understanding of cancer biology has been significantly improved over the past few decades. Being a major obstacle to effective cancer therapy, the spread of cancer tumor cells from one location to many other locations within the same organism, called *metastasis*, accounts for the majority of cancer-related deaths [14]. Unfortunately, the process of metastasis remains the least understood aspect of cancer biology [27,41]. For example, some recent studies [54] on breast cancer suggest that, contrary to more conventional thinking, metastatic dissemination can occur rather early in the course of cancer development, but the mechanism of such early dissemination is yet unknown. Such early occult metastasis [17] are undetectable by any standard diagnosis modalities. For instance, one may take a size of 10^8 cells (i.e., about 100 mm^3) as the threshold for a metastasis to be visible. Hence, it is very important to equip the clinicians with suitable tools to determine or measure the number or mass of metastases that are not or barely visible with current medical imaging techniques. For more advances on understanding metastasis, we refer to the recent review paper [27].

Quantitative approaches based on mathematical and computational modeling have become increasingly important in cancer treatment research [12,26,32,42], since insightful mathematical models built upon laboratory data can augment experimental and clinical studies by deepening our understanding of mechanisms driving tumorigenesis and identifying possible better treatment strategies [1,10]. In this paper, we will focus on studying a size-structured transport PDE model that generalizes the ODE-PDE dynamical model in [23] for characterizing the density or distribution of metastases. The advantage of the ODE-PDE model in [23] with dedicated numerical algorithms [6,15] is that we may be able to see beyond the current imaging technology and accurately estimate the number or mass of both visible and non-visible metastases. Since the pioneering work in [23], this type of metastatic model has been mathematically analyzed and experimentally studied and generalized in many different ways, such as adding appropriate treatment components to describe the interaction between tumor cells and chemotherapy/antiangiogenic therapy [8]. Remarkably, this metastatic model without treatments has been recently validated through laboratory experiments with tumor-bearing mice [16]. We believe this

☆ Jun Liu's research was supported by a Seed Grants for Transitional and Exploratory Projects (STEP) Award (FY2019) from the Graduate School at Southern Illinois University Edwardsville (SIUE).

* Corresponding author.
E-mail addresses: juliu@siue.edu (J. Liu), xswang@louisiana.edu (X.-S. Wang).

https://doi.org/10.1016/j.mbs.2019.06.001
Received 29 November 2018; Received in revised form 20 March 2019; Accepted 4 June 2019
Available online 07 June 2019

metastatic model will soon make its way to clinical trials.

To facilitate our following discussion, we now briefly introduce this elegant ODE-PDE metastatic model [23], which belongs to the family of age-structured McKendrick-Von Foerster equation [3,22]. It describes the dynamics of the colony size distribution of multiple metastatic tumors through two biological components:

(1) The growth of the primary tumor size, denoted by $x_p(t)$, follows a nonlinear ODE:

$$x_p'(t) = g_p(x_p(t)), \quad x_p(0) = x_0, \tag{1}$$

where x_0 denotes the initial tumor size, and $g_p(x)$ denotes the growth rate corresponding to different growth models. In [23], a Gompertz growth model with $g_p(x) = ax\ln(b/x)$ is used, where a is growth rate constant and $b > 1$ is the maximum tumor size (in number of cells).

(2) The evolution of the size-structured metastatic density, denoted by $\bar{\rho}(x, t)$, satisfies a transport PDE:

$$\bar{\rho}_t(x, t) + (g_m(x)\bar{\rho}(x, t))_x = 0, \quad t > 0, x \in (1, b), \tag{2}$$

endowed with a zero initial condition

$$\bar{\rho}(x, 0) = 0 \tag{3}$$

and a non-local boundary condition at $x = 1$

$$g_m(1)\bar{\rho}(1, t) = \beta(x_p(t)) + \int_1^b \beta(x)\bar{\rho}(x, t)dx, \quad t > 0. \tag{4}$$

Here $g_m(x)$ denotes the growth rate of metastases, which is usually assumed to be identical to $g_p(x)$. The colonization (birth) rate is chosen as $\beta(x) = \mu x^\alpha$, where the parameter $\alpha \in (0, 1]$ corresponds to an emission proportional to a fractal dimension of the emitting tumor. For example, it is chosen to be $\alpha = 2/3$ for tumors of sphere shape.

In the above coupled ODE-PDE model, from modeling perspective it is reasonably assumed that:

(i) There is no metastatic tumor at the initial time $t = 0$, which gives the initial condition $\bar{\rho}(x, 0) = 0$;
(ii) The primary and metastatic tumors emit new metastases at the same colonization (birth) rate $\beta(x)$;
(iii) The newly created metastases has a single cell, this leads to the boundary condition by noticing that the number of newly created metastatic cells per unit time at time t (the left-hand side of (4)) is the total rate of occurrences of metastases from both the primary tumor and metastatic tumors (corresponding to the first and second term of right-hand side of (4), respectively).

It is usually very difficult to find an analytic solution of such a PDE model. Only in some simple cases could one obtain an exact solution via Laplace transform; see [23]. Such an exact solution, however, is defined in terms of infinity series, and thus is very cumbersome for practical use and efficient computation. Therefore, it is more desirable to develop an efficient and accurate algorithm to solve the model system numerically. The main challenge in numerical algorithms for solving metastatic PDE model is not only caused by the nonlocal boundary condition, but also by the huge domain size. It is observed from clinical data that the tumor size x usually ranges from 1 cell to $b = O(10^8 \sim 10^{12})$ cells. Besides, the duration of metastatic process is typically measured in years, say, $T = 10 \sim 20$ years. Thus, many standard discretization schemes, whose approximation accuracy depends on the mesh step sizes, will become prohibitively expensive, since the computational domain $[1, b] \times [0, T]$ is too large to use any reasonably fine mesh.

In the literature, there are very few research works on efficient numerical methods for solving the above ODE-PDE model, although the numerical methods for general age-structured population models are well studied. In [6], following an earlier work [2] on general size-structured population models, the authors conducted a thoroughly mathematical analysis on the asymptotic behavior of the unique weak solution of the model, and also constructed a simple convergent scheme by computing the numerical solutions along the characteristic curves. Different from such characteristic schemes, the authors in [11] proposed to use a logarithmic change of variable $y = \ln(b/x)$ to map the original large domain $[1, b]$ into a far smaller domain $[0, \ln b]$. To solve the transformed PDE, they suggest combining a 5th-order WENO (Weighted Essentially Non-Oscillatory) scheme in space with a 3rd-order Runge–Kutta time discretization, which, however, lead to some noticeable discrepancy between the numerical solution and the exact solution. We provide in the Appendix A further discussion on why the proposed scheme in [11] would lead to inaccurate approximations. In more recent works [15,16], the authors proposed to reformulate the original PDE model into a Volterra integral equation, which was shown to have some advantages of getting more accurate numerical solutions with less computational costs (based on FFT techniques). However, such an integral equation approach also has its limitation. If treatment input is modelled with time-dependent growth rates, then the obtained Volterra equation will be not of convolution type and hence become more expensive to solve, due to unstructured dense system upon discretization of the integral equation.

To the best of our knowledge, the only references regarding the optimal treatment based on the above metastatic PDE model are Benzekry's two papers [8,9], where various combinations of the chemotherapy and anti-angiogenic therapy are considered. In this paper, we propose a different treatment model based on a unified PDE model and optimal control formulation, and also develop an efficient algorithm built upon the characteristic scheme for its accurate numerical solutions. The use of deterministic ODE optimal control theory to optimize cancer treatment (chemotherapy) is an old topic, see e.g. [31,45–47] and the references therein, but there are much less research work on PDE-based optimal control models for size-structured cancer treatment optimization, except a few marginally related papers (e.g., [21,29,40,50]). Our current paper contributes to advocate the application of PDE optimal control theory with numerical optimization algorithms to improve the treatment of metastatic cancer.

This paper is organized as follows. In Section 2, we present a unified optimal control model and discuss the uniqueness of optimal control under different objective functionals. In Section 3, we derive its first-order necessary optimality system and introduce a projection gradient descent (PGD) algorithm, where the characteristic schemes are extensively used for accurate discretization. Numerical examples are presented in Section 4. Finally, some conclusions are drawn in Section 5 and some additional discussion are included in the Appendices.

2. A unified optimal treatment model

Motivated by the assumption that both the primary tumor and metastatic tumor emit new metastases at the same colonization rate ($g_m = g_p = :g$; see [19,23,48]), it becomes more convenient to not treat them separately. Especially, in the context of treatment model, the contribution of primary tumor is of no more significance to the model system than other metastatic cells. It is thus reasonable to include the primary tumor into a unified density function of all tumor cells, rather than consider a separate equation for the primary tumor. More specifically, by introducing a Dirac delta density function for the primary tumor

$$\hat{\rho}(x, t) = \delta_{x_p(t)}(x),$$

and then defining a unified tumor density function (including both primary tumor and metastatic tumors)

$$\rho(x, t) = \bar{\rho}(x, t) + \hat{\rho}(x, t),$$

the aforementioned coupled ODE-PDE model (1)–(4) can be reformulated into a single PDE model

$$\begin{aligned} &\rho_t(x,\,t) + (g(x)\rho(x,\,t))_x = 0, \qquad 1 \le x \le b,\, 0 < t < T,\\ &g(1)\rho(1,\,t) = \int_1^b \beta(x)\rho(x,\,t)dx, \qquad t > 0,\\ &\rho(x,\,0) = \delta_{x_0}(x), \qquad 1 \le x \le b, \end{aligned} \tag{5}$$

where the initial condition $\delta_{x_0}(x)$ is derived from the initial condition $x_p(0) = x_0$, and

$$g(x) = ax\ln(b/x), \quad \beta(x) = \mu x^{\alpha}. \tag{6}$$

Such a unified PDE model would allow us to more easily model the situation with the metastatic tumors being present at the initial time, by choosing the initial condition to be a density function $\rho_0(x)$ profiling the initial tumors' size distribution. Based the solution regularity analysis in [6], the above PDE (5) in general admits only a unique weak solution $\rho \in C([0,\,T],\,L^1(1,\,b))$, which indeed becomes a strong solution if the compatibility condition (at $x = 1$) between the initial condition and the boundary condition holds. It was further shown in [15] that the unique weak solution ρ is continuously differentiable on the two domains separated by the characteristic curve starting at $x = 1$.

We now introduce a bilinear optimal treatment (control) model based on the above unified PDE model (5), where the treatment with chemotherapy drug dosage (or intensity) over time is denoted as a control/decision variable $u(t)$. For simplicity, we assume the chemotherapy drug has a uniform treatment effectiveness on all tumors that is independent of the tumor's size, which means the treatment term $u(t)$ only depends on time. This uniform treatment effectiveness assumption is typical from the modeling point of view; see, for example, [33]. For practical purpose, we also assume the treatment term $u(t)$ to be bounded by a given maximum dosage $\bar{u}$, i.e. $0 \le u(t) \le \bar{u}$. More specifically, we consider the following simple form of objective functional

$$\begin{aligned} \min_{0\le u(t)\le \bar{u}} \quad J_\theta(\rho,\,u) :=& \int_1^b w(x)\rho(x,\,T)dx + \theta\int_0^T\int_1^b w(x)\rho(x,\,t)dxdt\\ &+ \gamma\int_0^T u(t)dt, \end{aligned} \tag{7}$$

where $\rho(x,\,t)$ satisfies the unified PDE that describes the migration and growth of metastatic tumors

$$\begin{aligned} &\rho_t(x,\,t) + (g(x)\rho(x,\,t))_x = -d(u(t))\rho(x,\,t),\\ &1 \le x \le b,\, 0 < t < T,\\ &g(1)\rho(1,\,t) = \int_1^b \beta(x)\rho(x,\,t)dx, \qquad t > 0,\\ &\rho(x,\,0) = \rho_0(x), \qquad 1 \le x \le b. \end{aligned} \tag{8}$$

Here $w(x) \ge 0$ is a given weight function. Note that the weighted total metastases $\int_1^b w(x)\rho(x,\,t)dx$ is the total number of tumor cells (resp. the total metastatic mass) if $w(x) = 1$ (resp. $w(x) = x$). We denote by $d(u(t))$ the mortality (death) rate of metastatic tumors induced by drug treatment. For simplicity, we assume $d(u(t)) = u(t)$. Another major difference between the tumor growth model without treatment (5) and the treatment model (8) lies in the initial conditions. For the treatment model, we let $t = 0$ be the time when drug therapy is applied. In practice, the primary tumor has grown to a significant size (say 10^8 cells) and metastasis has already taken place at the time of cancer detection and drug treatment. Thus, the initial profile for treatment model is no longer a simple delta function, but a very general distribution function $\rho_0(x) \ge 0$. In the objective functional J_θ, we assume $\gamma \ge 0$ and $\theta \ge 0$. The first term of J_θ gives the weighted total metastases at the final time T. The second term of J_θ provides the weighted total metastases integrated over the whole treatment time interval $[0,\,T]$. The third term of J_θ denotes the weighted total drug dosage during the course of treatment. If the patient only cares about the outcomes at the end of treatment period, one may take $\theta = 0$, which allows the possible large growth of tumor during the course of treatment and also leads to non-uniqueness of optimal control; see below. We point out that our treatment model is very different from the 2D treatment model developed in [9], where the control (treatment) $u(t)$ is assumed to reduce the growth rate $g(x)$ to $g(x) - \mu x u(t)$ for some positive constant $\mu > 0$. Similar optimal control models have been studied in the context of age-structured population models [3,4], which are usually referred to as the maximum harvesting problem. But the analysis and algorithms developed in [3] can not be directly applied to our current treatment model, due to the significant difference in the objective functional.

The optimal treatment model (7) and (8) leads to a bilinear optimal control problem with a transport PDE as the state equation and also with boxed control constraints. Before presenting the optimization algorithms for solving (7) and (8), we need to first discuss the existence of an optimal solution to the optimal control problem, and then derive the first order necessary optimality conditions on the optimal control. The existence of an optimal control can be shown by following the similar arguments in [3], but its uniqueness will depend on the objective functional as discussed further below.

2.1. Non-uniqueness of optimal control for the J_0 model

We will solve the transport PDE (8) along its characteristic curves. First, we define a characteristic function

$$\chi(z) = b^{1-e^{-az}} \tag{9}$$

such that $\chi'(z) = g(\chi(z))$ and $\chi(0) = 1$. It is also clear that $\chi(z) \to b$ as $z \to \infty$. Next, we define $v(z,\,t) = g(\chi(z))\rho(\chi(z),\,t)$ to rewrite (8) as

$$\begin{aligned} &v_t(z,\,t) + v_z(z,\,t) = -u(t)v(z,\,t), \qquad z \ge 0,\ 0 < t < T,\\ &v(0,\,t) = \int_0^\infty \eta(z)v(z,\,t)dz, \qquad t > 0,\\ &v(z,\,0) = v_0(z), \qquad x \ge 0, \end{aligned} \tag{10}$$

where $\eta(z) = \beta(\chi(z)) = mb^{\alpha-\alpha e^{-az}}$ and $v_0(z) = g(\chi(z))\rho_0(\chi(z))$. Note that $g(\chi(z)) = (a\ln b)e^{-az}b^{1-e^{-az}}$. If we make a change of variable $x = \chi(z) = b^{1-e^{-az}}$, then we have $dx = g(x)dz$, $b^{-e^{-az}} = x/b$ and $\chi(t+z) = b^{1-e^{-az-at}} = b(x/b)^{e^{-at}}$. The first equation in (10) can be exactly solved along the characteristic lines:

$$v(z,\,t) = \begin{cases} v(0,\,t-z)e^{-\int_{t-z}^t u(s)ds}, & z \le t,\\ v(z-t,\,0)e^{-\int_0^t u(s)ds}, & z > t. \end{cases} \tag{11}$$

It then follows from the boundary condition and initial condition in (10) that

$$\begin{aligned} v(0,\,t) =& \int_0^t \eta(z)v(0,\,t-z)e^{-\int_{t-z}^t u(s)ds}dz\\ &+ \int_t^\infty \eta(z)v_0(z-t)e^{-\int_0^t u(s)ds}dz, \end{aligned} \tag{12}$$

which gives a Volterra-type integral equation [5,28] for $v(0,\,t)$. We introduce a function $\phi(t) = v(0,\,t)e^{\int_0^t u(s)ds}$. The above integral equation can be simplified as

$$\phi(t) = \int_0^t \eta(z)\phi(t-z)dz + f(t) = \int_0^t \eta(t-z)\phi(z)dz + f(t), \tag{13}$$

where the inhomogeneous term

$$\begin{aligned} f(t) =& \int_t^\infty \eta(z)v_0(z-t)dz = \int_0^\infty \eta(t+z)v_0(z)dz\\ =& \int_1^b mb^{\alpha}(x/b)^{\alpha e^{-at}}\rho_0(x)dx \end{aligned} \tag{14}$$

only depends on the initial condition. It is remarkable that $\phi(t)$ is independent of the treatment function $u(t)$. Moreover, we note that

$$\begin{aligned}\int_1^b w(x)\rho(x,t)dx &= \int_0^\infty w(\chi(z))v(z,t)dz\\ &= \int_0^t w(\chi(z))v(0,t-z)e^{-\int_{t-z}^t u(s)ds}dz\\ &\quad + \int_t^\infty w(\chi(z))v_0(z-t)e^{-\int_0^t u(s)ds}dz\\ &= e^{-\int_0^t u(s)ds}dz\Big[\int_0^t w(\chi(z))\phi(t-z)dz +\\ &\quad \int_t^\infty w(\chi(z))v_0(z-t)dz\Big]\\ &=: e^{-U(t)}\Phi(t),\end{aligned}$$

where

$$U(t)=\int_0^t u(s)ds \tag{15}$$

is the cumulative drug dosage during $[0, t]$, and

$$\begin{aligned}\Phi(t) &= \int_0^t w(\chi(z))\phi(t-z)dz + \int_0^\infty w(\chi(z+t))v_0(z)dz\\ &= \int_0^t w(\chi(z))\phi(t-z)dz + \int_1^b w(b(x/b)^{e^{-at}})\rho_0(x)dx\end{aligned} \tag{16}$$

is a quantity independent of the treatment function $u(t)$. Thus, the objective functional can be rewritten as

$$J_\theta = e^{-U(T)}\Phi(T) + \theta\int_0^T e^{-U(t)}\Phi(t)dt + \gamma U(T).$$

If $\theta = 0$, the objective functional J_0 is minimized when $U(T) = \ln[\Phi(T)/\gamma]$ and the minimum is given as $J_{\min} = \gamma(1 + U(T))$. This proves that the optimal strategy (i.e., the allocation of dosage within the time interval) is not unique because the same total dosage $U(T)$ with different dosage allocations $u(t)$, $t \in [0, T]$ have the same effect on the objective functional J_0. The non-uniqueness of optimal control will also be observed in numerical examples, where two different optimal treatment plans give the same minimal objective functional. In particular, we have proved that different treatment strategies with the same total drug dosage $U(T)$ will lead to the same final metastatic size $\rho(x, T)$. It thus seems more realistic to consider the J_θ model with $\theta > 0$ in clinical use; see [37].

2.2. Uniqueness of optimal control for the J_1 model

The optimal control of J_θ model with $\theta > 0$ turns out to be unique due to the extra path-dependent integral term in the objective functional. Denote $W(t) = e^{-U(t)}$. To show uniqueness mathematically, we have to prove that the objective functional

$$J_\theta(W) = W(T)\Phi(T) + \theta\int_0^T W(t)\Phi(t)dt - \gamma\ln W(T)$$

has a unique minimum for all continuous, positive, and non-increasing functions $W(t)$ such that $W(0) = 1$. Assuming that $u(t) \le \bar{u}$, we have an additional constraint that $W(t) \ge e^{-\bar{u}t}$. First, we show that the minimum of J_θ must be achieved at a bang-bang control (as a minimizer). For any $0 \le u(t) \le \bar{u}$ and fixed $T > 0$, we define

$$\tilde{u}(t) = \begin{cases}\bar{u}, & t \le \tau,\\ 0, & t > \tau,\end{cases}$$

where $\tau = U(T)/\bar{u}$. Denote

$$\tilde{W}(t) = e^{-\int_0^t \tilde{u}(s)ds}.$$

It is readily seen that $\tilde{W}(t) \le W(t)$ for all $t \in [0, T]$ and $\tilde{W}(T) = W(T)$. Therefore, $J_\theta(\tilde{W}) \le J_\theta(W)$ and we only need to consider minimizing the objective functional over the admissible set of bang-bang functions $\tilde{u}(t)$ with only one jump (switching) point τ. Note that

$$\tilde{W}(t) = \begin{cases}e^{-\bar{u}t}, & t \le \tau,\\ e^{-\bar{u}\tau}, & t \le \tau.\end{cases}$$

We can rewrite the objective functional as a single variable function of τ:

$$J_\theta(\tau) = e^{-\bar{u}\tau}\Phi(T) + \theta\int_0^\tau e^{-\bar{u}t}\Phi(t)dt + \theta\int_\tau^T e^{-\bar{u}\tau}\Phi(t)dt + \gamma\bar{u}\tau.$$

By taking its first-order and second-order derivatives with respective to τ, we observe

$$\begin{aligned}\frac{dJ_\theta}{d\tau} &= -\bar{u}e^{-\bar{u}\tau}\Phi(T) - \bar{u}\theta\int_\tau^T e^{-\bar{u}\tau}\Phi(t)dt + \gamma\bar{u}, \frac{d^2J_\theta}{d\tau^2}\\ &= \bar{u}^2e^{-\bar{u}\tau}\Phi(T) + \bar{u}\theta e^{-\bar{u}\tau}\Phi(\tau) + \bar{u}^2\theta\int_\tau^T e^{-\bar{u}\tau}\Phi(t)dt > 0.\end{aligned}$$

Thus, J_θ has a unique minimum at some $\tau^* \in [0, T]$. Note that

$$\begin{aligned}\frac{dJ_\theta}{d\tau}|_{\tau=0} &= -\bar{u}\Phi(T) - \bar{u}\theta\int_0^T \Phi(t)dt + \gamma\bar{u}, \frac{dJ_\theta}{d\tau}|_{\tau=T}\\ &= -\bar{u}e^{-\bar{u}T}\Phi(T) + \gamma\bar{u}.\end{aligned}$$

If $\gamma \ge \Phi(T) + \theta\int_0^T \Phi(t)dt$, then $\tau^* = 0$. This implies that no drug should be used if the side effect is too strong. If $\gamma \le e^{-\bar{u}T}\Phi(T)$, then $\tau^* = T$. This means that all possible drugs should be applied if the side effect is too weak. If $e^{-\bar{u}T}\Phi(T) < \gamma < \Phi(T) + \theta\int_0^T \Phi(t)dt$, then τ^* (i.e., the stopping time of drug treatment) is the unique root of the following nonlinear transcendental function

$$F(\tau) := \Phi(T) + \theta\int_\tau^T \Phi(t)dt - \gamma e^{\bar{u}\tau}.$$

It is easy to check that

$$F'(\tau) = -\theta\Phi(\tau) - \gamma\bar{u}e^{\bar{u}\tau}.$$

To numerically find the unique root τ^*, we may use a standard Newton iteration (e.g., starting with $\tau_0 = T/2$):

$$\tau_{n+1} = \tau_n + \frac{\Phi(T) + \theta\int_{\tau_n}^T \Phi(t)dt - \gamma e^{\bar{u}\tau_n}}{\theta\Phi(\tau_n) + \gamma\bar{u}e^{\bar{u}\tau_n}}.$$

Such a Newton iteration, however, is computationally expensive due to the integral term and also only locally convergent. Alternatively, we suggest to use more efficient direct search algorithm to find the unique minimum for the objective function $J_\theta(\tau)$. First, we discretize the interval $[0, T]$ by uniform grids: $\tau_i = i\Delta\tau$, and calculate $\Phi_i = \Phi(\tau_i)$ by solving the integral equation of ϕ. Then we may use trapezoidal rule and recurrence relation to calculate the two integrals at $\tau = \tau_i$:

$$I_1(\tau) = \theta\int_0^\tau e^{-\bar{u}t}\Phi(t)dt, \quad I_2(\tau) = \theta\int_\tau^T \Phi(t)dt.$$

Next, we evaluate $J_\theta(\tau_i) = e^{-\bar{u}\tau_i}(\Phi(T) + I_2(\tau_i)) + I_1(\tau_i) + \gamma\bar{u}\tau_i$. Finally, we use a loop to find the unique optimal index $i = i^*$ such that $J_\theta(\tau_{i^*}) \le J_\theta(\tau_i)$ for all i, and then set $\tau^* = \tau_{i^*}$. Once all $\Phi(\tau_i)$ are computed, the remaining computation cost is only linear in the number of mesh points. Obviously, the obtained τ^* by direct search has an accuracy of $\Delta\tau$, which is mainly used for verifying our following numerical algorithms.

3. Optimality system and numerical algorithm

We are now ready to derive the first-order necessary optimality system (also known as the Karush–Kuhn–Tucker (KKT) conditions in optimization community), which will be useful in the development of adjoint-based gradient descent optimization solvers [30]. Moreover, it shows the bang-bang structure [25,44,49] of the optimal control, which further validates our analysis on the uniqueness of optimal control.

3.1. First-order necessary KKT optimality system

We will use the standard Lagrangian approach [49] to derive the first-order necessary optimality system, which characterizes the optimal control and state. It is interesting to notice the nonlocal boundary condition in the state equation leads a nonstandard (nonlocal) adjoint-state equation that contains its boundary values.

Define the Lagrangian with a Lagrange multiplier p (using integration by parts in both x and t)

$$\begin{aligned} L(\rho, u, p) &= J(\rho, u) + \int_0^T \int_1^b (\rho_t + (g\rho)_x + u\rho)p dx dt \\ &= J(\rho, u) + \left[\int_1^b \rho(x, t)p(x, t)dx \right]_0^T - \int_0^T \int_1^b \rho p_t dx dt \\ &\quad + \left[\int_0^T g(x)\rho(x, t)p(x, t)dt \right]_1^b - \int_0^T \int_1^b g\rho p_x dx dt \\ &\quad + \int_0^T \int_1^b u\rho p dx dt \end{aligned}$$

By applying the initial and boundary conditions in (8), the objective functional in (7), and the fact that $g(b) = 0$, we have

$$\begin{aligned} L(\rho, u, p) &= \int_1^b (w(x)\rho(x, T) + \rho(x, T)p(x, T) - \rho_0(x)p(x, 0))dx \\ &\quad + \gamma \int_0^T u(t)dt \\ &\quad - \int_0^T \int_1^b (p_t + gp_x - up + \beta(x)p(1, t) - \theta w(x))\rho dx dt. \end{aligned}$$

Then, for its partial derivative w.r.t. ρ, in the direction y with $y(x, 0) = 0$, we have

$$\begin{aligned} 0 = L_\rho(\rho, u, p)y &= \int_1^b (w(x) + p(x, T))y(x, T)dx \\ &\quad - \int_0^T \int_1^b (p_t + gp_x - up + \beta(x)p(1, t) - \theta w(x))y \\ &\quad dx dt, \end{aligned}$$

which (due to the arbitrary choice of y) gives the so-called adjoint state equation:

$$\begin{aligned} p(x, T) &= -w(x), p_t + gp_x - up + \beta(x)p(1, t) \\ &= \theta w(x). \end{aligned}$$

Unlike the tumor growth equation in (8), no boundary condition is needed when we solve the adjoint state equation in backward time. This is because all characteristic curves of the above equation cannot exceed the bound $x = b$ and the final condition at $t = T$ is sufficient.

The optimality of u under the control constraints $u \in U_{ad}$: $=\{u \in L^\infty(0, T)$: $0 \le u(t) \le \bar{u}\}$ implies $L(\rho, u, p)$ is increasing in all admissible directions $(v - u)$ with $v \in U_{ad}$ (since U_{ad} is convex), that is

$$\begin{aligned} 0 \le L_u(\rho, u, p)(v - u) &= \gamma \int_0^T (v(t) - u(t))dt \\ &\quad + \int_0^T \int_1^b (v(t) - u(t))\rho p dx dt \\ &= \int_0^T \left(\gamma + \int_1^b \rho p dx \right)(v(t) - u(t))dt. \end{aligned}$$

Define $z(t) = \int_1^b \rho(x, t)p(x, t)dx$. We have

$$\left(\gamma + \int_1^b \rho(x, t)p(x, t)dx \right)(v - u(t)) \ge 0 \qquad \forall\ v \in [0, \bar{u}]$$

for almost every $t \in [0, T]$. This implies that the optimal control u is given by

$$u(t) = \begin{cases} 0 & \text{if } \gamma + z(t) > 0, \\ \in [0, \bar{u}] & \text{if } \gamma + z(t) = 0, \\ \bar{u} & \text{if } \gamma + z(t) < 0, \end{cases} \tag{17}$$

which is of bang-bang type if the set $\mathcal{A}_\gamma$: $=\{t \in [0, T]$: $\gamma + z(t) = 0\}$ is assumed to have a zero Lebesgue measure. In general, an optimal control $u(t)$ is called *bang-bang* if $u(t) \in \{0, \bar{u}\}$ for a.e. $t \in [0, T]$, i.e., the control constraints are active for almost all the time. To summarize, we have derived the following first necessary optimality KKT conditions.

Theorem 3.1. *A control $u \in U_{ad}$: $=\{u \in L^\infty(0, T)$: $0 \le u(t) \le \bar{u}\}$ with the associate state ρ is optimal for the optimal control problem (7–8) if and only if the corresponding Lagrange multiplier (adjoint state) p satisfies the following adjoint equation (marching backward in time from the right endpoint $x = b$)*

$$\begin{aligned} & p_t(x, t) + g(x)p_x(x, t) - u(t)p(x, t) + \beta(x)p(1, t) = \theta w(x), \\ & 1 \le x \le b,\ 0 < t < T, \\ & p(x, T) = -w(x), \qquad 1 \le x \le b, \end{aligned} \tag{18}$$

and the optimal control u is of bang-bang type as given in (17). Moreover, we have $\rho(x, t) > 0$ and $p(x, t) < 0$ for all $x \in [1, b)$ and $t \in [0, T]$.

Proof. We only need to prove the statement in the last sentence. From the Volterra integral Eq. (12), we have $v(0, t) > 0$, which together with (11) implies that $v(z, t) = g(\xi(z))\rho(\xi(z), t) > 0$ for all $z \ge 0$ and $t \ge 0$. The positiveness of $\rho(x, t)$ follows. Similarly, we define $P(z, t) = e^{-U(t)}p(\chi(z), t)$ and rewrite the adjoint Eq. (18) as

$$\begin{aligned} & P_t(z, t) + P_z(z, t) = -\beta(\chi(z))P(0, t) + \theta w(\chi(z))e^{-U(t)}, \\ & 0 \le z < \infty,\ 0 < t < T, \\ & P(z, T) = -w(\chi(z))e^{-U(T)}, \quad 0 \le z < \infty. \end{aligned}$$

Solving the above equation along the characteristic lines gives

$$\begin{aligned} P(z, t) &= -w(\chi(z + T - t))e^{-U(T)} \\ &\quad + \int_0^{T-t} [\beta(\chi(z + s))P(0, t + s) - \theta w(\chi(z + s))e^{-U(t+s)}]ds. \end{aligned} \tag{19}$$

Especially, by choosing $z = 0$, we have the Volterra integral equation

$$\begin{aligned} P(0, t) &= -w(\chi(T - t))e^{-U(T)} \\ &\quad + \int_0^{T-t} [\beta(\chi(s))P(0, t + s) - \theta w(\chi(s))e^{-U(t+s)}]ds, \end{aligned}$$

from which we have $P(0, t) < 0$ for all $t \in [0, T]$. Substituting this into (19) gives $P(z, t) < 0$ for all $z \ge 0$ and $t \in [0, T]$. Consequently, $p(x, t) < 0$ for all $x \in [1, b]$ and $t \in [0, T]$ □

It is readily seen that $\gamma = 0$ (side effect is ignored) should give the optimal control $u(t) = \bar{u}$ for all $t \in [0, T]$, which agrees with the intuition that all possible drugs should be used. On the other hand, if $\gamma > 0$ is large enough (the drug is too poisonous) such that $\gamma + z(t) > 0$ for all $t \in [0, T]$, then we would expect the trivial optimal control $u(t) = 0$. The values of $z(t)$, however, is not easy to compute since it depends globally on the values of ρ and p. For an appropriately chosen $\gamma > 0$, we anticipate to observe the typical bang-bang control with only one switching point.

3.2. A projection gradient descent (PGD) algorithm

For numerically solving PDE-constrained optimal control problems, there are two types of algorithms: (i) optimize-then-discretize (OD) and (ii) discretize-then-optimize (DO). In OD algorithms, one discretizes the first-order necessary optimality system and then solves the discretized linear/nonlinear optimality system with efficient iterative linear/nonlinear solvers. While in DO algorithms, one directly discretizes the original optimization problem to obtain a fully discretized large-scale

finite-dimensional optimization problem, which is then solved by any existing optimization solvers. Both the OD and DO algorithms have their own advantages. Intuitively, the DO approach is more straightforward to implement, since the optimization can be automatically operated by black-box optimization solvers, but it may become less efficient unless the gradient and Hessian information can be analytically provided, especially for large scale problems. We will develop an OD type algorithm, which seems to be more efficient for our considered problems.

Recall the first-order necessary optimality KKT PDE system reads

$$\begin{cases} \rho_t(x,\,t) + (g(x)\rho(x,\,t))_x + u(t)\rho(x,\,t) = 0, \\ 1 \le x \le b,\ 0 < t < T, \\ g(1)\rho(1,\,t) = \int_1^b \beta(x)\rho(x,\,t)dx, \qquad t > 0, \\ \rho(x,\,0) = \rho_0(x), \qquad 1 \le x \le b, \\ p_t(x,\,t) + g(x)p_x(x,\,t) - u(t)p(x,\,t) + \beta(x)p(1,\,t) = \theta w(x), \\ 1 \le x \le b,\ 0 < t < T, \\ p(x,\,T) = -w(x), \qquad 1 \le x \le b, \\ (\gamma + \int_1^b \rho(x,\,t)p(x,\,t)dx)(v - u(t)) \ge 0 \qquad \forall\ v \in [0,\,\bar{u}], \end{cases} \tag{20}$$

where the last variational inequality couples ρ with p through the control constraints. In one-shot OD algorithms, all variables in (20) will be discretized and solved simultaneously, which would lead to a huge discretized system that is very expensive to store and solve. Hence, we will adopt the cheaper projection gradient descent method [24,39] to solve (20) by iteratively updating the control u with a forward solving of ρ and a backward solving of p, where the variational inequality is implemented as a projection mapping.

Given a time interval $[0,\,T]$, we define a uniform mesh $\{t_n = nh\}_{n=0}^N$ with a step size $h = T/N$. For the space domain $[1,\,b]$, we compute a non-uniform mesh $\{x_j = x(t_j)\}_{j=0}^N$ from solving the characteristic ODE

$$x'(t) = g(x), \qquad \text{with} \quad x(0) = 1.$$

In general, one may use some numerical methods for solving the above ODE. For example, one can use the second-order accurate modified Euler method ($j = 0,\,1,\,\cdots, N-1$)

$$x_{j+1} = x_j + hg\left[x_j + 0.5hg(x_j)\right], \qquad \text{with } x_0 = 1,$$

Notice the spatial mesh $\{x_j\}_{j=0}^N$ needs to be computed only once, since the growth rate function $g(x)$ does not explicitly depend on time t. Otherwise, the spatial mesh should be recomputed at each time step.

Let $\bar{x}(t;\,x,\,s)$ be a characteristic curve starting at $\bar{x}(s) = x$ and solving the characteristic ODE

$$\frac{d\bar{x}}{dt}(t;\,x,\,s) = g(\bar{x}).$$

For the given $g(x) = ax\ln(b/x)$, the solution of above equation is explicitly given as

$$\bar{x}(t;\,x,\,s) = b(x/b)^{e^{-a(t-s)}}.$$

Along such a characteristic curve $\bar{x}(t;\,x,\,s)$, the above transport PDE is then reduced to an ODE

$$\frac{d\rho}{dt}(\bar{x}(t;\,x,\,s),\,t) = -(g'(\bar{x}(t;\,x,\,s)) + u(t))\rho(\bar{x}(t;\,x,\,s),\,t),$$

which has the following explicit solution expression (starting at $\rho(\bar{x}(s;\,x,\,s),\,s) = \rho(x,\,s)$)

$$\begin{aligned} \rho(\bar{x}(t;\,x,\,s),\,t) &= \rho(x,\,s)\exp\left(-\int_s^t (g'(\bar{x}(\tau;\,x,\,s)) + u(\tau))d\tau\right) \\ &= \frac{\rho(x,\,s)g(x)}{g(\bar{x}(t;\,x,\,s))}e^{-\int_s^t u(\tau)d\tau}. \end{aligned} \tag{21}$$

Let $y_j^n = \rho(x_j,\,t_n)$ and $g_n = g(x_n)$. Now, we choose $x = x_j$, $s = t_n$ and $t = t_{n+1}$ in the above equation. It is noted that $\bar{x}(t_n;\,x_j,\,t_n) = x_j$ and $\bar{x}(t_{n+1};\,x_j,\,t_n) = x_{j+1}$. Therefore, $\rho(x,\,s) = \rho(x_j,\,t_n) = y_j^n$, $\rho(\bar{x}(t;\,x,\,s),\,t) = \rho(\bar{x}(t_{n+1};\,x_j,\,t_n),\,t_{n+1}) = \rho(x_{j+1},\,t_{n+1}) = y_{j+1}^{n+1}$, $g(\bar{x}(t_n;\,x,\,s)) = g(\bar{x}(t_n;\,x_j,\,t_n)) = g(x_j)$, and $g(\bar{x}(t_{n+1};\,x,\,s)) = g(\bar{x}(t_{n+1};\,x_j,\,t_n)) = g(x_{j+1})$. To evaluate the integral $\int_s^t (g'(\bar{x}(\tau;\,x,\,s))d\tau$, we make a change of variable $\tau \to \bar{x}$. The end points $\tau = t_n$ and $\tau = t_{n+1}$ correspond to the end points $\bar{x}(t_n;\,x,\,s) = \bar{x}(t_n;\,x_j,\,t_n) = x_j$ and $\bar{x}(t_{n+1};\,x,\,s) = \bar{x}(t_{n+1};\,x_j,\,t_n) = x_{j+1}$, respectively. Then we can obtain a discrete recursion formula along the characteristic curves:

$$\begin{aligned} y_{j+1}^{n+1} &= y_j^n \exp\left(-\int_{t_n}^{t_{n+1}} (g'(\bar{x}(\tau;\,x,\,s)) + u(\tau))d\tau\right) \\ &= y_j^n \exp\left(-\int_{x_j}^{x_{j+1}} \frac{g'(\bar{x}(\tau;\,x,\,s))}{g(\bar{x}(\tau;\,x,\,s))}d\bar{x}\right)\exp\left(-\int_{t_n}^{t_{n+1}} u(\tau)d\tau\right) \\ &= y_j^n \frac{g(x_j)}{g(x_{j+1})}\exp\left(-\int_{t_n}^{t_{n+1}} u(\tau)d\tau\right) \\ &\approx y_j^n \frac{g_j}{g_{j+1}}\exp(-0.5h(u(t_n) + u(t_{n+1}))), \end{aligned}$$

where the last approximation is based on the trapezoidal rule. This characteristic recursion formula starts with the initial condition $\rho(x,\,0) = \rho_0(x)$, that is $y_j^0 = \rho_0(x_j)$, $0 \le j \le N$. Finally, the non-local integral boundary condition $g(1)\rho(1,\,t) = \int_1^b \beta(x)\rho(x,\,t)dx$ at time $t = t_{n+1}$ can also be approximated with the composite trapezoidal rule

$$g(1)y_0^{n+1} = \sum_{j=0}^{N-1} \frac{s_j}{2}\left(\beta_j y_j^{n+1} + \beta_{j+1}y_{j+1}^{n+1}\right) + \int_{x_N}^b \beta(\bar{x})\rho(\bar{x},\,t_{n+1})d\bar{x},$$

where $s_j = (x_{j+1} - x_j)$ and $\beta_j = \beta(x_j)$. This can be further simplified into

$$(g(1) - \frac{s_0\beta_0}{2})y_0^{n+1} = \sum_{j=1}^{N-1} \frac{s_{j-1} + s_j}{2}\beta_j y_j^{n+1} + \frac{s_{N-1}}{2}\beta_N y_N^{n+1} + R_N(t_{n+1}).$$

The last truncated integration term (was neglected in the characteristic scheme developed in [6])

$$R_N(t_{n+1}) = \int_{x_N}^b \beta(\bar{x})\rho(\bar{x},\,t_{n+1})d\bar{x}$$

may be approximately zero if x_N is sufficiently close to b. In general, x_N is far less than b when T is small. In this case, we shall make use of (21) with $s = 0$ and $\bar{x}(t;\,x,\,s) = \bar{x}(t;\,x,\,0) = \bar{x} \ge x_N$ to compute $R_N(t_{n+1})$. It is easily seen that $x = b(\bar{x}/b)^{e^{at}}$ and

$$\rho(\bar{x},\,t) = \rho_0(b(\bar{x}/b)^{e^{at}})\frac{g(b(\bar{x}/b)^{e^{at}})}{g(\bar{x})}e^{-\int_0^t u(r)dr}$$

for $\bar{x} \in [x_N,\,b]$. Thus, we obtain an explicit expression (assume u is known)

$$\begin{aligned} R_N(t_{n+1}) &:= \int_{x_N}^b \beta(\bar{x})\rho(\bar{x},\,t_{n+1})d\bar{x} \\ &= \left(\int_{x_N}^b \beta(\bar{x})\rho_0\left(b(\bar{x}/b)^{e^{at_{n+1}}}\right)\frac{g(b(\bar{x}/b)^{e^{at_{n+1}}})}{g(\bar{x})}d\bar{x}\right)e^{-\int_0^{t_{n+1}} u(r)dr}, \end{aligned}$$

which can be then approximated by numerical quadrature (e.g. the *integral* function in MATLAB).

The characteristic scheme for p will be slightly different due to the extra term $\beta(x)p(1,\,t)$ with unknown boundary value $p(1,\,t)$. Let $\bar{x}(t;\,x,\,s)$ be the same characteristic curve starting at $\bar{x}(s) = x$. Along such a characteristic curve $\bar{x}(t;\,x,\,s)$, the above adjoint state PDE is then reduced to an ODE (treating $p(1,\,t)$ as a known function depends on t)

$$\frac{dp}{dt}(\bar{x}(t; x, s), t) = u(t)p(\bar{x}(t; x, s), t) - \beta(\bar{x}(t; x, s))p(1, t) + \theta w(\bar{x}(t; x, s)),$$

which has the following explicit solution expression (starting at s with $p(\bar{x}(s; x, s), s) = p(x, s)$)

$$\begin{aligned} &p(\bar{x}(t; x, s), t) \\ &= \exp\left(\int_s^t u(\tau)d\tau\right)\Big(p(x, s) - \int_s^t \exp(-\int_s^\tau u(r)dr)[\beta(\bar{x}(\tau; x, s))\\ &\quad p(1, \tau) - \theta w(\bar{x}(\tau; x, s))]d\tau\Big). \end{aligned} \tag{22}$$

By letting $s = t_n$ and $t = t_{n+1}$, we can obtain the recursion formula (used trapezoidal rule twice)

$$\begin{aligned} p_{j+1}^{n+1} &= \exp\left(\int_{t_n}^{t_{n+1}} u(\tau)d\tau\right)\Big(p_j^n - \int_{t_n}^{t_{n+1}} \exp\Big(-\int_{t_n}^{\tau} u(r)dr\\ &\quad \Big)[\beta(\bar{x}(\tau; x, t_n))p(1, \tau) - \theta w(\bar{x}(\tau; x, t_n))]d\tau\Big)\\ &\approx \exp\left(\int_{t_n}^{t_{n+1}} u(\tau)d\tau\right)\Big(p_j^n - 0.5h\Big([\beta_j p(1, t_n) - \theta w_j] + \exp\\ &\quad \Big(-\int_{t_n}^{t_{n+1}} u(r)dr\Big)[\beta_{j+1}p(1, t_{n+1}) - \theta w_{j+1}]\Big)\Big)\\ &= (p_j^n - 0.5h[\beta_j p_1^n - \theta w_j])\exp\left(\int_{t_n}^{t_{n+1}} u(\tau)d\tau\right)\\ &\quad - 0.5h[\beta_{j+1}p_1^{n+1} - \theta w_{j+1}]\\ &\approx (p_j^n - 0.5h[\beta_j p_1^n - \theta w_j])\exp(0.5h(u(t_n) + u(t_{n+1})))\\ &\quad - 0.5h[\beta_{j+1}p_1^{n+1} - \theta w_{j+1}] \end{aligned}$$

which can be marched backward according to (with $j \geq 2$ and the end condition $p_j^N = -w_j$)

$$\begin{aligned} &p_1^n\\ &= [(p_2^{n+1} + 0.5h[\beta_2 p_1^{n+1} - \theta w_2])\exp(-0.5h(u(t_n) + u(t_{n+1})))\\ &\quad - 0.5h\theta w_1]/(1 - 0.5h\beta_1), p_j^n\\ &= (p_{j+1}^{n+1} + 0.5h[\beta_{j+1}p_1^{n+1} - \theta w_{j+1}])\exp(-0.5h(u(t_n) + u(t_{n+1})))\\ &\quad + 0.5h[\beta_j p_1^n - \theta w_j]. \end{aligned}$$

Using characteristic curves, the effective boundary condition at $x = x_N$ is

$$\begin{aligned} p_N^n &= p(\chi(2T - t_n), T)e^{-\int_{t_n}^T u(s)ds}\\ &\quad + \int_0^{T-t_n} [\beta(\chi(T + s))p(1, t + s) - \theta w(\chi(T + s))]e^{-\int_{t_n}^{t_n+s} u(r)dr}ds\\ &= -w(\chi(h(2N - n)))e^{-U_N + U_n}\\ &\quad + \sum_{k=0}^{N-n} Q_{N-n,k}[\beta(\chi(h(N + k)))p_1^{n+k} - \theta w(\chi(h(N + k)))]e^{-U_{n+k}+U_n}, \end{aligned}$$

where $U_k = \int_0^{t_k} u(s)ds$ and $Q_{n,k} = h(1 - \delta_{k,n}/2 - \delta_{k,0}/2)$ denotes the quadrature weights of trapezoidal rule.

Based on the above described characteristic schemes, a projection gradient descent (PGD) algorithm for solving (20) is to construct a fixed point iteration for iteratively updating $u(t)$ along the gradient descent direction. In each iteration it requires time-marching along the characteristic curves to solve the decoupled independent PDE for ρ and p, respectively. The complete PGD algorithm is summarized in the following Algorithm 1, where the step size α_k is selected by the standard Armijo rule [30]. Based on our following numerical simulations, we have observed a linear convergence rate of the proposed PGD algorithm. However, we remark that a rigorous convergence analysis of the above PGD algorithm is beyond the scope of this paper, which will be left as our future work.

4. Numerical results

In this section, we will provide several numerical examples to validate our obtained theoretical conclusions and also demonstrate the approximate accuracy of our proposed characteristic schemes. All simulations are implemented using MATLAB 2017b on a Dell Precision Workstation with Intel(R) Core(TM) i7-7700K CPU@4.2GHz and 32GB RAM. The CPU time (in seconds) is estimated by timing functions `tic/toc`.

Following [23], we will use the Gompertz growth rate $g(x) = ax\ln(b/x)$ and birth rate $\beta(x) = \mu x^\alpha$. The stopping tolerance of the PGD algorithm is $tol = 10^{-8}$ and $k_{\max} = 5000$. Throughout this section, we set $w(x) = x$ and define the total metastatic masses at time t as

$$M(t) = \int_1^b x\rho(x, t)dx. \tag{23}$$

Example 1. In our first toy example, we choose the following academic parameters

$$a = 1/day,\ b = e,\ \mu = 1/day,\ \alpha = 1,\ \gamma = 0.1,\ \bar{u} = 2.$$

We also set $T = 10$ and $N = 100$. Given the same initial colony $\rho(x, 0) = e^{-x}$, we compare the metastatic masses $M(t)$, the final colony $\rho(x, T)$ and the optimal controls $u(t)$ for the J_0 and J_1 models, respectively, in Fig. 1. The treatment outcomes of the J_0 and J_1 model are very different, where the J_1 model gives much smaller total metastatic masses $M(t)$ and the corresponding computed optimal control matches closely with the exact bang-bang control. Notice the computed optimal control in the J_0 model is uniform and not of bang-bang type, which is not surprising due to its non-uniqueness. The V-shape trajectory of the metastatic mass $M(t)$ in J_1 model implies the tumor will continue to grow at the normal growth rates once the drug dosage becomes zero. We remark that, in the right bottom panel of Fig. 1, the computed optimal control for J_1 model (blue dashed line) will become more vertical across the exact switching point and eventually approach the exact optimal control (green dotted line) if we choose a smaller tolerance and allow more iterations. Similar behavior is also observed among the following related figures.

In Table 1, we report the computed objective functional (J_0^h and J_1^h) and the corresponding total drug dosage $U(T)$. In the last row of Table 1, we give the benchmarking exact optimal objective functional and total drug dosage, which are computed via solving the integral equation derived in Section 2 with a much smaller mesh size. Due to the approximation errors in numerical quadrature used for computing the provided exact reference values, we should understand the 'exact' values are accurate only up to about 10^{-3} accuracy. It is also worthwhile to notice that the J_1 model takes more iterations to reach convergence, and the convergence rates of J_0 model seem to be mesh-independent, which is expected according to [24].

Figs. 2 and 3 provide the surface plot (with $N = 100$) of the computed optimal state ρ and adjoint state p in the J_0 and J_1 model, respectively, where we clearly observe $\rho(x, t) > 0$ and $p(x, t) < 0$, as shown in Theorem 3.1. The incompatible initial-boundary condition at $x = 1$ and $t = 0$ leads to the discontinuity of ρ at the corner (1,0). Comparing J_0 and J_1 models, the dynamics of the optimal state ρ are similar, but it is interesting to note that the adjoint state p demonstrates very different dynamics. This may partially explain why the required numbers of iterations are significantly different for the J_0 and J_1 model.

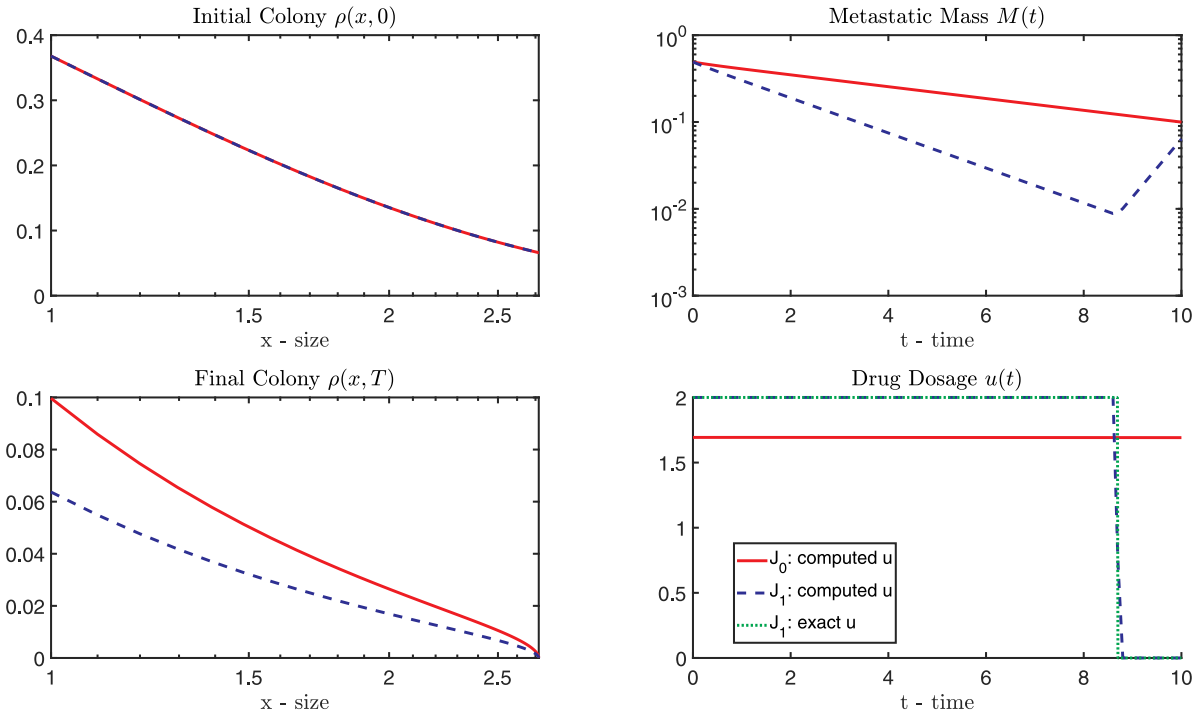

Fig. 1. Ex 1: Results for the J_0 model (red solid line) and the J_1 (blue dashed line) model: initial metastatic density (left top), final metastatic density (left bottom), metastatic mass dynamics (right top), and the computed optimal drug dosage treatment plans (right bottom). The green dotted line in the right bottom panel illustrates the exact optimal control for J_1 model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Now, we fix the optimal total drug dosage to be $U(T) = 17.37$ (with $T = 10$) and compare the treatment outcomes of the bang-bang control

$$u(t) = \begin{cases} 2, & t \in [0, \tau^*], \\ 0, & t \in (\tau^*, 10], \end{cases}$$

with $\tau^* = U(T)/\bar{u} = 17.37/2 = 8.685$ and the uniform control

$$u(t) = U(T)/T = 1.737, \quad t \in [0, 10].$$

It is observed from Fig. 4 that the final metastatic masses of two treatment plans are the same, which agrees with our theoretical result that the final metastatic masses only depends on the total drug dosage. However, the metastatic masses for the (optimal) bang-bang treatment over $(0, T)$ are always smaller than that for the simple uniform dosage treatment. The biological interpretation of this result is that the maximum drug dosage should be administrated from the beginning of the treatment period, otherwise the metastatic mass may become too high for the patient to survive until the treatment ends (see also Example 2).

Example 2. We choose the following parameters in a preclinical scenario [15]:

$a = 0.08/day$, $b = 6 \times 10^8$, $\mu = 10^{-5}/day$, $\alpha = 2/3$, $\gamma = 0.1$, $\bar{u} = 2$.

We also set $T = 15$, $N = 150$ and $\rho(x, 0) = e^{-x}$ in the simulation. The comparison between J_0 and J_1 models is illustrated in Fig. 5. We notice the J_0 model shows the undesirable growth in the total metastatic masses during the early stage of the treatment period, which also illustrates the importance of choosing a realistic treatment model. We remark that the huge spatial domain $[1, b = 6 \times 10^8]$ will lead to numerical difficulty for any standard finite difference and finite element discretization schemes that introduce mesh-based discretization errors in space. However, the characteristic scheme adopted in our numerical simulation does not suffer from possible large discretization errors in space, because it uses exact formulas along the characteristic curves and the only approximation errors are introduced in the quadrature rule. It is also worthwhile to mention again the shape of computed optimal control used by the PGD algorithm only approximates the exact bang-bang control, whose sharpness in switching behavior is controlled by the chosen stopping tolerance *tol*. If a smaller tolerance were used, the computed optimal control would better approximate the bang-bang shape of the exact control, but the drawback is that more iterations and thus more computation time would be needed. **Example 3.** We choose the following parameters in a clinical scenario [9]:

$a = 0.0084/day$, $b = 6.25 \times 10^8$, $\mu = 0.001/day$, $\alpha = 2/3$, $\gamma = 0.1$, $\bar{u} = 0.5$.

Table 1
Comparison of the J_0 model and the J_1 model (with $T = 10$).

J_0 model					J_1 model				
N	Iter	J_0^h	$U(T)$	CPU	N	Iter	J_1^h	$U(T)$	CPU
100	9	1.792320	16.926107	0.07	100	1313	2.853595	17.373909	5.79
200	9	1.792249	16.923214	0.14	200	1630	2.853041	17.369267	18.98
300	9	1.792236	16.922678	0.23	300	2075	2.852935	17.370359	47.14
400	9	1.792231	16.922491	0.35	400	1765	2.852903	17.370896	66.91
500	9	1.792229	16.922404	0.56	500	1910	2.852885	17.369670	109.36
Exact		1.792225	16.922400	6.69	Exact		2.852847	17.370000	6.26

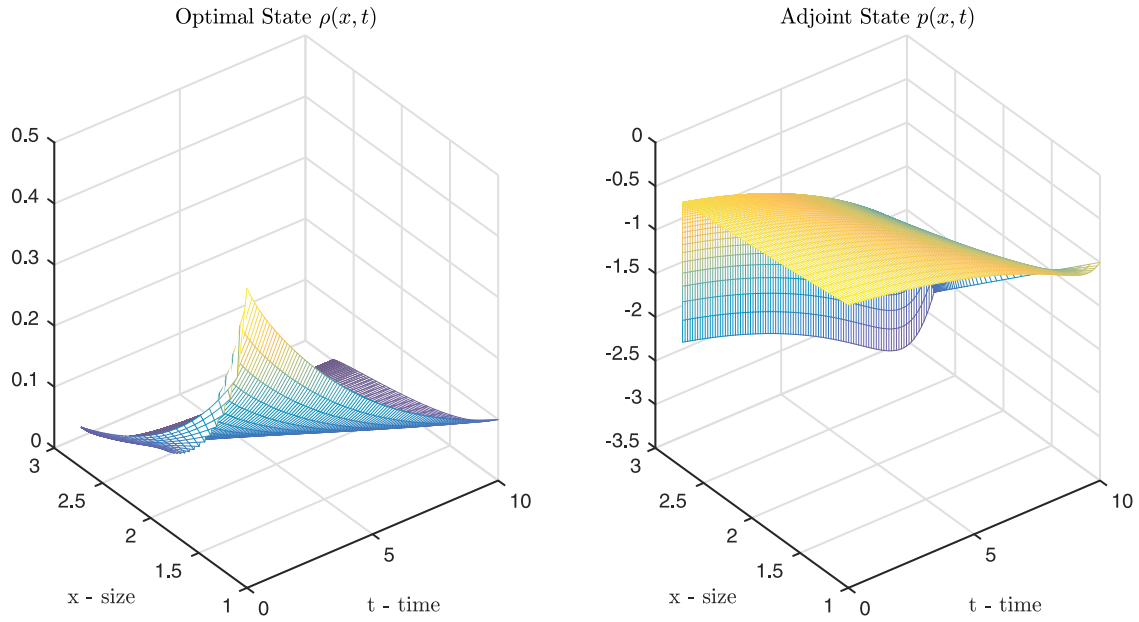

Fig. 2. Ex 1: Surface plots of the optimal state ρ and the adjoint state p in J_0 model.

We first set $T = 30$, $N = 300$ and $\rho(x, 0) = e^{-x}$ in the simulation. It is observed from Fig. 6 that the J_1 model has much better treatment outcomes and the optimal switching time for the bang-bang control is $\tau^* \approx 16.7$. Different from Example 2, a much smaller a gives much slower growth rate $g(x)$, which requires a much larger time T to generate a spatial mesh adequately resolving the domain $[1, b]$.

To illustrate the effects of different initial conditions, we also test the case with a constant initial colony $\rho_0(x) = 0.5$ and a larger time $T = 90$ (set $N = 900$), as shown in Fig. 7. We observe from Figs. 6 and 7 that the distribution of initial colony does not affect the treatment outcomes significantly. For the given two initial conditions, the final colony will take similar shape of a decreasing function; see left bottom panels of Figs. 6 and 7. We also note that the total drug dosage of optimal control for J_1 model is slightly more than that in J_0 model, but the treatment outcome of J_1 model is significantly better than the J_0 model in the sense that the metastatic mass is much smaller over the treatment period.

5. Conclusions

In this paper, we have proposed a unified optimal treatment model based on a validated metastatic cancer growth PDE model. The first-order necessary optimality condition is derived for characterizing the optimal control. After that, a projection gradient descent algorithm built upon the characteristic scheme is developed for its efficient numerical solutions. Numerical results confirm our theoretical conclusions regarding the non-uniqueness and uniqueness of optimal control in the J_0 and J_1 model, respectively.

Based on our analytic and computational studies, we have learned how the side effects of drug affect the optimal strategy in cancer

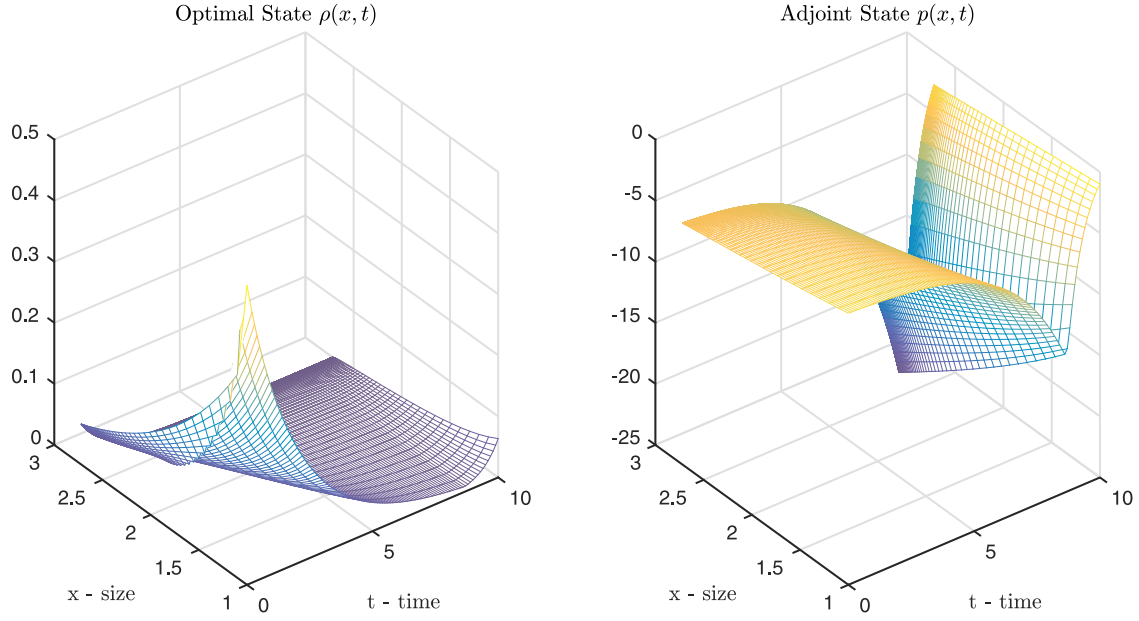

Fig. 3. Ex 1: Surface plots of the optimal state ρ and the adjoint state p in J_1 model.

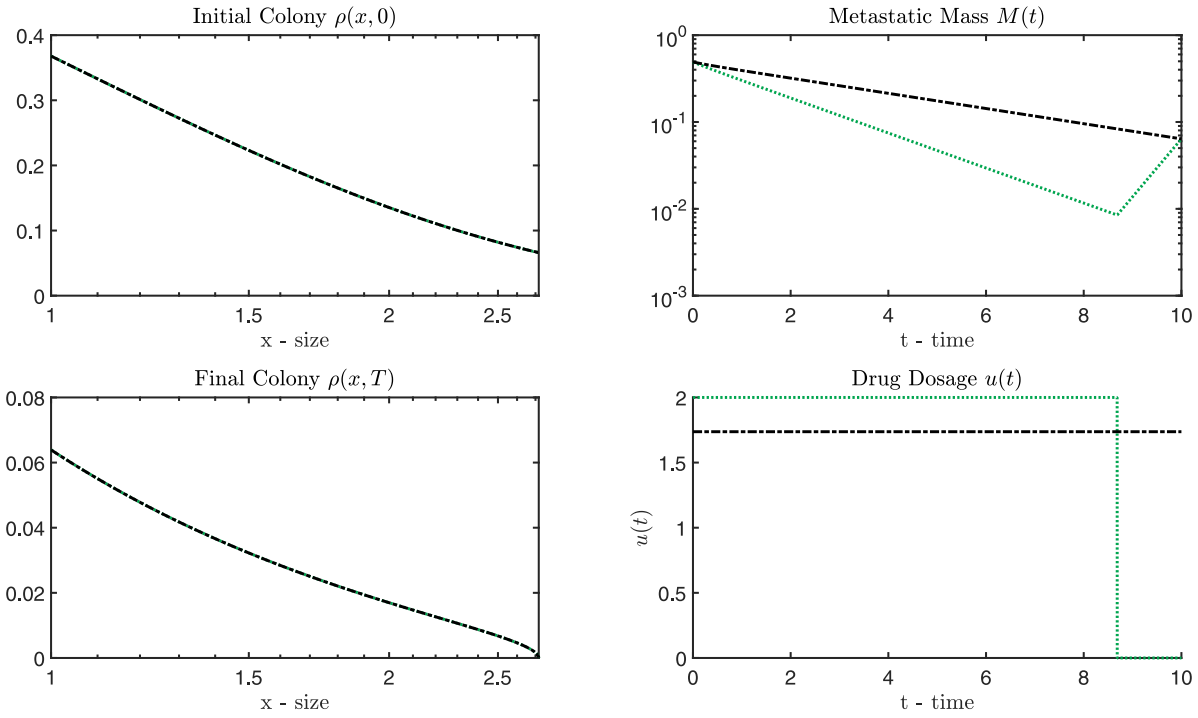

Fig. 4. Ex 1: Results for bang-bang treatment (green dotted line) and uniform treatment (black dash-dotted line) with the same optimal total drug dosage $U(T) = 17.37$: initial metastatic density (left top), final metastatic density (left bottom), metastatic mass dynamics (right top), and the drug dosage treatment plans (right bottom). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

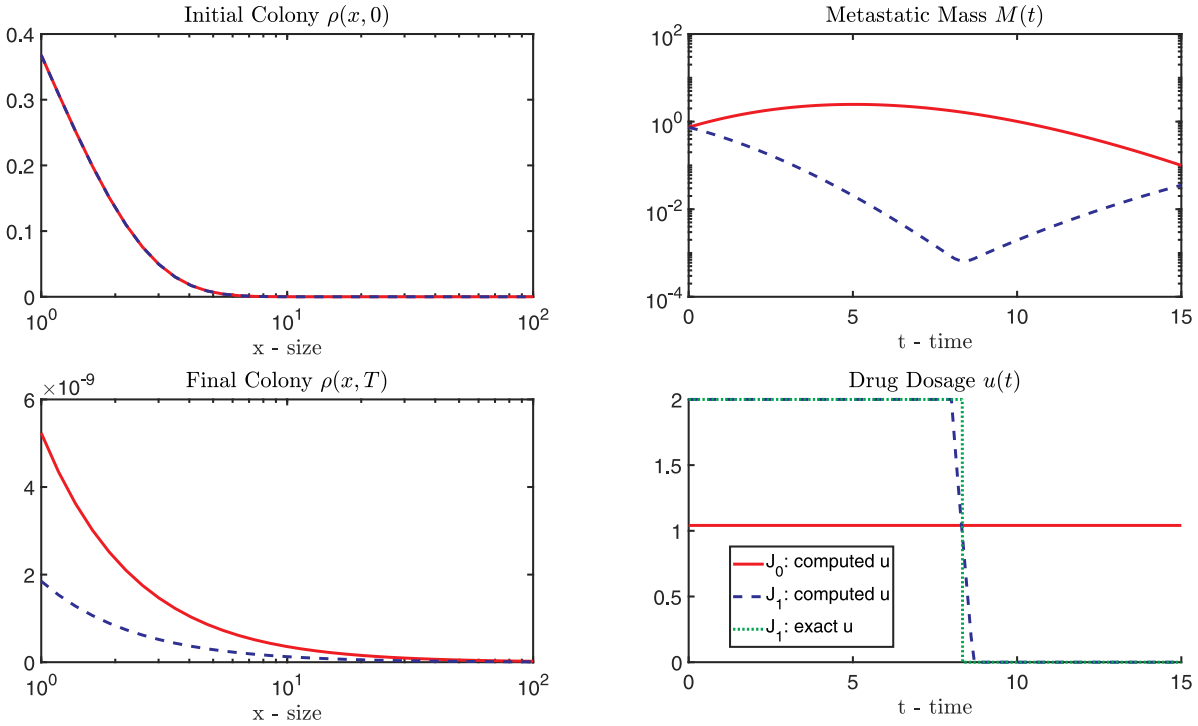

Fig. 5. Ex 2: Results for the J_0 model (red solid line) and the J_1 (blue dashed line) model: initial metastatic density (left top), final metastatic density (left bottom), metastatic mass dynamics (right top), and the computed optimal drug dosage treatment plans (right bottom). The green dotted line in the right bottom panel illustrates the exact optimal control for J_1 model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

treatment. If the drug is too poisonous, one should always avoid using it. On the other hand, if the side effect of the drug is negligible, we should use all possible drug dosages to optimize the outcome. However, in the case when the side effect of the drug is neither too strong nor too weak, we need to balance the tumor treatment with drug abuse. So, we need to consider different types of objective functionals. If the tumor sizes during the treatment are ignored and we only want to reduce the final tumor size, it is proved in this paper that the optimal total dosage

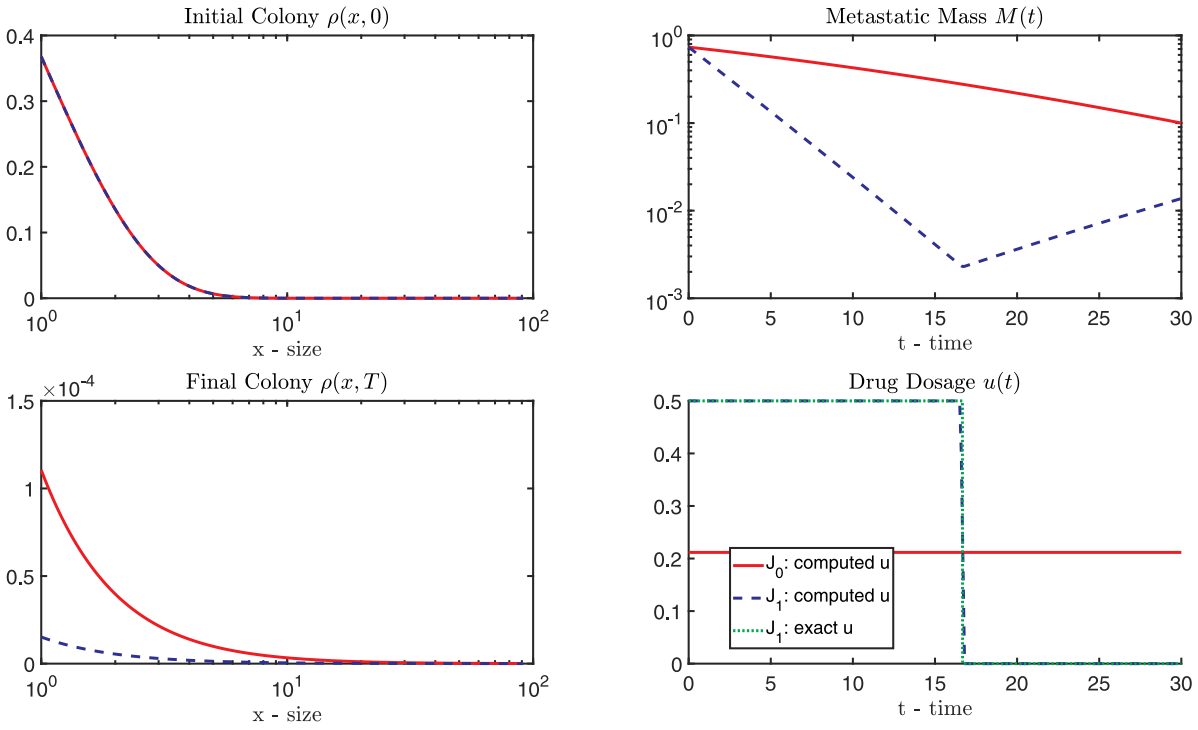

Fig. 6. Ex 3 ($\rho_0(x) = e^{-x}$, $T = 30$): Results for the J_0 model (red solid line) and the J_1 (blue dashed line) model: initial metastatic density (left top), final metastatic density (left bottom), metastatic mass dynamics (right top), and the computed optimal drug dosage treatment plans (right bottom). The green dotted line in the right bottom panel illustrates the exact optimal control for J_1 model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

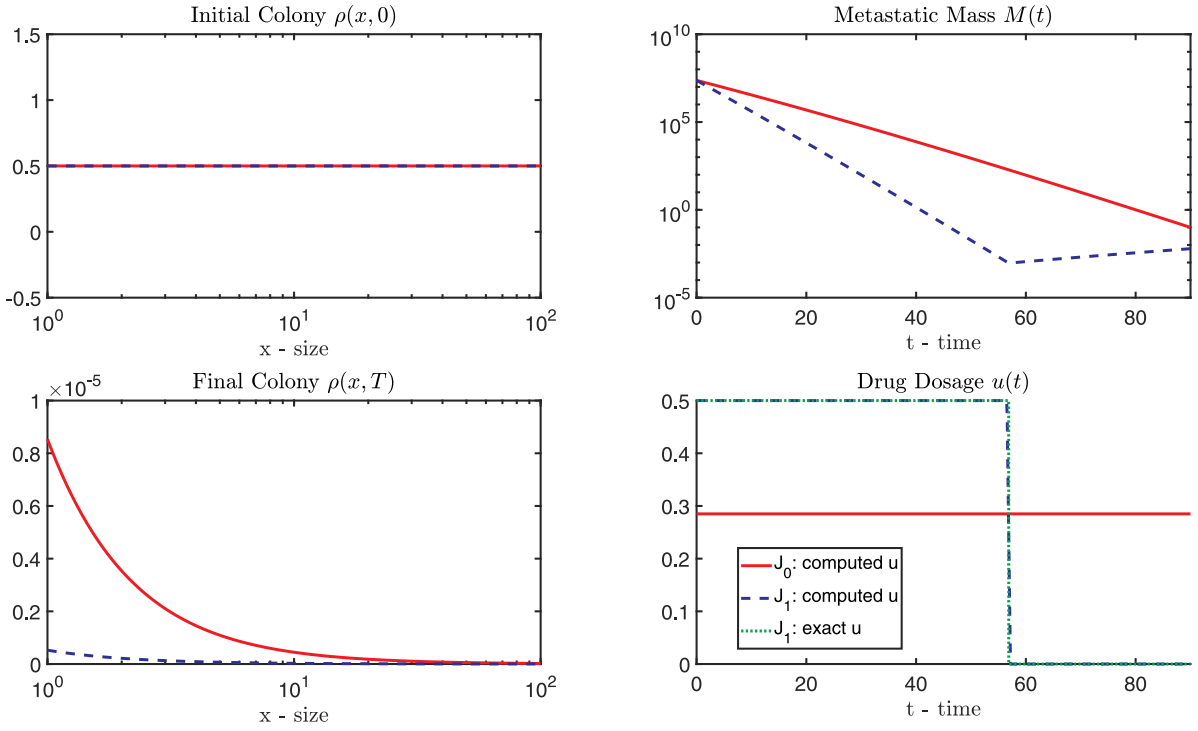

Fig. 7. Ex 3 ($\rho_0(x) = 0.5$, $T = 90$): Results for the J_0 model (red solid line) and the J_1 (blue dashed line) model: initial metastatic density (left top), final metastatic density (left bottom), metastatic mass dynamics (right top), and the computed optimal drug dosage treatment plans (right bottom). The green dotted line in the right bottom panel illustrates the exact optimal control for J_1 model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

procedure u^*=PGD($\theta, \gamma, N, h, tol, k_{\max}$)

choose an initial guess $u_n^{(k)} = 0,\ 0 \le n \le N$;

for k=0:$k_{\max}$ **do**

solve for $y^{(k)} = \{y_j^n : 0 \le n, j \le N\}$ using the characteristic scheme (forward with $y_j^0 = \rho_0(x_j)$)

$$y_{j+1}^{n+1} = (g_j/g_{j+1}) y_j^n \exp\left(-0.5h(u_n^{(k)} + u_{n+1}^{(k)})\right), \quad 0 \le n, j \le N-1$$

$$y_0^{n+1} = \frac{1}{(g_0 - \frac{s_0\beta_0}{2})}\left(\sum_{j=1}^{N-1} \frac{s_{j-1}+s_j}{2}\beta_j y_j^{n+1} + \frac{s_{N-1}}{2}\beta_N y_N^{n+1} + R_N(t_{n+1})\right), \quad 0 \le n \le N-1.$$

solve for $p^{(k)} = \{p_j^n : 0 \le n, j \le N\}$ using the scheme (backward with $p_j^N = -\theta w(x_j)$)

$$p_1^n = \left[(p_2^{n+1} + 0.5h[\beta_2 p_1^{n+1} - \theta w_2]) \exp\left(-0.5h(u_n^{(k)} + u_{n+1}^{(k)})\right) - 0.5h\theta w_1\right]/(1 - 0.5h\beta_1),$$

$$p_j^n = (p_{j+1}^{n+1} + 0.5h[\beta_{j+1} p_1^{n+1} - \theta w_{j+1}]) \exp\left(-0.5h(u_n^{(k)} + u_{n+1}^{(k)})\right) + 0.5h[\beta_j p_1^n - \theta w_j],$$

$$p_N^n = -w(\chi(h(2N-n)))e^{-U_N+U_n} + \sum_{j=0}^{N-n} Q_{N-n,j}[\beta(\chi(h(N+j)))p_1^{n+j} - \theta w(\chi(h(N+j)))]e^{-U_{n+j}+U_n}.$$

compute $z^{(k)}$ by the trapezoidal rule

$$z_n^{(k)} = \int_1^b \rho^{(k)}(x, t_n) p^{(k)}(x, t_n) dx \approx \sum_{j=0}^{N} Q_{N,j}[y^{(k)}]_j^n [p^{(k)}]_j^n$$

update $u^{(k)}$ to get $u^{(k+1)}$ along the negative gradient direction (based on optimality condition):

$$u^{(k+1)} = \max\{0, \min\{\bar{u}, u^{(k)} - \alpha_k(\gamma + z^{(k)})\}\},$$

where α_k is the step size computed with the Armijo rule or just taken as $\alpha_k = 1$.

if $\|u^{(k+1)} - u^{(k)}\| \le tol$ **then**

return $u^* = u^{(k+1)}$; //return the optimal control

end if

end for

end procedure

Algorithm 1. A projection gradient descent (PGD) algorithm: **Initialization**: θ, γ, N, $h = T/N$, tol, $k_{\max}$ **Output**: The optimal control vector $u_n^*,\ 0 \le n \le N$.

is unique but the optimal drug therapies are not unique. In other words, it does not matter when the drug is applied, as long as by the end of cancer treatment, the total drug dosage is maintained at a unique optimal value. If we take into consideration the tumor sizes along the whole treatment time, our analysis demonstrates that the optimal drug therapy becomes unique and the unique optimal solution is of bang-bang type; namely, one should apply all possible drug dosages at the beginning of treatment. The J_1 model seems to be more realistic than J_0 model because the patient may not be able to survive from the high concentration of tumor cells before the drug treatment completes. The bang-bang optimal control for J_1 model provides mathematical justification on the importance and effectiveness of early diagnosis and early treatment for cancer disease.

Our model can be improved in many different directions. For example, we may add nonlinear drug resistance and immunotherapy effects. We can enforce state constraints to keep a safe level of drug concentration within the body. For clinical purposes, it is also interesting to consider the time-optimal control problem, where the objective is to minimize the total treatment time for the given treatment goals (such as a fixed minimal amount of tumor mass reduction). We observe from Table 1 that the CPU time of solving the J_1 model is much higher than that of the J_0 model. It seems that the computational cost of the J_0 model is roughly of order $O(N)$, while that of the J_1 model is about $O(N^2)$, where N is the number of mesh points. Regarding the efficiency of the developed numerical algorithms, it is possible to further speed up our current optimization algorithms by combining multilevel optimization techniques [20,34] and Nesterov's accelerated gradient descent methods [7,35].

Finally, we remark that the size-structured metastatic density $\rho(x, t)$ itself may not be of obvious biological interests, hence it is also desirable to calculate the biological quantities directly without first computing $\rho(x, t)$. For instance, in [15], the authors reformulated the PDE into a Volterra integral equation of the second kind of convolution type regarding the biologically observable quantity $M(t)$ only, which allows more efficient numerical algorithms, such as FFT. Hence, it is very interesting to build optimal control models upon such integral equations, where the control term $u(t)$ acts on $M(t)$ in a collective manner.

Acknowledgments

The authors want to thank the editor and two anonymous referees for their valuable comments and detailed suggestions that have greatly contributed to improving the quality of this paper. The first author would also like to thank Prof. Michael Hinze and Prof. Suzanne Lenhart for some helpful discussions, which were made possible by the travel support from NSF DMS-1743826 Award, namely the NSF-CBMS Conference: Computational Methods in Optimal Control, held at Jackson State University during July 23–27, 2018.

Appendix A. revisit the technique of change of variables

In [11], the authors suggested to handle the large spatial domain by introducing a change of variable

$$y = \ln\frac{b}{x} \qquad i.e., \qquad x = be^{-y}$$

and then a change of function

$v(y, t) = x\rho(x, t) = be^{-y}\rho(be^{-y}, t),$

which lead to the following transformed model (with $f(y) = -ay$ and $\bar{b} = \ln b$)

$$v_t(y, t) + (f(y)v(y, t))_y = 0, \qquad 0 \le y \le \bar{b}, 0 < t < T,$$
$$f(\bar{b})v(\bar{b}, t) = \int_1^b \beta(be^{-y})v(y, t)dy, \qquad t > 0,$$
$$v(y, 0) = ab^{-y}\rho_0(be^{-y}), \qquad 1 \le y \le \bar{b}.$$

In the finite difference scheme proposed in [11], a uniform spaced mesh in y is used, which creates an exponentially nonuniform mesh in x that concentrates most of the grid points near the left boundary $x = 1$. It seems reasonable to accurately resolve the left boundary condition due to its importance, but their numerical results indicate such a nonuniform mesh in x has difficulty or lower accuracy in approximating the numerical solutions near the right boundary $x = b$. In view of the success of the characteristic scheme, which in fact stretches most of the grid points to the right boundary $x = b$, we believe a nonuniform mesh in y is needed so that the numerical solutions at large x can be more accurately resolved. Notice that the nonlocal left boundary condition will not be accurate anyway if the numerical solution is inaccurate near the right boundary.

Based on the above transformed model, we can also develop its characteristic scheme or discretize the y variable in a nonuniform way so that it produces better grid mesh in the x variable. For example, we may choose the following nonuniform mesh points

$y_j = \bar{b}e^{-at_j}, \qquad t_j = jh, h = T/N,$

which corresponds to the following nonuniform mesh in x (clustering towards b)

$x_j = b\exp\left(-\bar{b}e^{-at_j}\right) = b^{1-e^{-at_j}} = \chi(t_j), \qquad t_j = jh, h = T/N.$

Due to the large mesh step size in y_j or x_j near b, any standard finite difference scheme will provide very lower approximation accuracy, while the nonuniform mesh based characteristic scheme leads to more accurate approximations since it is exact along the characteristic curves and the approximations to the integral terms are of second-order accuracy. Hence, characteristic schemes are more suitable for optimal control models.

Appendix B. optimal control with Tikhonov regularization

In general, the bang-bang optimal control problem may be ill-posed, in the sense that its optimal solution does not depend continuously on the data, which hence often needs careful numerical treatments to get numerically stable approximations. For practical numerical approximations, it is a standard technique (see e.g. [43,51,52] and the references therein) to add a Tikhonov regularization term $\frac{\lambda}{2}\int_0^T u^2(t)dt$ to the original objective functional to get a regularized problem of minimizing

$$J_{\theta,\lambda}(\rho, u) := J_\theta(\rho, u) + \frac{\lambda}{2}\int_0^T u^2(t)dt$$

where $\lambda > 0$ denotes the Tikhonov regularization parameter. For the treatment purpose, we are mainly interested in the limit problem with $\lambda \to 0$, which can be implemented by taking $\lambda = h^2$ for a discretization with a mesh step size h. Such an extra regularization term will turn the last 'bang-bang' type variational inequality into a semi-smooth projection formula. Specifically, we arrive at the following optimality condition

$$0 \le L_u(\rho, u, p)(v - u) = \lambda\int_0^T u(t)(v(t) - u(t))dt + \gamma\int_0^T (v(t) - u(t))dt + \int_0^T\int_1^b (v(t) - u(t))\rho p dx dt$$
$$= \int_0^T \left(\lambda u(t) + \gamma + \int_1^b \rho p dx\right)(v(t) - u(t))dt,$$

which further leads to the following projection formula for characterizing the optimal control

$$u_\lambda(t) = \mathcal{P}_{U_{ad}}\left(\frac{-1}{\lambda}(z(t) + \gamma)\right) := \begin{cases} 0 & \text{if } (\lambda u(t) + \gamma) + z(t)) > 0, \\ \frac{-1}{\lambda}(z(t) + \gamma), & \text{if } (\lambda u(t) + \gamma) + z(t)) = 0, \\ \bar{u} & \text{if } (\lambda u(t) + \gamma) + z(t)) < 0. \end{cases}$$

For simplicity, such a Tikhonov regularization term was not implemented in our numerical simulations, but our proposed PGD algorithm can easily combine such a Tikhonov regularization term if needed.

Appendix C. A discretize-then-optimize (DO) algorithm

For completeness, we also present a discretize-then-optimize (DO) algorithm for solving our optimal control problem. Let $w_j = w(x_j)$. The discretized objective functional with the composite trapezoidal rule reads

$$J_\theta^h(\boldsymbol{y}, \boldsymbol{u}) = [\boldsymbol{c}_1 \quad \gamma\boldsymbol{c}_2]^\intercal[\boldsymbol{y}^N\boldsymbol{u}] + \theta\boldsymbol{c}_3^\intercal\boldsymbol{y} = [\theta\boldsymbol{c}_3 + \hat{\boldsymbol{c}_1} \quad \gamma\boldsymbol{c}_2]^\intercal[\boldsymbol{y}\boldsymbol{u}] = \boldsymbol{c}^\intercal[\boldsymbol{y}\boldsymbol{u}]$$

where $\boldsymbol{y} = vec(\{y_j^n : j = 0, \cdots, N, n = 0, \cdots, N\})$, $\boldsymbol{y}^N = [y_0^N, y_1^N, \cdots, y_N^N]^\intercal$ is the tumor colony size distribution vector at the final time $t = t_N = T$, $\boldsymbol{u} = [u_0, u_1, \cdots, u_N]^\intercal$ is the drug dosage decision vector, and the constant quadrature weights vectors

$$\boldsymbol{c}_1 = [\frac{s_0}{2}w_0, \frac{s_0 + s_1}{2}w_1, \cdots, \frac{s_{j-1} + s_j}{2}w_j, \cdots, \frac{s_{N-2} + s_{N-1}}{2}w_{N-1}, \frac{s_{N-1}}{2}w_N], \quad \boldsymbol{c}_2 = [\frac{h}{2}, h, \cdots, h, \frac{h}{2}],$$

$\boldsymbol{c}_3 = \boldsymbol{c}_1 \otimes \boldsymbol{c}_2$ corresponds to the double integral term, and $\hat{\boldsymbol{c}_1} = [\boldsymbol{0}_{N(N+1)} \quad \boldsymbol{c}_1]$ denotes the obvious zero extension of $\boldsymbol{c}_1$ to match with the length of $\boldsymbol{c}_3$. Here $\otimes$ denotes the standard matrix Kronecker product.

Putting all pieces together, we obtain the following discretized nonlinear optimization problem

$$\begin{aligned}
\min_{0\le \boldsymbol{u}\le \bar{u}} \quad & J_\theta^h(\boldsymbol{y},\boldsymbol{u}) = \boldsymbol{c}^\intercal \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{u} \end{bmatrix} \\
\text{s.t.} \quad & \\
& y_j^0 = \rho_0(x_j), \quad 0 \le j \le N \\
& g_{j+1} y_{j+1}^{n+1} - g_j y_j^n \exp(-0.5h(u_n + u_{n+1})) = 0, \quad 0 \le n, j \le N-1 \\
& (g_0 - \frac{s_0\beta_0}{2}) y_0^{n+1} - \sum_{j=1}^{N-1} \frac{s_{j-1}+s_j}{2}\beta_j y_j^{n+1} - \frac{s_{N-1}}{2}\beta_N y_N^{n+1} = R_N(t_{n+1}), \quad 0 \le n \le N-1.
\end{aligned} \tag{24}$$

For better computational efficiency in numerical optimization solvers, we may further linearize the nonlinear constraints in the optimization problem (24) by introducing a change of variable $z_j^n = \ln(y_j^n)$, which will then lead to a nonlinear objective function. More specifically, we get a 'linearized' problem

$$\begin{aligned}
\min_{0\le \boldsymbol{u}\le \bar{u}} \quad & J_\theta^h(\boldsymbol{z},\boldsymbol{u}) = \boldsymbol{c}^\intercal \begin{bmatrix} \exp(\boldsymbol{z}) \\ \boldsymbol{u} \end{bmatrix} \\
\text{s.t.} \quad & \\
& z_j^0 = \ln(\rho_0(x_j)), \qquad 0 \le j \le N \\
& z_{j+1}^{n+1} - z_j^n + 0.5h(u_n + u_{n+1}) = \ln\left(\frac{g_j}{g_{j+1}}\right), \qquad 0 \le n, j \le N-1 \\
& (g_0 - \frac{s_0\beta_0}{2}) \exp(z_0^{n+1}) - \sum_{j=1}^{N-1} \frac{s_{j-1}+s_j}{2}\beta_j \exp(z_j^{n+1}) - \frac{s_{N-1}}{2}\beta_N \exp(z_N^{n+1}) = R_N(t_{n+1}),
\end{aligned}$$

where we need to assume $\rho_0(x_j) > 0$. The above linearized formulation is expected to be computationally more favorable to be used in the standard nonlinear optimization solvers, since it has a smaller number of nonlinear constraints and hence requires fewer computational costs in calculating the constraints' Jacobian matrix to be used in optimization solvers (e.g., the SQP algorithm in SNOPT [13] and the interior point algorithm in IPOPT [53]). The above discretized optimization model treats the states z_j^n and control u_n at all time steps as decision variables (of $O(N^2)$), which leads to a large-scale sparse nonlinear optimization that can not be easily handled by any black-box optimization solvers. It is often necessary to develop a specific optimization solver by making use of the sparsity structure of the Jacobian matrix and some efficient preconditioning techniques [18,36,38].

References

[1] P.M. Altrock, L.L. Liu, F. Michor, The mathematics of cancer: integrating quantitative models, Nat. Rev. Cancer 15 (2015) 730–745.
[2] O. Angulo, J. López-Marcos, Numerical schemes for size-structured population equations, Math. Biosci. 157 (1999) 169–188.
[3] S. Aniţa, Analysis and Control of Age-Dependent Population Dynamics, Springer, Netherlands, 2000.
[4] S. Aniţa, V. Arnăutu, V. Capasso, An Introduction to Optimal Control Problems in Life Sciences and Economics: From Mathematical Models to Numerical Simulation with MATLAB®, Modeling and Simulation in Science, Engineering And Technology, Birkhäuser, Boston, 2011.
[5] C.T. Baker, A perspective on the numerical treatment of Volterra equations, J. Comput. Appl. Math. 125 (2000) 217–249.
[6] D. Barbolosi, A. Benabdallah, F. Hubert, F. Verga, Mathematical and numerical analysis for a model of growing metastatic tumors, Math. Biosci. 218 (2009) 1–14.
[7] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009) 183–202.
[8] S. Benzekry, Mathematical and numerical analysis of a model for anti-angiogenic therapy in metastatic cancers, ESAIM 46 (2011) 207–237.
[9] S. Benzekry, P. Hahnfeldt, Maximum tolerated dose versus metronomic scheduling in the treatment of metastatic cancers, J. Theor. Biol. 335 (2013) 235–244.
[10] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J.M.L. Ebos, L. Hlatky, P. Hahnfeldt, Classical mathematical models for description and prediction of experimental tumor growth, PLoS Comput. Biol. 10 (2014) E1003800.
[11] A. Devys, T. Goudon, P. Lafitte, A model describing the growth and the size distribution of multiple metastatic tumors, Discrete Continuous Dyn. Syst. Ser.B 12 (2009) 731–767.
[12] W. Dominik, K. Natalia, Dynamics Of Cancer: Mathematical Foundations Of Oncology, World Scientific Publishing Company, 2014.
[13] P.E. Gill, W. Murray, M.A. Saunders, E. Wong, User's Guide for SNOPT 7.7: Software for Large-Scale Nonlinear Programming, Center for Computational Mathematics Report CCoM 18-1, Department of Mathematics, University of California, San Diego, La Jolla, CA, 2018.
[14] G.P. Gupta, J. Massagué, Cancer metastasis: building a framework, Cell 127 (2006) 679–695.
[15] N. Hartung, Efficient resolution of metastatic tumor growth models by reformulation into integral equations, Discrete Continuous Dyn. Syst. Ser.B 20 (2015) 445–467.
[16] N. Hartung, S. Mollard, D. Barbolosi, A. Benabdallah, G. Chapuisat, G. Henry, S. Giacometti, A. Iliadis, J. Ciccolini, C. Faivre, F. Hubert, Mathematical modeling of tumor growth and metastatic spreading: validation in tumor-bearing mice, Cancer Res. 74 (2014) 6397–6407.
[17] D. Hawes, A.M. Neville, R. Cote, Occult metastasis, Biomed. Pharmacother. 55 (2001) 229–242.
[18] R. Herzog, J.W. Pearson, M. Stoll, Fast iterative solvers for an optimal transport problem, Adv. Comput.Math., 45 (2019) 495–517.
[19] H.C. Hoover, A.S. Ketcham, Metastasis of metastases, Am. J. Surg. 130 (1975) 405–411.
[20] V. Hovhannisyan, P. Parpas, S. Zafeiriou, MAGMA: multilevel accelerated gradient mirror descent algorithm for large-scale convex composite minimization, SIAM J Imaging Sci 9 (2016) 1829–1857.
[21] N. Hritonenko, Y. Yatsenko, R.U. Goetz, A. Xabadia, A bang–bang regime in optimal harvesting of size-structured populations, Nonlinear Anal. 71 (2009) e2331–e2336.
[22] M. Iannelli, Mathematical theory of age-structured population dynamics, Applied mathematics monographs, Giardini editori e stampatori in pisa, 1995.
[23] K. Iwata, K. Kawasaki, N. Shigesada, A dynamical model for the growth and size distribution of multiple metastatic tumors, J. Theor. Biol. 203 (2000) 177–186.
[24] C.T. Kelley, E.W. Sachs, Mesh independence of the gradient projection method for optimal control problems, SIAM J. Control Optim. 30 (1992) 477–493.
[25] D. Kirk, Optimal Control Theory: An Introduction, Dover Books on Electrical Engineering, Dover Publications, 2012.
[26] Y. Kuang, J. Nagy, S. Eikenberry, Introduction to Mathematical Oncology, Chapman & Hall/CRC Mathematical and Computational Biology, CRC Press, 2016.
[27] A.W. Lambert, D.R. Pattabiraman, R.A. Weinberg, Emerging biological principles of metastasis, Cell 168 (2017) 670–691.
[28] P. Linz, Analytical and Numerical Methods for Volterra Equations, Society for Industrial and Applied Mathematics, 1985.
[29] R. Liu, G. Liu, Optimal contraception control for a size-structured population model with extra mortality, Appl. Anal. (2018) 1–14.
[30] J.C.D.l. Reyes, Numerical PDE-Constrained Optimization, Springer International Publishing, 2015.
[31] R. Martin, Optimal control drug scheduling of cancer chemotherapy, Automatica 28 (1992) 1113–1123.
[32] R. Martin, K. Teo, Optimal Control of Drug Administration in Cancer Chemotherapy, World Scientific, 1994.
[33] J.M. Murray, Optimal control for a cancer chemotheraphy problem with general growth and loss functions, Math. Biosci. 98 (1990) 273–287.
[34] S. Nash, A multigrid approach to discretized optimization problems, Optim. Methods Softw. 14 (2000) 99–116.
[35] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, Soviet Math. Dokl. 27 (1983) 372–376.
[36] J.W. Pearson, Fast Iterative Solvers for PDE-Constrained Optimization Problems, University of Oxford, 2013 Phd thesis.
[37] L.G.D. Pillis, A. Radunskaya, A mathematical tumor model with immune resistance

and drug therapy: an optimal control approach, J. Theor. Med. 3 (2001) 79–100.
[38] M. Porcelli, V. Simoncini, M. Stoll, Preconditioning PDE-constrained optimization with L1-sparsity and control constraints, Comput. Math. Appl. 74 (2017) 1059–1075.
[39] J. Preininger, P.T. Vuong, On the convergence of the gradient projection method for convex optimal control problems with bang–bang solutions, Comput. Optim. Appl. 70 (2018) 221–238.
[40] J. Pyy, A. Ahtikoski, A. Lapin, E. Laitinen, Solution of optimal harvesting problem by finite difference approximations of size-structured population model, Math. Comput. Appl. 23 (2018) 22.
[41] J.J. Qian, E. Akçay, Competition and niche construction in a model of cancer metastasis, PLoS ONE 13 (2018) e0198163.
[42] H. Schättler, U. Ledzewicz, Optimal control for mathematical models of cancer therapies: an application of geometric methods, Interdisciplinary Applied Mathematics, Springer, New York, 2015.
[43] M. Seydenschwanz, Convergence results for the discrete regularization of linear-quadratic control problems with bang–bang solutions, Comput. Optim. Appl. 61 (2015) 731–760.
[44] J. Speyer, D. Jacobson, Primer on optimal control theory, Advances in Design and Control, SIAM, 2010.
[45] G.W. SWAN, Optimal control analysis of a cancer chemotherapy problem, Math. Med. Biol. 4 (1987) 171–184.
[46] G.W. SWAN, General applications of optimal control theory in cancer chemotherapy, Math. Med. Biol. 5 (1988) 303–316.
[47] G.W. Swan, Role of optimal control theory in cancer chemotherapy, Math. Biosci. 101 (1990) 237–284.
[48] J.E. Talmadge, S.R. Wolman, I.J. Fidler, Evidence for the clonal origin of spontaneous metastasis, Science 217 (1982) 361–363.
[49] F. Tröltzsch, J. Sprekels, Optimal control of partial differential equations: theory, methods, and applications, Graduate Studies in Mathematics, AMS, 2010.
[50] V.M. Veliov, T.E. Simos, G. Psihoyios, C. Tsitouras, Numerical optimal control of size-structured systems, AIP Conference Proceedings, AIP, (2007).
[51] N. von Daniels, Tikhonov regularization of control-constrained optimal control problems, Comput Optim Appl 70 (2017) 295–320.
[52] N. von Daniels, M. Hinze, Variational discretization of a control-constrained parabolic bang-bang optimal control problem, J. Comput. Math. 37 (2018) 361–387.
[53] A. Wächter, L.T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Math. Program. 106 (2005) 25–57.
[54] X.H. Yang, Metastasis: slipping control, Cell 168 (2017) 547–549.