Distributed parameters deterministic model for treatment of brain tumors using Galerkin finite element method

Siddhartha P. Chakrabarty [a,*] Floyd B. Hanson [b]

[a] *Department of Mathematics, Indian Institute of Technology Guwahati 781 039, Assam, INDIA*

[b] *Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan St. (m/c 249), Chicago, IL 60607, USA*

Abstract

In this paper, we present a distributed parameters deterministic model for treatment of brain tumors using Galerkin finite element method. The dynamic model comprises of system of three coupled reaction–diffusion models, involving the tumor cells, the normal tissues and the drug concentration. An optimal control problem is formulated with the goal of minimizing the tumor cell density and reducing the side effects of the drug. A distributed parameters method based on the application of variational calculus is used on an integral-Hamiltonian, which is then used to obtain an optimal coupled system of forward state equations and backward co-state equations. The Galerkin finite element method is used to realistically represent the brain structure as well as to facilitate computation. Finally a three–dimensional test case is considered and partitioned into a set of spherical finite elements, using tri–linear basis functions, except for the elements effected by singularities of polar and azimuthal angles, as well the origin.

Key words: Optimal control; Drug delivery; Brain tumors, Galerkin finite element method

* Corresponding author. Tel.: +91 361 2582606; fax: +91 361 258 2649.

Email addresses: pratim@iitg.ernet.in (Siddhartha P. Chakrabarty), hanson@math.uic.edu (Floyd B. Hanson).

URLs: http://www.iitg.ernet.in/pratim/ (Siddhartha P. Chakrabarty), http://www.math.uic.edu/∼hanson (Floyd B. Hanson).

1. Introduction

The growth and control of brain tumors have been the subject of medical and scientific scrutiny for a very long time as described in Murray's book [35, vol. II]. As in case of other cancers, brain tumors can originate from a single cell, which then proliferates and starts affecting the neighboring healthy normal tissues. As the tumor cells become malignant they become more dangerous for the host and take life threatening proportions. Understanding the mechanism that augment and abet tumor progression is necessary for its diagnosis and optimal treatment. The most common and deadly form of brain tumors are the gliomas, which account for more than half of the newly reported brain tumor cases. Gliomas are highly invasive and severely infiltrate the surrounding tissues. Araujo and McElwain [3] did a historical survey of solid tumor growth and the contribution of mathematical modeling. Baxter and Jain [5] look at the transport mechanism taking into account the interstitial pressure and convection factors. Engelhard's [13] paper provides a nice description of the difficulties that arise in drug delivery due to the presence of the blood brain barrier (BBB). Gatenby and Gawlinski [14] and Mansuri's thesis [33] dwell on some general reaction–diffusion model for cancer growth. Murray's [34,35] classical texts are an excellent reference on growth and treatment of brain tumors, with a historical prelude. Swanson's thesis [37] is arguably the most seminal work on the mathematical modeling of brain tumors using real clinical data. This work is very important because it encompasses both the mathematical and medical issues at work.

Despite improved diagnostic procedures such as computerized tomography (CT) scan and magnetic resonance imaging (MRI), the benefits of such modern accessories have been restricted by the treatment options available. According to Engelhard [13], one major problem of administering the drugs to the brain tumor site is the blood brain barrier (BBB), which exists in the human brain as a desirable protection for the brain cells against the transport of water soluble toxic substances between the blood and the central nervous system. Another problem that arises is the residue of tumor cells which remains, after the core mass of the tumor has been surgically removed, i.e., the brain has been resectioned. Note that these residual cells which are above the detection threshold in a CT/MRI scan are called *tentacles* due to their appearances on the original tumor. One way to deliver the drug is to conjugated it with a polymer and place it in the brain cavity for controlled release. One such antibody agent used is VGEF or Vascular Endothelial Growth Factor [41]. But the approach we will focus on in this paper is the optimal drug delivery at the original tumor site using invasive methods like the catheter.

While working on the mathematical representation of the brain tumor problem, researchers primary looked at two things. Firstly, the question was how

to accurately represent the growth of tumor. Secondly, the question is what is the mechanism of drug transport to the site of a brain tumor?

(1) The first aspect about the tumor growth has been elaborated in the work of Murray [35] and Swanson [37]. They point out that unlike most cancerous cells which primarily grow, gliomas are highly diffusive. Their model of brain tumor growth is given by a reaction–diffusion PDE and involves the diffusion of tumor accompanied by an exponential growth and a natural no-flux boundary condition. It is important to note that the diffusion for white matter is five times that of grey matter [37].
(2) On the other hand Wang et al. [38,39] have worked extensively on drug delivery to tumors in three dimensions for drugs like IgG and BCNU [1]. They also used a PDE for the drug transport, akin to the mass conservation equation for porous media. Their model also involves the supply and loss of drugs and also a chemical reaction term. The interstitial velocity in the porous media is mathematically approached using Darcy's law.

The mathematical model used in this work, was influenced (but is different) by the models of Gatenby and Gawlinski [14] and Mansuri [33]. Both of these papers, while not dealing directly with brain tumors, have models which closely resemble the growth of brain tumors. Both these work not only take diffusion into account but also the effects of competition for resources between the cancerous cells and healthy tissues. The different kinds of cancerous growth (exponential, logistic and Gompertz being the common ones) are detailed in the books by Goldie and Coldman [16] and Murray [34,35] and also the work of Westman et al. [40]. Also, Woodward et al. [42] study a model of glioma growth and the effects of surgical resection.

The paper is organized as follows. Section 2 presents the mathematical tumor drug delivery model. Section 3 gives the formulation of the optimal drug control problem. Section 4 concerns the Galerkin finite element method for the computational model. Section 5 sets up the customized spherical finite element configuration. Section 6 describes the forward state to backward co-state double iterations. Section 7 gives details of the forward–backward computational algorithm. Finally, conclusions and future directions are presented in Section 8.

2. Mathematical model

This work while being motivated by a biomedical problem will be mostly mathematical and computational in content. We will focus primarily on control for the optimal distribution of the drug about the original tumor site after the bulk of the tumor has been surgically removed. In the PDE driven distributed parameters control model of Chakrabarty and Hanson [8–10] and Chakrabarty

[7], the density of tumor cells and normal tissues and the drug concentration are taken as the state variables. Let $T(\mathbf{x},t)$ be the density of tumor cells, $N(\mathbf{x},t)$ be the density of normal tissues and $C(\mathbf{x},t)$ be the drug concentration at any vector position $\mathbf{x}$ and time $t \in [0,t_f]$, in the interior of the domain Ω. Since this is a PDE driven control model it is called a *distributed parameters model*, as opposed to a *lumped parameter model* in the case for ODE models. While previous models only study the behavior of solutions we also study the control of drug delivery. This spatiotemporal model is a system of three coupled reaction–diffusion equations.

2.1. Tumor cells

It is assumed that the density of tumor cells, $T(\mathbf{x},t)$, satisfies a reaction–diffusion equation subject to competition with the normal cells with density $N(\mathbf{x},t)$ and killing due to drug concentration $C(\mathbf{x},t)$,

$$\frac{\partial T}{\partial t} = D_T \nabla_x^2[T] + a_T g_T(T) T - (\alpha_{T,N} N + \kappa_{T,C} C)\, T, \tag{2.1}$$

where the constant tumor diffusivity is D_T. The term $a_T g_T(T) T$ is the logistic growth rate of the tumor cells, i.e.,

$$g_T(T) = (1 - T/k_T)\,,$$

where a_T is the tumor cell intrinsic growth rate and k_T is the tumor cell carrying capacity. The interaction coefficient $\alpha_{T,N}$ is the death rate of the tumor cells due to competition for resources with the normal tissue, while $\kappa_{T,C}$ is the death rate of tumor cells due to drug treatment and could also be a non-linear function of the localized drug concentration $C(\mathbf{x},t)$ at the tumor site. The initial and boundary conditions for the tumor cells are given by $T(\mathbf{x},0) = T_0$ (T_0 will be specified later) and $D_T(\hat{\mathbf{n}}\cdot\nabla_x)[T] = 0$ respectively.

2.2. Normal tissues

Similar assumptions are made for the density of normal cells $N(\mathbf{x},t)$ with similar coefficients. Thus, the reaction–diffusion equation for normal tissue evolution is as follows,

$$\frac{\partial N}{\partial t} = D_N \nabla_x^2[N] + a_N g_N(N) N - (\alpha_{N,T} T + \kappa_{N,C} C)\, N, \tag{2.2}$$

where the normal tissue diffusivity (negligible compared to that of tumor cells) is D_N. As in the previous case, $a_N g_N(N)N$ is the logistic growth rate of normal tissue, i.e.,

$$g_N(N) = (1 - N/k_N)\,,$$

where a_N is the intrinsic growth rate and k_N is the normal tissue carrying capacity. $\alpha_{N,T}$ is the death rate of normal tissues due to competition with tumor cells. Note that the $-\kappa_{N,C}CN$ term indicates that some normal tissues die as a result of the treatment. The term $\kappa_{N,C}$ could also be a non-linear function of the localized drug concentration $C(\mathbf{x}, t)$. The initial and boundary conditions for the normal tissue are given by $N(\mathbf{x}, 0) = N_0$ (N_0 will be specified later) and $D_N(\hat{\mathbf{n}}\cdot\nabla_x)[N] = 0$ respectively.

2.3. Drug concentration

The drug shows a diffusive behavior and that there is a reabsorption at the rate a_C. Also $U = U(\mathbf{x}, t)$ is the rate at which the drug is being injected through catheter or released through a drug polymer. The choice of the symbol U indicates that we will use it as the control variable when dealing with the optimal control system. We would like to point out that it is reasonable (as is usually the case in such biomedical problems) to have the control to be temporally dependent only. However in this case, the rationale behind taking a distributed control is that physically the drug is being supplied using a catether or a drug polymer. In case of the former the dependece on time only would be prudent. But for a drug polymer the release would be spatially dependent too and hence the usage of the distributed control. The equation for drug concentration at position $\mathbf{x}$ and time t is,

$$\frac{\partial C}{\partial t} \;=\; D_C\nabla_x^2[C] + a_C C g_C(C) + U, \tag{2.3}$$

where $Cg_C(C) = -C$ is the linear reabsorption function [39] and D_C is the constant concentration diffusivity. The intrinsic growth terms, like $Cg_C(C)$, have been kept simple, since the main computational demand is the optimal control problem. The initial and boundary conditions for the drug concentration are given by $C(\mathbf{x}, 0) = C_0$ (C_0 will be specified later) and $D_C(\hat{\mathbf{n}}\cdot\nabla_x)[C] = 0$ respectively.

2.4. Vector formulation

For compact notation, the global state vector is defined to be,

$$\mathbf{Y}(\mathbf{x},t) = [Y_i(\mathbf{x},t)]_{3\times 1} = \begin{bmatrix} T(\mathbf{x},t) \\ N(\mathbf{x},t) \\ C(\mathbf{x},t) \end{bmatrix}, \tag{2.4}$$

at position $\mathbf{x}$ in the interior of the domain Ω (the whole brain) at time $t \in [0, t_f]$, the governing non-linear vector PDE is given by,

$$\frac{\partial \mathbf{Y}}{\partial t}(\mathbf{x},t) = D\nabla_x^2[\mathbf{Y}](\mathbf{x},t) + (A+B)(\mathbf{Y}(\mathbf{x},t))\mathbf{Y}(\mathbf{x},t) + \mathbf{U}(\mathbf{x},t), \tag{2.5}$$

where

$$\begin{aligned} D &= D_T\mathbf{e}_1\mathbf{e}_1^\top + D_N\mathbf{e}_2\mathbf{e}_2^\top + D_C\mathbf{e}_3\mathbf{e}_3^\top, \\ A(\mathbf{Y}) &= a_T(1-T/k_T)\mathbf{e}_1\mathbf{e}_1^\top + a_N(1-N/k_N)\mathbf{e}_2\mathbf{e}_2^\top - a_C\mathbf{e}_3\mathbf{e}_3^\top, \\ B(\mathbf{Y}) &= -(\alpha_{T,N}N + \kappa_{T,C}C)\mathbf{e}_1\mathbf{e}_1^\top - (\alpha_{N,T}T + \kappa_{N,C}C)\mathbf{e}_2\mathbf{e}_2^\top, \end{aligned}$$

and

$$\mathbf{U}(\mathbf{x},t) = U_3(\mathbf{x},t)\mathbf{e}_3. \tag{2.6}$$

Note that the bilinear component functions, A and B, should be explicit functions of $\mathbf{x}$ and t as well as $\mathbf{Y}(\mathbf{x},t)$. Here, $\mathbf{e}_i$ is the ith unit vector and $U_3(\mathbf{x},t) = U(\mathbf{x},t)$. It should be noted that the diffusion operator $D\nabla_x^2$ will have to be replaced by $\nabla_x^\top \cdot D(\mathbf{x},t)\nabla_x$ in case of inhomogeneous brain matter. The initial conditions for the state equations are given by,

$$\mathbf{Y}(\mathbf{x},0) = \begin{bmatrix} T(\mathbf{x},0) \\ N(\mathbf{x},0) \\ C(\mathbf{x},0) \end{bmatrix} = \begin{bmatrix} T_0(\mathbf{x}) \\ N_0(\mathbf{x}) \\ C_0(\mathbf{x}) \end{bmatrix} \equiv \mathbf{Y}_0(\mathbf{x}), \tag{2.7}$$

for $\mathbf{x}$ in Ω. Murray [35] recommends using Gaussian distribution for the initial distributions of tumors, but a conical distribution of compact support will be used to avoid spurious spreading of cancer cells due to exponential tails of the Gaussian. The no-flux boundary conditions are,

$$-D(\widehat{\mathbf{n}}\cdot\nabla_x)[\mathbf{Y}](\mathbf{x},t) = \begin{bmatrix} -D_T(\widehat{\mathbf{n}}\cdot\nabla_x)[T] \\ -D_N(\widehat{\mathbf{n}}\cdot\nabla_x)[N] \\ -D_C(\widehat{\mathbf{n}}\cdot\nabla_x)[C] \end{bmatrix}(\mathbf{x},t) = \mathbf{0}, \tag{2.8}$$

assuming that the diffusion coefficients are not zero (or else they would not be used in the condition), for $\mathbf{x} \in \mathbf{\Gamma} = \partial\Omega$, i.e., on the boundary of the domain, and for $t \in [0, t_f]$. Here $\widehat{\mathbf{n}}(\mathbf{x}, t)$ is the normal to the boundary. The diffusion matrix,

$$D = \begin{bmatrix} D_T & 0 & 0 \\ 0 & D_N & 0 \\ 0 & 0 & D_C \end{bmatrix}$$

is diagonal, but could be inhomogeneous depending on the brain matter [37]. At this point we note that for the purpose of computation in this paper we have taken the diffusion coefficient to be homogeneous even though the code could be set up using the same scheme to handle the case of inhomogeneous diffusion coefficient. Note that the no-flux condition at the boundary is a natural condition motivated by the physical reality that the brain is a finite and closed domain.

The interested reader (looking for a more mathematical flavor for the problem) may refer to the distributed parameter system references of Ahmed and Teo [2], Banks and Kunish [4] or Lions [32]. Also, the reader may refer to the Ph.D. thesis of Joshi [27] for a readable account on how to prove the existence of distributed parameter control state system. In addition, the authors are aware of the proof of the existence of our system that appeared as an abstract [29]. The interested readers may refer to a paper cited [30] in this abstract for a proof of the existence of the solution and the control itself for such systems.

3. Optimal control problem

In this section, we formulate an optimal control problem for general dimensions. The objective functional is taken to be a quadratic form of running and terminal costs,

$$\begin{aligned} J[\mathbf{x}, U] = \frac{1}{2} \int_0^{t_f} dt \int_\Omega d\mathbf{x} \left(r_T T^2(\mathbf{x}, t) + s_U \left(U - U_0 \right)^2 (\mathbf{x}, t) \right) \\ + \int_\Omega d\mathbf{x} \left(q_T T^2(\mathbf{x}, t_f) + q_C C^2(\mathbf{x}, t_f) \right). \end{aligned} \tag{3.9}$$

The goal is to minimize this functional with respect to the drug input rate $U(\mathbf{x}, t)$ relative to some threshold rate $U_0(\mathbf{x}, t)$ and the terminal costs at t_f, i.e., $\min_U [J[\mathbf{x}, U]]$. In other words, we are trying to minimize the density of tumor cells and the drug delivery quadratic control term $(U(\mathbf{x}, t) - U_0(\mathbf{x}, t))^2$, as also the tumor density and drug concentration at final time to reduce the effects of toxicity. The concentration term is not included in the running

costs, keeping in mind some amount of side effects is an inevitable part of the treatment. Note that here $r_T > 0$ is the tumor burden cost coefficient and $s_U > 0$ is the drug delivery cost coefficient, while $q_T > 0$ and $q_C > 0$ are the corresponding tumor and drug final costs. We could have chosen a linear control which would not only have been less realistic, but would have given rise to singular control complications. In addition no assumption is made about the control constraints, but $U_0(\mathbf{x}, t)$ serves as physical restriction on the amount and costs of drugs that can be administered. While we have U as unconstrained regular control, there is a physical upper limit in terms of drug that can be injected.

The quadratic objective functional in more general vector form is given by,

$$\begin{aligned} J[\mathbf{Y}, \mathbf{U}] = \frac{1}{2}\int_0^{t_f} dt \int_\Omega d\mathbf{x} \left(\mathbf{Y}^\top R \mathbf{Y} + (\mathbf{U}–\mathbf{U}_0)^\top S \, (\mathbf{U}–\mathbf{U}_0) \right)(\mathbf{x}, t) \\ + \frac{1}{2}\int_\Omega d\mathbf{x} \left(\mathbf{Y}^\top Q \mathbf{Y} \right)(\mathbf{x}, t_f), \end{aligned} \tag{3.10}$$

where $R = r_T \mathbf{e}_1 \mathbf{e}_1^\top$, $S = s_U \mathbf{e}_3 \mathbf{e}_3^\top$, $Q = q_T \mathbf{e}_1 \mathbf{e}_1^\top + q_C \mathbf{e}_3 \mathbf{e}_3^\top$, and $\mathbf{U}_0 = U_0(\mathbf{x}, t)\mathbf{e}_3$. To define an *integral-Hamiltonian*, we use three vectors for the *Lagrange multipliers*, to include the optimization constraints in the extended objective for the state PDE (2.5), the initial condition (2.7) and the boundary condition (2.8). Two multipliers are functions of space and time, while one is independent of time,

$$\boldsymbol{\xi}(\mathbf{x}, t) = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}, \quad \boldsymbol{\eta}(\mathbf{x}, t) = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}, \quad \boldsymbol{\chi}(\mathbf{x}) = \begin{bmatrix} \chi_1 \\ \chi_2 \\ \chi_3 \end{bmatrix}, \tag{3.11}$$

i.e., $\xi_i = \xi_i(\mathbf{x}, t), \eta_i = \eta_i(\mathbf{x}, t)$ and $\chi_i = \chi_i(\mathbf{x})$, for $i = 1 : 3$. Letting $\mathbf{Z} = (\mathbf{Y}, \mathbf{U}, \boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\chi})$, be an extended state vector we define the inclusive *integral-Hamiltonian* as in Gunzburger [18],

$$
\begin{aligned}
\mathcal{H}(\mathbf{Z}) \equiv & \frac{1}{2}\int_0^{t_f} dt \int_\Omega d\mathbf{x} \left(\mathbf{Y}^\top R \mathbf{Y} + (\mathbf{U}-\mathbf{U}_0)^\top S \, (\mathbf{U}-\mathbf{U}_0) \right) \\
& + \frac{1}{2} \int_\Omega d\mathbf{x} \left(\mathbf{Y}^\top Q \mathbf{Y} \right)(\mathbf{x}, t_f) \\
& + \int_0^{t_f} dt \int_\Omega d\mathbf{x} \; \boldsymbol{\xi}^\top \left(\frac{\partial \mathbf{Y}}{\partial t} - D \nabla_x^2 [\mathbf{Y}] - (A+B)(\mathbf{Y}) \mathbf{Y} - \mathbf{U} \right) \\
& + \int_0^{t_f} dt \int_{\partial\Omega} d\boldsymbol{\Gamma} \; \boldsymbol{\eta}^\top \left(-D \left(\hat{\mathbf{n}} \cdot \nabla_x \right) [\mathbf{Y}] \right) \\
& + \int_\Omega d\mathbf{x} \; \left(\boldsymbol{\chi}^\top (\mathbf{Y} - \mathbf{Y}_0) \right)(\mathbf{x}, 0).
\end{aligned}
\tag{3.12}
$$

3.1. Derivation of optimal trajectory equations

We then modify the standard method of *calculus of variations* for this integral formulation to find the differential equations for the optimal controls, states and the co-states (adjoint variables or Lagrange multipliers) by seeking the functional critical point necessary conditions for the first variation of the *integral-Hamiltonian* $\mathcal{H}(\mathbf{Z})$ (Gunzburger [18] and Kirk [31]). The extended state vector is perturbed about the optimal trajectory $\mathbf{Z}^*$, so that $\mathbf{Z} = \mathbf{Z}^* + \delta\mathbf{Z}$, where $\delta\mathbf{Z}$ is the perturbation. Next, simplifying and modifying the approach in Gunzburger [18], we expand the *integral-Hamiltonian* as,

$$
\mathcal{H}(\mathbf{Z}^* + \delta\mathbf{Z}) = \mathcal{H}(\mathbf{Z}^*) + \delta\mathcal{H}(\mathbf{Z}^*, \delta\mathbf{Z}) + O\left((\delta\mathbf{Z})^2 \right).
$$

Neglecting the quadratic order Hamiltonian perturbation terms, including the second variation of $\mathcal{H}$, the first variation is given by functional terms linear in $\delta\mathbf{Z}$ using (3.12),

$$
\begin{aligned}
\delta\mathcal{H}(\mathbf{Z}^*,\delta\mathbf{Z}) = &\int_0^{t_f} dt \int_\Omega d\mathbf{x}\left((\mathbf{Y}^*)^\top R\delta\mathbf{Y} + (\mathbf{U}^* - \mathbf{U}_0)^\top S\delta\mathbf{U}\right)\\
&+\int_\Omega d\mathbf{x}\left((\mathbf{Y}^*)^\top Q\delta\mathbf{Y}\right)(\mathbf{x},t_f)\\
&+\int_0^{t_f} dt \int_\Omega d\mathbf{x}\left((\boldsymbol{\xi}^*)^\top\left(\delta\frac{\partial\mathbf{Y}}{\partial t} - D\nabla_x^2[\delta\mathbf{Y}]\right.\right.\\
&\left.-(A+B)(\mathbf{Y}^*)\delta\mathbf{Y} - (\delta\mathbf{Y}\cdot\nabla_Y)[A+B](\mathbf{Y}^*) - \delta\mathbf{U}\right)\\
&\left.+\delta\boldsymbol{\xi}^\top\left(\frac{\partial\mathbf{Y}^*}{\partial t} - D\nabla_x^2[\mathbf{Y}^*] - (A+B)(\mathbf{Y}^*)\mathbf{Y}^* - \mathbf{U}^*\right)\right)\\
&-\int_0^{t_f} dt \int_{\partial\Omega} d\boldsymbol{\Gamma}\left((\boldsymbol{\eta}^*)^\top D\left(\hat{\mathbf{n}}\cdot\nabla_x\right)[\delta\mathbf{Y}] + \delta\boldsymbol{\eta}^\top D\left(\hat{\mathbf{n}}\cdot\nabla_x\right)[\mathbf{Y}^*]\right)\\
&+\int_\Omega d\mathbf{x}\left((\boldsymbol{\chi}^*)^\top\delta\mathbf{Y} + \delta\boldsymbol{\chi}^\top(\mathbf{Y} - \mathbf{Y}_0)\right)(\mathbf{x},0).
\end{aligned}
\tag{3.13}
$$

Before the critical conditions for first variation in (3.13) can be applied to obtain the extended state equations, the higher order derivatives in time and state of the extended state perturbations must be eliminated in favor of independent lower order terms, first using integrations by parts,

$$
\int_0^{t_f} dt(\boldsymbol{\xi}^*)^\top\delta\mathbf{Y}_t = (\boldsymbol{\xi}^*)^\top\delta\mathbf{Y}\Big|_0^{t_f} - \int_0^{t_f} dt\delta\mathbf{Y}^\top\boldsymbol{\xi}_t^*
$$

and secondly by using the Green's formula as in Haberman [19],

$$
\begin{aligned}
\int_\Omega d\mathbf{x}(\boldsymbol{\xi}^*)^\top D\nabla_x^2[\delta\mathbf{Y}] = &\int_\Omega d\mathbf{x}\delta\mathbf{Y}^\top\nabla_x^2[D\boldsymbol{\xi}^*]\\
&+\int_{\partial\Omega} d\boldsymbol{\Gamma}\left((\hat{\mathbf{n}}\cdot\nabla_x)[\delta\mathbf{Y}^\top]D\boldsymbol{\xi}^* - \delta\mathbf{Y}^\top(\hat{\mathbf{n}}\cdot\nabla_x)[D\boldsymbol{\xi}^*]\right).
\end{aligned}
$$

Note that here we have used the Green's second identity,

$$
\int_\Omega d\mathbf{x}\left(u\nabla^2[v] - v\nabla^2[u]\right) = \int_{\partial\Omega} d\boldsymbol{\Gamma}\left(u\hat{\mathbf{n}}\cdot\nabla[v] - v\hat{\mathbf{n}}\cdot\nabla[u]\right).
$$

Merging these identities with (3.13), rearranging inner products and collecting terms the extended state equations yields the following intermediate form:

$$
\begin{aligned}
\delta\mathcal{H}(\mathbf{Z}^*,\delta\mathbf{Z}) = &\int_0^{t_f} dt \int_\Omega d\mathbf{x}\, \delta\mathbf{Y}^\top \Big(R\mathbf{Y}^* - \frac{\partial \boldsymbol{\xi}^*}{\partial t} - \nabla_x^2[D\boldsymbol{\xi}^*] - (A+B)(\mathbf{Y}^*)\boldsymbol{\xi}^* \\
&- \nabla_Y[A+B](\mathbf{Y}^*) : (\boldsymbol{\xi}^*(\mathbf{Y}^*)^\top \Big) + \int_0^{t_f} dt \int_\Omega d\mathbf{x}\, \delta\mathbf{U}^\top (S\,(\mathbf{U}^* - \mathbf{U}_0) - \boldsymbol{\xi}^*) \\
&+ \int_0^{t_f} dt \int_\Omega d\mathbf{x}\, \delta\boldsymbol{\xi}^\top \Big(\frac{\partial \mathbf{Y}^*}{\partial t} - D\nabla_x^2[\mathbf{Y}^*] - (A+B)(\mathbf{Y}^*)\mathbf{Y}^* - \mathbf{U}^* \Big) \\
&- \int_0^{t_f} dt \int_{\partial\Omega} d\boldsymbol{\Gamma}\, \delta\boldsymbol{\eta}^\top D\,(\widehat{\mathbf{n}}\cdot\nabla_x)\,[\mathbf{Y}^*] + \int_0^{t_f} dt \int_{\partial\Omega} d\boldsymbol{\Gamma}\, \delta\mathbf{Y}^\top (\widehat{\mathbf{n}}\cdot\nabla_x)\,[D\boldsymbol{\xi}^*] \\
&- \int_0^{t_f} dt \int_{\partial\Omega} d\boldsymbol{\Gamma} (\widehat{\mathbf{n}}\cdot\nabla_x) \Big[\delta\mathbf{Y}^\top\Big] D(\boldsymbol{\eta}^* + \boldsymbol{\xi}^*) \\
&+ \int_\Omega d\mathbf{x} \left(\delta\boldsymbol{\chi}^\top\, (\mathbf{Y}^* - \mathbf{Y}_0(\mathbf{x})) \right)(\mathbf{x},0) \\
&+ \int_\Omega d\mathbf{x} \left(\delta\mathbf{Y}^\top (\boldsymbol{\chi}^* - \boldsymbol{\xi}^*) \right)(\mathbf{x},0) + \int_\Omega d\mathbf{x} \left(\delta\mathbf{Y}^\top (\boldsymbol{\xi}^* + Q\mathbf{Y}^*) \right)(\mathbf{x},t_f),
\end{aligned}
$$

where $A{:}B$ denotes the trace of the matrix AB or the double-dot product, e.g.,

$$
A : B = \sum_{j=1}^{3}\sum_{k=1}^{3} A_{ij}B_{jk} = \text{Trace}[AB].
$$

Using the fundamental theorem of *calculus of variations* [23,31] we set the coefficients of the independent variations to be equal to zero to obtain the optimal state, control and the co-state equations.

3.2. State equations

The optimal state equation is recovered by setting the coefficient of $(\delta\boldsymbol{\xi})^\top$ to zero:

$$
\frac{\partial \mathbf{Y}^*}{\partial t} = D\nabla_x^2[\mathbf{Y}^*] + (A+B)(\mathbf{Y}^*)\mathbf{Y}^* + \mathbf{U}^* \tag{3.14}
$$

on $\Omega \times (0, t_f]$, with boundary conditions on $\partial\Omega \times [0, t_f]$ from the coefficient of $(\delta\boldsymbol{\eta})^\top$, i.e.,

$$
-D(\widehat{\mathbf{n}}\cdot\nabla_x)[\mathbf{Y}^*](\mathbf{x},t) = \mathbf{0}, \quad (\mathbf{x},t) \in \partial\Omega \times [0,t_f] \tag{3.15}
$$

and with initial conditions on the interior Ω from the coefficient of $(\delta\boldsymbol{\chi})^\top$,i.e.,

$$
\mathbf{Y}^*(\mathbf{x},0) = \mathbf{Y}_0(\mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{3.16}
$$

Due to the presence of the functions $(A+B)(\mathbf{Y}^*)\mathbf{Y}^*$ the forward PDE (3.14) is non-linear.

3.3. Regular optimal control

Since the control has been defined in (2.6) as only having one component (the other components are automatically zero), only the coefficient of δU_3 is set to zero giving the corresponding regular control

$$U_3^*(\mathbf{x},t) = U_0(\mathbf{x},t) + \xi_3^*(\mathbf{x},t)/s_3, \quad (\mathbf{x},t)\in\Omega\times[0,t_f], \tag{3.17}$$

provided $s_3 \neq 0$. Note that this control law only requires solving for the third component of the first co-state vector $\boldsymbol{\xi}^*(\mathbf{x},t)$, since $\delta U_1 \equiv 0$ and $\delta U_2 \equiv 0$.

3.4. Co-state equations

Setting the functional coefficient of $(\delta\mathbf{Y})^\top$ to zero yields the primary co-state backward PDE,

$$\begin{aligned}\mathbf{0} = \frac{\partial\boldsymbol{\xi}^*}{\partial t} &+ \nabla_x^2[D\boldsymbol{\xi}^*] + (A+B)(\mathbf{Y}^*)\boldsymbol{\xi}^* \\ &+ \nabla_Y[A+B](\mathbf{Y}^*):\Big(\boldsymbol{\xi}^*(\mathbf{Y}^*)^\top\Big) - R\mathbf{Y}^*,\end{aligned} \tag{3.18}$$

for $(\mathbf{x},t)\in\Omega\times[0,t_f)$. This PDE (3.18) is unidirectionally coupled to the state PDE (3.14), except that only the third component $\xi_3^*(\mathbf{x},t)$ is needed for the regular optimal control $U_3^*(\mathbf{x},t)$ from (3.17). The boundary condition follows from setting the coefficient of $\delta\mathbf{Y}(\mathbf{x},t)$, for $\mathbf{x}\in\boldsymbol{\Gamma}=\partial\Omega$ to zero, so

$$(\widehat{\mathbf{n}}\cdot\nabla_x)[D\boldsymbol{\xi}^*](\mathbf{x},t) = \mathbf{0}, \quad (\mathbf{x},t)\in\partial\Omega\times[0,t_f) \tag{3.19}$$

and the final condition for this backward PDE follows from forcing the coefficient of $\delta\mathbf{Y}(\mathbf{x},t_f)$ to be zero on Ω,

$$\boldsymbol{\xi}^*(\mathbf{x},t_f) = -Q\mathbf{Y}(\mathbf{x},t_f), \quad \mathbf{x}\in\Omega. \tag{3.20}$$

The two other co-state vectors should not be needed, but satisfy rather simple equations. The second co-state vector equation follows as the zero coefficient of $(\widehat{\mathbf{n}}\cdot\nabla_x)[\delta\mathbf{Y}^\top]$ on the state boundary $\boldsymbol{\Gamma}=\partial\Omega$,

$$\boldsymbol{\eta}^*(\mathbf{x},t) = -\boldsymbol{\xi}^*(\mathbf{x},t), \quad (\mathbf{x},t) \in \partial\Omega \times [0,t_f].$$

The third co-state vector equation follows as the zero coefficient of state initial condition $\delta\mathbf{Y}(\mathbf{x},0)$,

$$\boldsymbol{\chi}^*(\mathbf{x}) = \boldsymbol{\xi}^*(\mathbf{x},0), \quad \mathbf{x} \in \Omega.$$

4. Galerkin finite element method

The PDEs derived in the previous section are highly non-linear and coupled in nature. As such they require numerical methods instead of analytical methods. We have used modified Crank–Nicolson method [8] to obtain results in one–dimension. However, using finite difference methods like Crank–Nicolson method or alternating directions implicit method have serious drawbacks and they can not be directly used for non-linear problems using predictor–corrector modifications. Moreover finite difference techniques are more likely to have higher computational requirements, *i.e.,* they suffer from the *curse of dimensionality*. Finite element methods require a relatively smaller number of nodes as compared to the finite difference methods while maintaining the same level of accuracy. Also, the finite element methods can better handle irregular structure, such as the brain tumor and brain structure. Hanson [21,22] has worked extensively in high performance computing and has made a comparative study of different numerical methods for stochastic dynamic programming. For the problem under consideration, we use the Galerkin finite element method so as to reduce the number of state nodes.

4.1. Galerkin approximation and element integrals

In this section, we reduce the state and co-state PDEs to Galerkin ODEs. The Galerkin approximation for any state or control vector summation in the optimal global vector $\mathbf{Z}^*(\mathbf{x},t)$ is given by,

$$\mathbf{Z}^*(\mathbf{x},t) \simeq \widehat{\mathbf{Z}}(\mathbf{x},t) \equiv \sum_{i=1}^{\widehat{M}} \widehat{\mathbf{Z}}_i(t) \cdot \phi_i(\mathbf{x}),$$

where $[\phi_i(\mathbf{x})]_{\widehat{M}\times 1}$ is a set of $\widehat{M}$ linearly independent continuous basis functions, with the normalization property,

$$\phi_i(\mathbf{x}_j) = \delta_{j,i},$$

at the element node $\mathbf{x}_j$, implying the interpolation property that

$$\mathbf{Z}^*(\mathbf{x}_j, t) = \widehat{\mathbf{Z}}_j(t),$$

for $j = 1{:}\widehat{M}$ finite element nodes. We now define some element integral matrix coefficients that will be used in the Galerkin ODE formulation.

(1) The element mass integral for $i, j = 1{:}\widehat{M}$ is:

$$\mathcal{M}_{i,j} \equiv \int_{\Omega} d\mathbf{x} \phi_i(\mathbf{x}) \phi_j(\mathbf{x}). \tag{4.21}$$

(2) The element stiffness integral for $i, j = 1{:}\widehat{M}$ is:

$$\mathcal{K}_{i,j} \equiv \int_{\Omega} d\mathbf{x} \nabla_x^{\top}[\phi_i](\mathbf{x}) \nabla_x[\phi_j](\mathbf{x}). \tag{4.22}$$

(3) The triple basis element integral for $i, j, k = 1{:}\widehat{M}$ newly arising from the non-linear terms is:

$$\mathcal{T}_{i,j,k} \equiv \int_{\Omega} d\mathbf{x} \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) \phi_k(\mathbf{x}). \tag{4.23}$$

The Galerkin basis integral coefficients $(\mathcal{M}_{i,j}, \mathcal{K}_{i,j}, \mathcal{T}_{i,j,k})$ can be computed by exact symbolic methods or numerical quadrature if there is sufficient element complexity. In our case Maple™ is used to do this computation symbolically. The predictor–corrector modification is necessary to handle the non -linearities and multi-dimension. These coefficients were calculated on an element-by-element decomposition and element results are later reassembled to form the global solution as in Sewell [36].

4.2. Galerkin ODEs

The Galerkin ODEs that we will derive now, analogous to the optimal control problem derivation, is numerically solved by our extrapolated predictor–corrector Crank–Nicolson method [8], the details of which are outlined in a later section.

(1) Let the Galerkin approximation for the state be,

$$\mathbf{Y}^*(\mathbf{x}, t) \simeq \widehat{\mathbf{Y}}(\mathbf{x}, t) \equiv \sum_{i=1}^{\widehat{M}} \widehat{\mathbf{Y}}_i(t) \cdot \phi_i(\mathbf{x}), \tag{4.24}$$

along with a similar approximation for the optimal control,

$$\mathbf{U}^*(\mathbf{x},t) \simeq \widehat{\mathbf{U}}(\mathbf{x},t) \equiv \sum_{i=1}^{\widehat{M}} \widehat{\mathbf{U}}_i(t)\cdot\phi_i(\mathbf{x}). \tag{4.25}$$

The properties of $[\phi_i(\mathbf{x})]_{\widehat{M}\times 1}$ have already been described. Before applying the Galerkin approximation (4.24) to the state equation (3.14), the equation must be put into integral form on Ω with respect to a test function $\phi_i(\mathbf{x})$ taken from the basis and then further prepared for low order basis function by reducing the second order derivatives to first order derivatives using integration by parts (Green's formula [19]). Thus,

$$\begin{aligned}
\mathbf{0} &= \int_\Omega d\mathbf{x}\phi_i(\mathbf{x})\left(\mathbf{Y}_t^* - D\nabla_x^2[\mathbf{Y}^*] - (A+B)(\mathbf{Y}^*)\mathbf{Y}^* - \mathbf{U}^*\right)\\
&= \int_\Omega d\mathbf{x}\left(\phi_i\mathbf{Y}_t^* + D\nabla_x^\top[\phi_i]\nabla_x[\mathbf{Y}^*] - \phi_i\left((A+B)(\mathbf{Y}^*)\mathbf{Y}^* + \mathbf{U}^*\right)\right)\\
&\quad -\int_{\partial\Omega} d\mathbf{\Gamma}\phi_i D(\widehat{\mathbf{n}}\cdot\nabla_x)\left[\mathbf{Y}^*\right]\\
&= \int_\Omega d\mathbf{x}\left(\phi_i\mathbf{Y}_t^* + D\nabla_x^\top[\phi_i]\nabla_x[\mathbf{Y}^*] - \phi_i\left((A+B)(\mathbf{Y}^*)\mathbf{Y}^* + \mathbf{U}^*\right)\right),
\end{aligned}$$

for $i = 1{:}\widehat{M}$, where the exact no-flux boundary condition has been used in the last step. Note that for the Galerkin approximation to be compatible with this no-flux condition, the boundary basis functions $\phi_j(\mathbf{x})$ would best satisfy this condition on $\partial\Omega$, or the no-flux condition should be satisfied in the variational integral form neglected in the exact formulation above. We would like to point out that the no-flux condition in the implementation of the code was done using a discretization (see section 5.2). The proper theoretical setting for the no-flux should be for the usual finite elements, but for practical computational reasons for the spherical elements, we found that forcing the boundary elements to automatically satisfy no-flux was a more certain and efficient technique to prevent computational leakage at the boundary. Now, the Galerkin approximation (4.24) can be applied yielding,

$$\begin{aligned}
\mathbf{0} \simeq \sum_{j=1}^{\widehat{M}} \int_\Omega d\mathbf{x} &\left(\widehat{\mathbf{Y}}_j'\phi_i\phi_j + D\widehat{\mathbf{Y}}_j\left(\nabla_x^\top[\phi_i]\nabla_x[\phi_j]\right)\right.\\
&\left.-\left((A+B)\left(\widehat{\mathbf{Y}}\right)\widehat{\mathbf{Y}}_j + \widehat{\mathbf{U}}_j\right)\phi_i\phi_j\right),
\end{aligned}$$

for $i = 1{:}\widehat{M}$. Upon further reduction using the finite element integrals the compact state Galerkin ODE is given by:

$$
\begin{aligned}
\mathbf{0} = &\sum_{\mathrm{J}=1}^{\widehat{M}} \Big(\mathcal{M}_{i,j} \left(\widehat{\mathbf{Y}}'_j(t) - \Big(a_1 \mathbf{e}_1 \mathbf{e}_1^\top + a_2 \mathbf{e}_2 \mathbf{e}_2^\top \right. \\
&\left. - a_3 \mathbf{e}_3 \mathbf{e}_3^\top \Big) \widehat{\mathbf{Y}}_j(t) - \widehat{\mathbf{U}}_j(t) \Big) + \mathcal{K}_{i,j} D \widehat{\mathbf{Y}}_j(t) \right. \\
&+ \sum_{k=1}^{\widehat{M}} \mathcal{T}_{i,j,k} \Big(\frac{a_1}{k_1} \widehat{Y}_{1,j}(t) \widehat{Y}_{1,k}(t) \mathbf{e}_1 + \frac{a_2}{k_2} \widehat{Y}_{2,j}(t) \widehat{Y}_{2,k}(t) \mathbf{e}_2 \\
&+ \Big(\alpha_{1,2} \widehat{Y}_{2,k}(t) + \kappa_{1,3} \widehat{Y}_{3,k}(t) \Big) \widehat{Y}_{1,j}(t) \mathbf{e}_1 \\
&\left. + \Big(\alpha_{2,1} \widehat{Y}_{1,k}(t) + \kappa_{2,3} \widehat{Y}_{3,k}(t) \Big) \widehat{Y}_{2,j}(t) \mathbf{e}_2 \Big) \right),
\end{aligned} \tag{4.26}
$$

for $i = 1{:}\widehat{M}$.

(2) Similar to the state equation, a Galerkin approximation for the co-state equation using the same basis is given by,

$$
\boldsymbol{\xi}^*(\mathbf{x}, t) \simeq \widehat{\boldsymbol{\xi}}(\mathbf{x}, t) \equiv \sum_{j=1}^{\widehat{M}} \widehat{\boldsymbol{\xi}}_j(t) \cdot \phi_j(\mathbf{x}) \tag{4.27}
$$

for $t < t_f$. As with the state Galerkin variational formulation, the variation formulation for the co-state equation (3.18) is

$$
\begin{aligned}
\mathbf{0} = &\int_\Omega d\mathbf{x} \phi_i(\mathbf{x}) \left(\boldsymbol{\xi}^*_t + \nabla_x^2 [D \boldsymbol{\xi}^*] + (A+B)(\mathbf{Y}^*) \boldsymbol{\xi}^* \right. \\
&\left. + \nabla_Y [A+B](\mathbf{Y}^*) : (\boldsymbol{\xi}^* (\mathbf{Y}^*)^\top) - R \mathbf{Y}^* \right) \\
= &\int_\Omega d\mathbf{x} \left(\phi_i \left(\boldsymbol{\xi}^*_t + (A+B)(\mathbf{Y}^*) \boldsymbol{\xi}^* \right. \right. \\
&\left. + \nabla_Y [A+B](\mathbf{Y}^*) : (\boldsymbol{\xi}^* (\mathbf{Y}^*)^\top) - R \mathbf{Y}^* \right) \\
&\left. - \nabla_x^\top [\phi_i] \nabla_x [D \boldsymbol{\xi}^*] \right) + \int_{\partial\Omega} d\Gamma \phi_i \left(\hat{\mathbf{n}} \cdot \nabla_x \right) [\boldsymbol{\xi}^*].
\end{aligned}
$$

A form with reduced order derivatives is derived by eliminating the boundary integral by the no-flux condition (3.19) and then the Galerkin approximations are substituted for the state and co-state, thus producing,

$$
\begin{aligned}
\mathbf{0} \simeq &\sum_{j=1}^{\widehat{M}} \int_\Omega d\mathbf{x} \left(\left(\widehat{\boldsymbol{\xi}}'_j(t) + (A+B)(\widehat{\mathbf{Y}}) \widehat{\boldsymbol{\xi}}_j \right. \right. \\
&\left. \left. + \nabla_Y [A+B](\widehat{\mathbf{Y}}) : (\widehat{\boldsymbol{\xi}}_j (\widehat{\mathbf{Y}})^\top) - R \widehat{\mathbf{Y}}_j \right) \phi_i \phi_j - D \widehat{\boldsymbol{\xi}}_j \nabla_x^\top [\phi_i] \nabla_x [\phi_j] \right),
\end{aligned}
$$

except that the non-linear terms are only symbolically designated by $\widehat{\mathbf{Y}}$, for $i = 1{:}\widehat{M}$. Next by substituting the Galerkin approximation for $\widehat{\mathbf{Y}}$ in the non-linear terms and using the element Galerkin integral notation the compact co-state Galerkin ODEs are obtained:

$$
\begin{aligned}
\mathbf{0} = \sum_{j=1}^{\widehat{M}} \Big(& \mathcal{M}_{i,j} \left(\widehat{\boldsymbol{\xi}}_j'(t) + a_1 \widehat{\xi}_{1,j} \mathbf{e}_1 + a_2 \widehat{\xi}_{2,j} \mathbf{e}_2 \right. \\
& \left. - a_3 \widehat{\xi}_{3,j} \mathbf{e}_3 - R \widehat{\mathbf{Y}}_j(t) \right) - \mathcal{K}_{i,j} D \widehat{\boldsymbol{\xi}}_j(t) \\
& - \sum_{k=1}^{\widehat{M}} \mathcal{T}_{i,j,k} \Big(\frac{2a_1}{k_1} \widehat{Y}_{1,k}(t) \widehat{\xi}_{1,j}(t) \mathbf{e}_1 + \frac{2a_2}{k_2} \widehat{Y}_{2,k}(t) \widehat{\xi}_{2,j}(t) \mathbf{e}_2 \\
& + \alpha_{1,2} \left(\widehat{Y}_{2,k}(t) \mathbf{e}_1 + \widehat{Y}_{1,k}(t) \mathbf{e}_2 \right) \widehat{\xi}_{1,j}(t) \\
& + \kappa_{1,3} \left(\widehat{Y}_{3,k}(t) \mathbf{e}_1 + \widehat{Y}_{1,k}(t) \mathbf{e}_3 \right) \widehat{\xi}_{1,j}(t) \\
& + \alpha_{2,1} \left(\widehat{Y}_{2,k}(t) \mathbf{e}_1 + \widehat{Y}_{1,k}(t) \mathbf{e}_2 \right) \widehat{\xi}_{1,j}(t) \\
& + \kappa_{2,3} \left(\widehat{Y}_{3,k}(t) \mathbf{e}_2 + \widehat{Y}_{2,k}(t) \mathbf{e}_3 \right) \widehat{\xi}_{2,j}(t) \Big) \Big),
\end{aligned}
\tag{4.28}
$$

for $i = 1{:}\widehat{M}$. This Galerkin ODE (4.28) may be computed by the appropriate method using the same Galerkin integral basis coefficients.

5. Spherical finite element test configuration

This section will focus on a simple three–dimensional test configuration that is a sphere of radius R_r. We use a spherical configuration since it resembles the basic structure of the brain better than a cartesian configuration. Transforming the spherical coordinates in space as usual,

$$
(x, y, z) = r(\cos(\theta)\sin(\psi), \sin(\theta)\sin(\psi), \cos(\psi)), \tag{5.29}
$$

where r, θ, ψ are the radius, polar angle and azimuthal angle, respectively. Also

$$
0 \le r \le R_r, \quad 0 \le \theta \le 2\pi \quad \text{and} \quad 0 \le \psi \le \pi.
$$

The brick element grid in spherical coordinates is constructed of M_r radial sectors of width $\Delta r = R_r/M_r$, M_θ polar sectors of width $\Delta\theta = 2\pi/M_\theta$ and M_ψ azimuthal sectors of width $\Delta\psi = \pi/M_\psi$. The global nodal values are given by,

$$
(\theta_{i_e}, \psi_{j_e}, r_{k_e}) = ((i_e - 1)\Delta\theta, (j_e - 1)\Delta\psi, (k_e - 1)\Delta r),
$$

for $i_e = 1 : M_\theta + 1$, $j_e = 1 : M_\psi + 1$ and $k_e = 1 : M_r + 1$. The elements are numbered in (θ, ψ, r) linear priority order like the nodal values,

$$
e_{i_e, j_e, k_e} = i_e + (j_e - 1) \cdot M_\theta + (k_e - 1) \cdot M_\theta \cdot M_\psi,
$$

for $i_e = 1{:}M_\theta$, $j_e = 1{:}M_\psi$ and $k_e = 1{:}M_r$. Within element e_{i_e, j_e, k_e}, the element primary node with local node number $i = 1$ has the same global node number $\bar{k}_e = \{i_e, j_e, k_e\}$ as the element, i.e., $n_{\bar{k}_e, 1} = e_{i_e, j_e, k_e}$, for $i_e = 1{:}M_\psi$, $j_e = 1{:}M_\theta$

and $k_e = 1{:}M_r$. The element local node numbering is $i = 1{:}8$ as shown in the Fig. 1.

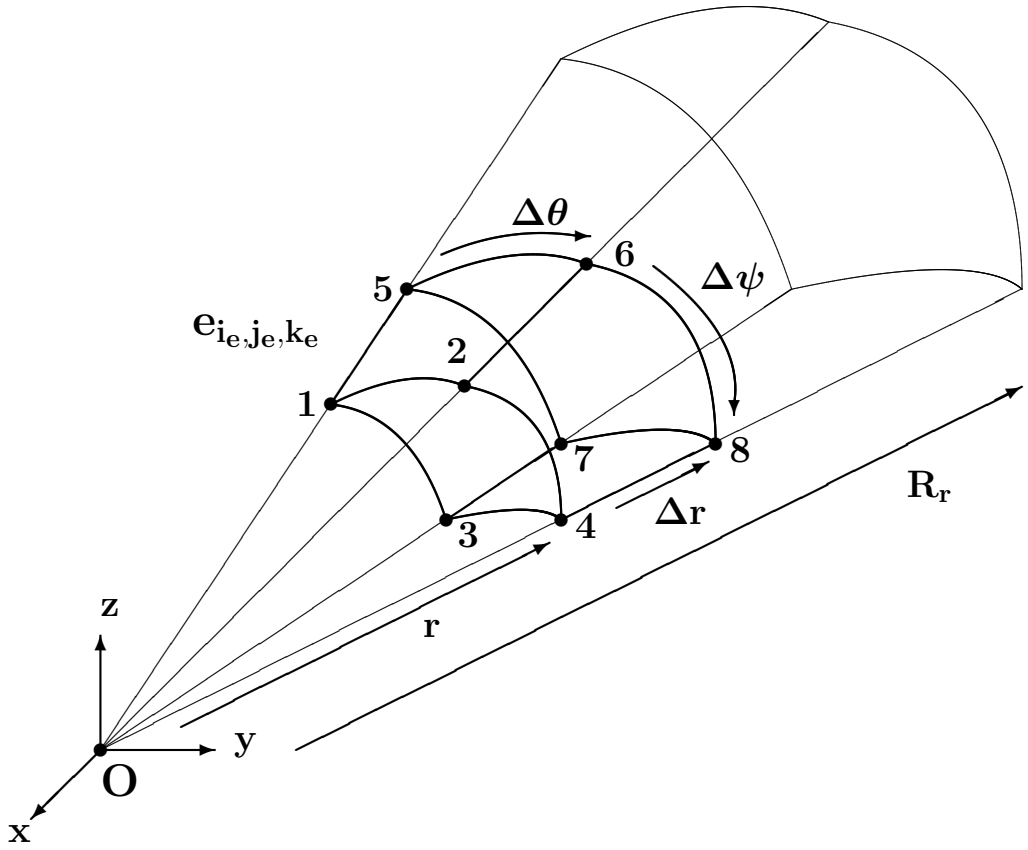

Fig. 1. Local node numbering with $i = 1{:}8$ for general element $\mathbf{e}_{i_e,j_e,k_e}$.

5.1. Tri–linear basis functions

For simplicity, tri–linear basis functions are used for all brick elements in spherical coordinates. The tri–linear bases in three dimensions are constructed from the more basic one–dimensional linear bases with only two nodes labeled 1 and 2

$$\phi_{1,r}^{(k_e)}(r) = \left(\frac{r_{k_e+1} - r}{\Delta r}\right), \qquad \phi_{2,r}^{(k_e)}(r) = \left(\frac{r - r_{k_e}}{\Delta r}\right), \tag{5.30}$$

$$\phi_{1,\theta}^{(i_e)}(\theta) = \left(\frac{\theta_{i_e+1} - \theta}{\Delta\theta}\right), \qquad \phi_{2,\theta}^{(i_e)}(\theta) = \left(\frac{\theta - \theta_{i_e}}{\Delta\theta}\right), \tag{5.31}$$

$$\phi_{1,\psi}^{(j_e)}(\psi) = \left(\frac{\psi_{j_e+1} - \psi}{\Delta\psi}\right), \qquad \phi_{2,\psi}^{(j_e)}(\psi) = \left(\frac{\psi - \psi_{j_e}}{\Delta\psi}\right). \tag{5.32}$$

For the general element,

$$e_{i_e,j_e,k_e} \text{ on } [\theta_{i_e}, \theta_{i_e+1}] \times [\psi_{j_e}, \psi_{j_e+1}] \times [r_{k_e}, r_{k_e+1}]$$

for $i_e = 1{:}M_\theta$, $j_e = 1{:}M_\psi$ and $k_e = 2{:}M_r$ in the $i = 1{:}8$ element node numbering,

$$\phi_1^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{1,r}^{(k_e)}(r) \cdot \phi_{1,\theta}^{(i_e)}(\theta) \cdot \phi_{1,\psi}^{(j_e)}(\psi), \tag{5.33}$$

$$\phi_2^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{1,r}^{(k_e)}(r) \cdot \phi_{2,\theta}^{(i_e)}(\theta) \cdot \phi_{1,\psi}^{(j_e)}(\psi), \tag{5.34}$$

$$\phi_3^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{1,r}^{(k_e)}(r) \cdot \phi_{1,\theta}^{(i_e)}(\theta) \cdot \phi_{2,\psi}^{(j_e)}(\psi), \tag{5.35}$$

$$\phi_4^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{1,r}^{(k_e)}(r) \cdot \phi_{2,\theta}^{(i_e)}(\theta) \cdot \phi_{2,\psi}^{(j_e)}(\psi), \tag{5.36}$$

$$\phi_5^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{2,r}^{(k_e)}(r) \cdot \phi_{1,\theta}^{(i_e)}(\theta) \cdot \phi_{1,\psi}^{(j_e)}(\psi), \tag{5.37}$$

$$\phi_6^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{2,r}^{(k_e)}(r) \cdot \phi_{2,\theta}^{(i_e)}(\theta) \cdot \phi_{1,\psi}^{(j_e)}(\psi), \tag{5.38}$$

$$\phi_7^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{2,r}^{(k_e)}(r) \cdot \phi_{1,\theta}^{(i_e)}(\theta) \cdot \phi_{2,\psi}^{(j_e)}(\psi), \tag{5.39}$$

$$\phi_8^{(i_e,j_e,k_e)}(r,\theta,\psi) = \phi_{2,r}^{(k_e)}(r) \cdot \phi_{2,\theta}^{(i_e)}(\theta) \cdot \phi_{2,\psi}^{(j_e)}(\psi). \tag{5.40}$$

The mapping (5.29) from spherical to cartesian coordinates is not unique since the mapping is many-to-one. The non-unique nodes arise at the *origin* for $r = r_1 = 0$ and any (θ, ψ), or at the *poles* for $\psi = 0, \pi$ and any (r, θ), or at the polar angle line of discontinuity for $\theta = 2\pi$ for any (r, ψ) (really a periodic boundary condition). The non-uniqueness or resulting over-determinism can simply be removed by adding together the appropriate bases in (5.33)-(5.36) corresponding to the same non-unique nodes, using the identities for the one–dimensional bases and their derivatives,

$$\phi_{1,\rho}^{(\ell_e)}(\rho) + \phi_{2,\rho}^{(\ell_e)}(\rho) = 1, \qquad \phi_{1,\rho}^{(\ell_e)\,\prime}(\rho) + \phi_{2,\rho}^{(\ell_e)\,\prime}(\rho) = 0, \tag{5.41}$$

for $\rho = r,\ \theta$ or ψ and $\ell_e = k_e,\ i_e$ or j_e, respectively. Details are given in Appendix A. While it may appear awkward to have to make this adjustment, the disadvantage is out-weighed by the ease of deforming a sphere into a brain geometry rather than deforming a brick or rectangular solid into a brain geometry. As far as we know, there does not exist a description of spherical elements, at least to this detail.

5.2. No-flux boundary condition for spherical Galerkin approximation

Another advantage of spherical coordinates is the ease of imposing the *no-flux boundary condition (BC)* at $r = R_r$, since on the element e_{i_e,j_e,k_e} the i_sth state solution for $i_s = 1{:}3$ is expressed as a preliminary Galerkin approximation,

$$Y_{i_s}^{(i_e,j_e,k_e)}(r,\theta,\psi,t) \simeq \sum_{j=1}^{8} \widetilde{Y}_{i_s,j}^{(i_e,j_e,k_e)}(t) \cdot \phi_j^{(i_e,j_e,k_e)}(r,\theta,\psi), \tag{5.42}$$

so the normal gradient, at boundary element $k_e = M_r$ with $r = R_r$ for local nodes $j = 5{:}8$, reduces to

$$Y_{i_s,r}^{(i_e,j_e,M_r)}(R_r,\theta,\psi,t) \simeq \sum_{j=5}^{8} \left(\widetilde{Y}_{i_s,j}^{(i_e,j_e,M_r)}(t) - \widetilde{Y}_{i_s,j-4}^{(i_e,j_e,M_r)}(t)\right) \cdot \phi_{j,r}^{(i_e,j_e,M_r)}(R_r,\theta,\psi),$$

and then

$$\widetilde{Y}_{i_s,j}^{(i_e,j_e,M_r)}(t) = \widetilde{Y}_{i_s,j-4}^{(i_e,j_e,M_r)}(t) \tag{5.43}$$

for $j = 5{:}8$ and arbitrary (θ,ψ), if $Y_{i_s,r}^{(i_e,j_e,M_r)}(R_r,\theta,\psi,t) = 0$ for no-flux.

Here the symmetries and asymmetries of the bases functions (5.30-5.40) and derivatives have been used. This version of the no-flux condition is much better and simpler to use than the the one we had previously used [9], even when dealing with a deformed sphere in the form of a brain case.

5.3. Local node spherical element matrix coefficients

The element matrices for local node numbers $i,j,k = 1{:}8$ are

$$\begin{aligned}\mathcal{M}_{i,j}^{(i_e,j_e,k_e)} = &\int_{\theta_{i_e}}^{\theta_{i_e+1}} d\theta \int_{\psi_{j_e}}^{\psi_{j_e+1}} d\psi \int_{r_{k_e}}^{r_{k_e+1}} dr\ r^2 \sin(\psi) \\ &\left(\phi_i^{(i_e,j_e,k_e)} \phi_j^{(i_e,j_e,k_e)}\right)(r,\theta,\psi),\end{aligned} \tag{5.44}$$

$$\begin{aligned}\mathcal{K}_{i,j}^{(i_e,j_e,k_e)} = &\int_{\theta_{i_e}}^{\theta_{i_e+1}} d\theta \int_{\psi_{j_e}}^{\psi_{j_e+1}} d\psi \int_{r_{k_e}}^{r_{k_e+1}} dr\ r^2 \sin(\psi) \\ &\left(\phi_{i,r}^{(i_e,j_e,k_e)} \phi_{j,r}^{(i_e,j_e,k_e)} + \frac{1}{r^2} \phi_{i,\psi}^{(i_e,j_e,k_e)} \phi_{j,\psi}^{(i_e,j_e,k_e)} \right. \\ &\left. + \frac{1}{r^2 \sin^2(\psi)} \phi_{i,\theta}^{(i_e,j_e,k_e)} \phi_{j,\theta}^{(i_e,j_e,k_e)}\right)(r,\theta,\psi),\end{aligned} \tag{5.45}$$

$$\begin{aligned}\mathcal{T}_{i,j,k}^{(i_e,j_e,k_e)} = &\int_{\theta_{i_e}}^{\theta_{i_e+1}} d\theta \int_{\psi_{j_e}}^{\psi_{j_e+1}} d\psi \int_{r_{k_e}}^{r_{k_e+1}} dr\ r^2 \sin(\psi) \\ &\left(\phi_i^{(i_e,j_e,k_e)} \phi_j^{(i_e,j_e,k_e)} \phi_k^{(i_e,j_e,k_e)}\right)(r,\theta,\psi),\end{aligned} \tag{5.46}$$

where in (5.45) $\phi_{i,\rho}^{(i_e,j_e,k_e)}$ denotes the partial derivative of $\phi_i^{(i_e,j_e,k_e)}$ with respect to generic spherical coordinate ρ. In the stiffness matrix (5.45), the mapping singularities of the gradient lead to reciprocal factors in r and $\sin(\psi)$, but

the r factors are simply cancelled by the Jacobian $r^2 \sin(\psi)$ and leave uncancelled $\sin(\psi)$–denominators in the θ–derivative term. However, these $\sin(\psi)$–denominators are completely eliminated in later analysis upon eliminating non-unique nodes by combining terms and associated ODEs.

6. Forward–backward computational iteration method

In this section, we present a double shot, forward–backward iteration algorithm This method is similar to the multiple shooting method of Hackbusch [20] used for solving parabolic PDEs with opposite orientation or to what Gunzburger [18] calls the one-shot method. Since the shots are forward and backward, we prefer to call them double shots. See Gunzburger [18] for more rigorous justification with Sobolev spaces in the more general abstract case. However the model here is quite concrete and much more appropriate for computation. The numerical implementation of the algorithm is akin to our previous work [8], except that here we use it solve for Galerkin ODEs instead of PDEs. The only problem which arises are the non-unique degeneracies of states and co-states due to aliases and boundary conditions. These are eliminated so that there are only $\widehat{M}$ linearly independent Galerkin coefficients, $\widehat{\mathbf{Y}}_{\hat{k}}(t)$ in ODE (4.26), $\widehat{\boldsymbol{\xi}}_{\hat{k}}(t)$ in ODE (4.28) and control $\widehat{\mathbf{U}}_{\hat{k}}(t)$. This non-uniqueness elimination keeps the system of ODEs from being over-determined, preserving the symmetry of the mass and other coefficient arrays, and eliminating the $1/\sin(\psi)$ singularity in the stiffness integrals by virtue of identities of the one–dimensional basis and their elements (5.41). A summary of the degeneracy removal by combining unknowns and the corresponding equations are given in the Appendix A and Appendix B (also see Chakrabarty's thesis [7]). This method uses a combination of Crank–Nicolson and prediction-correction methods developed in [21–24] for solving high dimensional stochastic control problems on supercomputers. The general method can handle both implicit and non-linear terms provided the modified parabolic mesh ratio is sufficiently small. We would like to point out here that we will use the symbol δ to indicate the iteration number or the number of double shots and γ for the correction number in the predictor–corrector step.

These methods have been extensively and successively developed by one of the authors [21–23] for a wide variety of problems in the biosciences, including non-linear ones. The authors considered Newton-like methods, but thought they were not practical due to the complexity of the non-linear terms and the computational cost of Hessian terms, whether exact or approximate. Moreover such schemes are unstable unless one is very close to the optimal and away from singular points. In addition, general fixed point method, other than fast Newton methods, are usually not considered for serious computational problems, although often used in purely theoretical treatment. We have found the

predictor–corrector approach most useful and dependable for a wide range of non-linear applications in optimal control, as long as the effective parabolic mesh ratio is kept sufficiently small. It would be naive to try to use a simple fixed point method on this. Our computational convergence has been regulated by our extrapolated, predictor–corrector methods and mesh-ratio conditions, else we would find divergence or very slow computational convergence. In fact convergence for this problem was acheived with three double shot iterations, showing that the method is computationally robust. Also the two-shot forward–backward iteration works because there is prediction and correction for the non-linear problem, as well as global corrections with each forward shot and backward shot, providing the mesh ratio is sufficiently small.

6.1. Forward step

The first step ($\delta = 1$) would be to make a guess about the control $U_3^*(\mathbf{x}, t) \simeq U_3^{(1)}(\mathbf{x}, t)$. We substitute it into the forward state equations and use the finite element method to solve for the state $\mathbf{Y}^*(\mathbf{x}, t) \simeq \mathbf{Y}^{(1)}(\mathbf{x}, t)$ for $t > 0$. Initially, $\mathbf{Y}^*(\mathbf{x}, 0) = \mathbf{Y}_0(\mathbf{x})$. For simplicity, the forward ODE (4.26) for the degeneracy removed i_sth-state $\widehat{Y}_{i_s,\hat{\jmath}}$ at node $\hat{\jmath}$ can be written symbolically, for $i_s = 1{:}3$ and nodes $\hat{\jmath} = 1{:}\widehat{M}$, as

$$\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \widehat{Y}'_{i_s,\hat{k}}(t) = \sum_{j_s=1}^{3} \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{A}_{i_s,j_s,\hat{\jmath},\hat{k}}(\widehat{\mathcal{Y}}(t)) \cdot \widehat{Y}_{j_s,\hat{k}}(t) - \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \widehat{U}_{\hat{k}}(t),$$

where $\mathcal{A}_{i_s,j_s,\hat{\jmath},\hat{k}}(\widehat{\mathcal{Y}}(t))$ symbolically represents the right-hand-side matrices including non-linear terms and $\widehat{\mathcal{Y}}(t)$ represents the combined states and nodes array. The essential set–up for a general Crank–Nicolson method is to use the midpoint approximation on the integral form of the differential equations followed by an average approximation of the midpoint values, producing from the state ODE for $\widehat{Y}_{i_s,\hat{k},\ell+1}$ at time $t_\ell = (\ell - 1) * \Delta t$ with $\ell = 1{:}N_t$, $\hat{\imath}, \hat{\jmath} = 1{:}\widehat{M}$ and $i_s = 1{:}3$,

$$\begin{aligned}\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \left(\widehat{Y}_{i_s,\hat{k},\ell+1} - \widehat{Y}_{i_s,\hat{k},\ell} \right) = &+\Delta t \sum_{j_s=1}^{3} \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{A}_{i_s,j_s,\hat{\jmath},\hat{k},\ell+0.5} \cdot \widehat{Y}_{j_s,\hat{k},\ell+0.5} \\ &-\Delta t \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \widehat{U}_{\hat{k},\ell+0.5},\end{aligned}$$

where the average approximation at the midpoint is

$$\widehat{Y}_{j_s,\hat{k},\ell+0.5} \simeq 0.5 * \left(\widehat{Y}_{i_s,\hat{k},\ell+1} + \widehat{Y}_{i_s,\hat{k},\ell}\right),$$
$$\widehat{U}_{\hat{k},\ell+0.5} \simeq 0.5 * \left(\widehat{U}_{\hat{k},\ell+1} + \widehat{U}_{\hat{k},\ell}\right),$$

and is compatible with the midpoint approximation in accuracy. The approximation $\mathcal{A}_{j_s,\hat{\jmath},\hat{k},\ell+0.5}$ is similarly computed.

6.2. Backward step

In the second shot of the double shot algorithm, the final condition (3.20),

$$\boldsymbol{\xi}^{(\delta)}(\mathbf{x}, t_f) \simeq -Q\widehat{\mathbf{Y}}^{(\delta)}(\mathbf{x}, t_f) = -Q\sum_{\hat{k}=1}^{M} \widehat{\mathbf{Y}}_{\hat{k}}^{(\delta)}(t_f)\phi_{\hat{k}}(\mathbf{x}),$$

for $\delta = 1{:}L$ double shots, is used to start the backward co-state solution. A similar backward ODE can be written for the co-state $\widehat{\xi}_{i_s\hat{k}}(t)$ with the state replacing the role of the control in the source while remaining in the general non-linear coefficient $\mathcal{B}_{i_s,j_s,\hat{\jmath},\hat{k}}(\widehat{\mathcal{Y}}(t))$ as in (4.28), i.e.,

$$\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}}\widehat{\xi}'_{i_s,\hat{k}}(t) = -\sum_{j_s=1}^{3}\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{B}_{i_s,j_s,\hat{\jmath},\hat{k}}(\widehat{\mathcal{Y}}(t)) \cdot \widehat{\xi}_{j_s,\hat{k}}(t) + \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} R\widehat{Y}_{i_s\hat{k}}(t).$$

In an manner analogous to the state equations, except for backward integration, the co-state $\widehat{\xi}_{i_s,\hat{k},\ell-1}$ satisfies

$$\begin{aligned}\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \left(\widehat{\xi}_{i_s,\hat{k},\ell-1} - \widehat{\xi}_{i_s,\hat{k},\ell}\right) = &+\Delta t \sum_{j_s=1}^{3}\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{B}_{i_s,j_s,\hat{\jmath},\hat{k},\ell-0.5} \cdot \widehat{\xi}_{j_s,\hat{k},\ell-0.5} \\ &-\Delta t \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} R\widehat{Y}_{\hat{k},\ell-0.5},\end{aligned}$$

where the average approximation at the midpoint is

$$\widehat{\xi}_{j_s,\hat{k},\ell-0.5} \simeq 0.5 * \left(\widehat{\xi}_{i_s,\hat{k},\ell-1} + \widehat{\xi}_{i_s,\hat{k},\ell}\right),$$

6.3. Predictor–corrector step

While we are running the iterations from $\delta = 1{:}L$, within each iteration the predictor–corrector procedure is used for $\gamma = 1{:}\gamma_{\max}$to handle the implicit and non-linear terms. For the sake of brevity we will provide an explanation only for the case of state equations. The zeroth corrector, given the final correction $\widehat{Y}_{i_s,\hat{\jmath},\ell}$ at time stage i is the predictor,

$$\mathrm{YC}^{(0)}_{i_s,\hat{\jmath},\ell+1} = \widehat{Y}_{i_s,\hat{\jmath},\ell},$$

for $\ell = 1{:}N_t$, $\hat{\imath},\hat{\jmath} = 1{:}\widehat{M}$ and $i_s = 1{:}3$. This initialization permits finding the $(\gamma+1)$th correction $\mathrm{YC}^{(\gamma+1)}_{\hat{\jmath},\ell+1}$ from

$$\begin{aligned}\sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \left(\mathrm{YC}^{(\gamma+1)}_{i_s,\hat{k},\ell+1} - \mathrm{YC}^{(\gamma)}_{i_s,\hat{k},\ell}\right) = &+\Delta t \sum_{j_s=1}^{3} \sum_{\hat{k}=1}^{\widehat{M}} \mathrm{AC}^{(\gamma)}_{j_s,\hat{\jmath},\hat{k},\ell+0.5} \cdot \mathrm{YC}^{(\gamma)}_{j_s,\hat{k},\ell+0.5} \\ &-\Delta t \sum_{\hat{k}=1}^{\widehat{M}} \mathcal{M}_{\hat{\jmath},\hat{k}} \mathrm{UC}^{(\gamma)}_{\hat{k},\ell+0.5},\end{aligned}$$

for $\ell = 1{:}N_t$, $\hat{\imath},\hat{\jmath} = 1{:}\widehat{M}$ and $i_s = 1{:}3$, where $YC^{(\gamma)}_{j_s,\hat{k},\ell+0.5}$ and other midpoint terms are evaluated as before by averaging. The final correction at the final time of the *state-shot forward iteration*, $\widehat{Y}_{\hat{\jmath},N_t+1}$ yields the starting or final-time condition for $\widehat{\xi}_{i_s,\hat{k},N_t+1}$ using (3.20). Then the final correction at the initial time of the *costate-shot backward iteration* produces the initial control condition when $i_s = 3$ which is used to begin another double shot.

6.4. Regular optimal control step

For each completed double shot for $\delta = 1:L$, the co-state approximation $\widehat{\boldsymbol{\xi}}^{(\delta)}(\mathbf{x},t) = \sum_{\hat{k}=1}^{\widehat{M}} \widehat{\boldsymbol{\xi}}^{(\delta)}_{\hat{k}}(t)\phi_{\hat{k}}(\mathbf{x})$ is used to determine the *regular optimal control* (3.17) updated value third component, with $\widehat{\xi}^{(\delta)}_3 = \mathbf{e}_3 \cdot \widehat{\boldsymbol{\xi}}^{(\delta)}(\mathbf{x},t)$,

$$\widehat{U}^{(\delta+1)}_3(\mathbf{x},t) = U_{0,3}(\mathbf{x},t) + \widehat{\xi}^{(\delta)}_3(\mathbf{x},t)/s_3.$$

6.5. Convergence criteria step

This process is repeated for $\delta = 2{:}L$ double shot iterations until a convergence criterion for sufficiently large L is reached, e.g., the relative criterion for the control,

$$\left|\left|U_3^{(\delta)}(\mathbf{x},t) - U_3^{(\delta-1)}(\mathbf{x},t)\right|\right| < \mathrm{tol}_u \left|\left|U_3^{(\delta-1)}(\mathbf{x},t)\right|\right|,$$

and,

$$\left|\left|\mathbf{Y}^{(\delta)}(\mathbf{x},t) - \mathbf{Y}^{(\delta-1)}(\mathbf{x},t)\right|\right| < \mathrm{tol}_y \left|\left|\mathbf{Y}^{(\delta-1)}(\mathbf{x},t)\right|\right|,$$

for $\delta = 2{:}L$ until satisfied, provided $||U_3^{(\delta-1)}(\mathbf{x},t)|| \neq 0$ and $||\mathbf{Y}^{(\delta-1)}(\mathbf{x},t)|| \neq 0$, where $\mathrm{tol}_u > 0$ and $\mathrm{tol}_y > 0$ are some prescribed tolerances. The algorithm was implemented in MATLAB™ on a desktop computer.

For the benefit of the readers whose interest is primarily in the mathematical flavor, we have included a compact matrix-vector notation in the Appendix.

7. Computational results

The data for the numerical parameters are drawn from various sources including Wang et al. [38,39], Swanson [37] and Murray [35], while unavailable parameters were estimated (see Table in Appendix C). The initial tumor spread and drug concentration are assumed to be conically distributed with scale and state dependent means and weights. Maple™ was used to exactly evaluate the integrals of the element matrices off-line. A sample history of the optimal relative tumor density $Y_1^*(r,\theta,\psi,t)$ for r over [0, 5] in centimeters at fixed $(\theta,\psi) = (\pi,\pi/2)$ in radians and at quartiles in time of a 5 day treatment schedule is given in Fig. 2. The initial tumor peak is at $(r,\theta,\psi) = (2.5,\pi,\pi/2)$. This test case shows significant reduction of tumor density over the treatment schedule. Fig. 3 shows the time-dependence of the optimal drug bolus density, with a significant decline coming at the end of the treatment period. The double shot computations took 7.3 h on a Macintosh G5 dual 2GHz, using three double shots and a maximum of two corrections per shot. The finite element mesh set-up takes less than one second. The double shot time is much longer than our prior reported results [10], mainly since the control quadratic cost coefficient $s_U = s_3$ was significantly reduced to get more satisfactory resolution for the drug delivery control $U = U_3$ decline near the final time. Nevertheless, the grid is somewhat coarse due to the high computational complexity of the

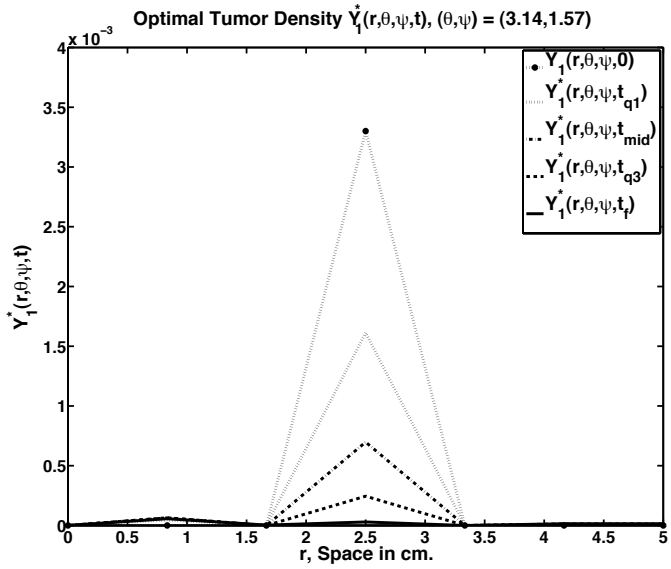

Fig. 2. The optimal, relative tumor density $Y_1^*(r, \theta, \psi, t)$ versus r at the quartile times, as a cross-section at $(\theta, \psi) = (\pi, \pi/2)$ radians, with the initial tumor density peak location at $(r, \theta, \psi) = (2.5, \pi, \pi/2)$. The grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$.

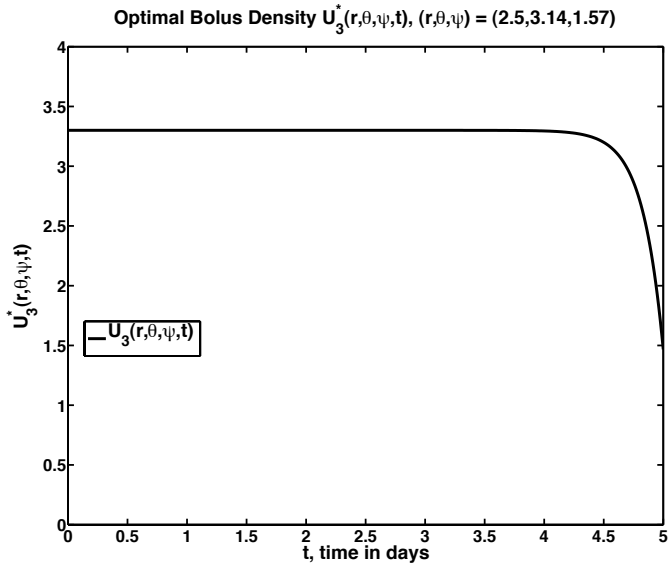

Fig. 3. On the right is the optimal drug bolus density $U_3^*(r, \theta, \psi, t)$ versus t at $(r, \theta, \psi) = (2.5, \pi, \pi/2)$. The grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$.

the numerical problem, so more high performance computing [22] would be needed for greater grid refinement.

A more detailed presentation of the initial and final tumor densities are presented for $t = t_0 = 0$ in Fig. 4 and $t = t_f = 5$ days is given in Fig. 5, both over the larger (r, θ) plane section with fixed $\psi = \pi/2$. They show that the final peak value is small and the tumor has not spread too much through the rest of

the plane in spherical coordinates, although somewhat in the θ–direction but not much in the r–direction. The final peak to peak tumor density reduction is 40.6%.

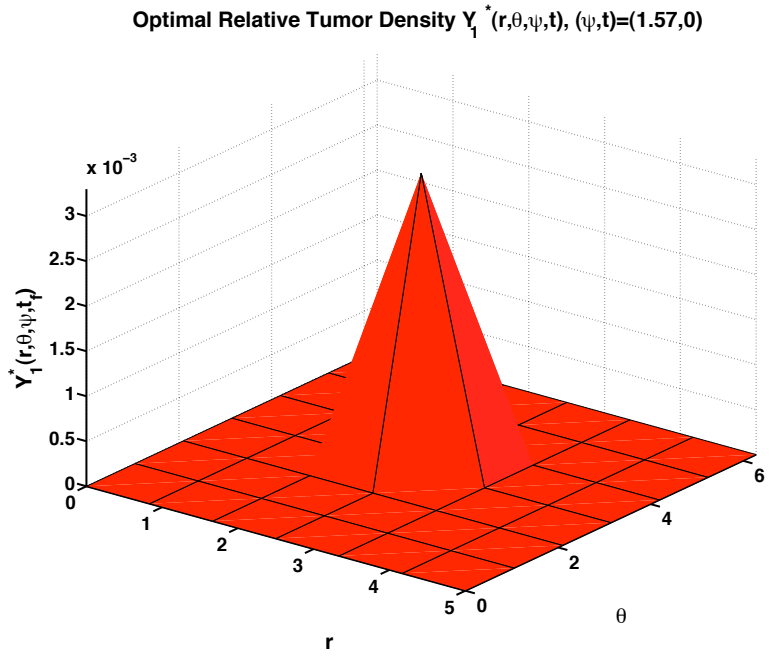

Fig. 4. The initial relative tumor density $Y_1^*(r, \theta, \psi, t)$, when $t = 0$ over (r, θ) plane $\psi = \pi/2$. The FEM grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$.

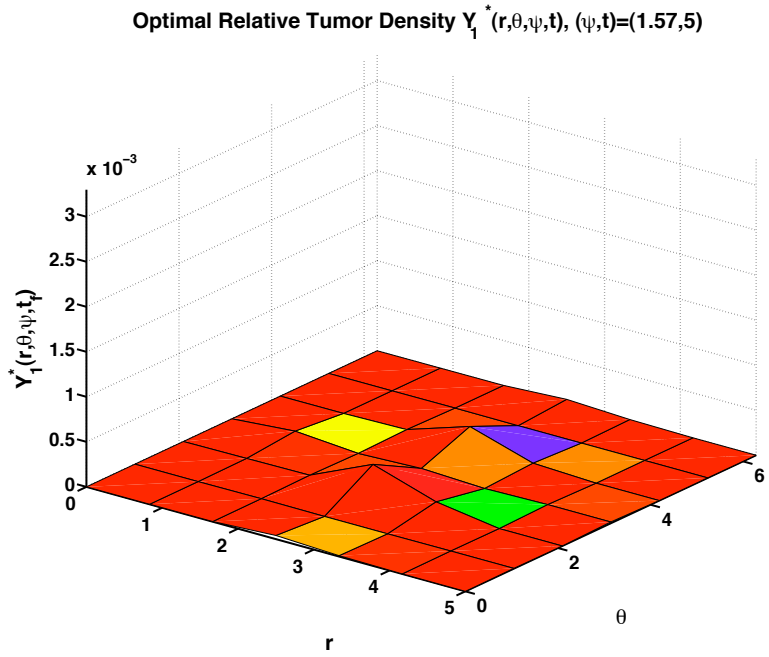

Fig. 5. The final relative tumor density $Y_1^*(r, \theta, \psi, t)$, when $t = 5$, over (r, θ) plane $\psi = \pi/2$. The FEM grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$.

The total decrease of the Galerkin integral of the optimal tumor density $\text{Int}[Y_1^*](t)$ relative to the initial state $\text{Int}[Y_1^*](0)$ over time is displayed in Fig. 6, showing a tapering off of the decay of the tumor as the end of the treatment period is reached. The Galerkin integral is the integral of the Galerkin approximation (5.42) of $Y_1^*(r, \theta, \psi, t)$ over the sphere of radius $r = R_r$ in spherical coordinates (r, θ, ψ).

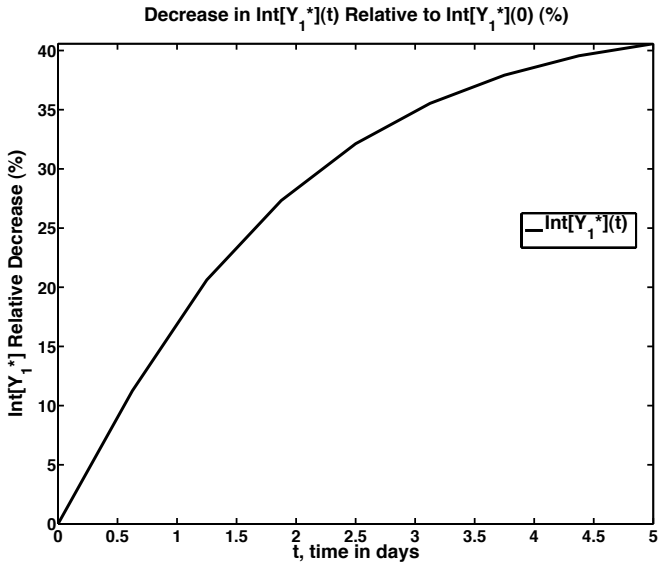

Fig. 6. The total decrease in the integral of the optimal tumor density $\text{Int}[Y_1^*](t)$ relative to the initial value $\text{Int}[Y_1^*](0)$ versus t at the quartile times. The grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$ and $wU_C = 8.0$.

There are some parameters associated with the drug delivery that will likely vary with the treatment, the particular drug and the patient condition. Hence, in Fig. 7 the sensitivity of the relative total decrease in the Galerkin integral of the optimal tumor density, $\text{Int}[Y_1^*](t)$, to the drug kill rate $\kappa_{T,C}$ when at end of the treatment, $t = 5$ days. Then, the accompanying Fig. 8 shows sensitivity of the relative $\text{Int}[Y_1^*](t)$ to the weighting parameter, wU_C, for the distribution of the threshold control variable $U_{0,3}$ and the sensitivity results are very similar to that of Fig. 7 for the unrelated parameter $\kappa_{T,C}$. Both sensitivity plots peak when the parameters have a common value, e.g., $\kappa_{T,C} = 8.0 = wU_C$ here.

8. Conclusion and future directions

The theory of Galerkin finite elements is used to develop approximations to the distributed parameter optimal control problem of cancer drug delivery to the brain governed by a coupled set of three reaction–diffusion PDEs in three space dimensions. The three state variables are the tumor cell density, nor-

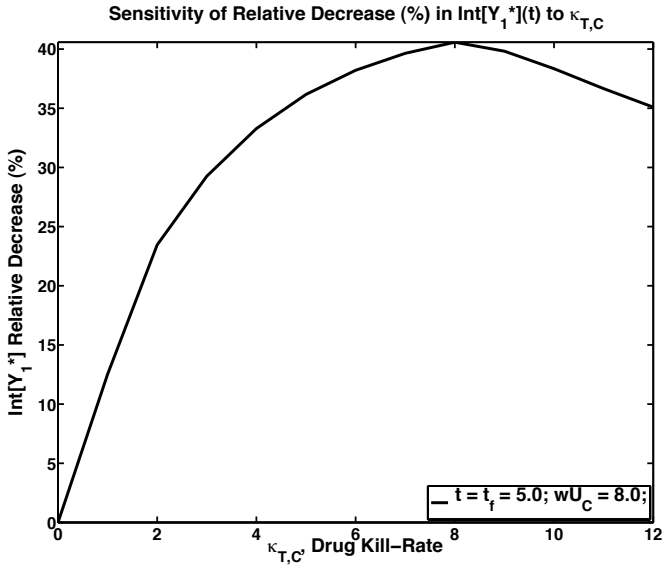

Fig. 7. The final optimal, relative tumor density $Y_1^*(r, \theta, \psi, t)$ versus the drug kill rate $\kappa_{T,C}$ is shown. The grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$ and $wU_C = 8.0$.

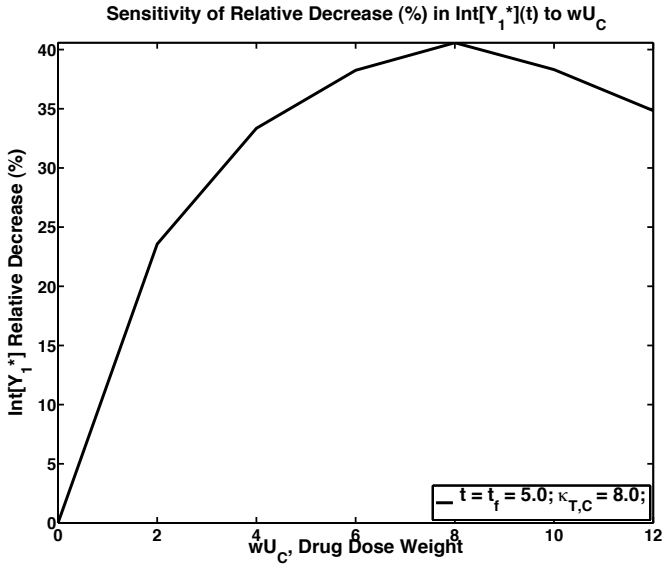

Fig. 8. The final optimal, relative tumor density $Y_1^*(r, \theta, \psi, t)$ versus the threshold weighting wU_C is shown. The grid size is $(M_r, M_\theta, M_\psi) = (6, 6, 6)$ and $\kappa_{T,C} = 8.0$.

mal cell density and cancer drug concentration. While the tumor and normal cells are highly coupled through competitive interactions, the concentration is directly controlled by the drug delivery control rate. The optimally controlled distributed parameter system is derived by a straight-forward calculus of variations technique without resort to an very abstract formulation, and that should be useful in other similar scientific or engineering applications. The resulting system of six optimal PDEs is reduced by Galerkin approximations of the state, co-state and control vectors to a system of six ODEs in time

with three fundamental element integral coefficient forms: the mass, the stiffness and non-linear coefficients. The finite element configuration is given for a spherical geometry that can be used to test the optimal drug delivery computations. This finite element configuration will be more amenable to complex brain structures and three-dimensional geometries than the finite difference method and low dimension of our earlier work.

The natural question that one might ask is what are the biomedical implications of the work. Firstly the model tries to capture the dynamics of interaction of gliomas and normal cells in the human brain in conjunction with the effects of drugs. To the best of our knowledge this is probably the first such work which tries to focus on this aspect taking into account the augmentation of the treatment and it's ramification due to the control. However the most crucial factor that would be of interest to researchers in biomedicine is that we try to make the problem more practical in terms of the geometry. Unlike the treatment of many mathematical biology problem such as viral dynamics and vector borne diseases, where the spatial factors are negligible, the spatial geometry is very very important for this problem. Hence our choice of a sphere (which is a smooth structure that is a characteristic of the human brain and a rough model of the resected tumor space). We see this work as a first step to extend this to more realistic regular structure like curvilinear coordinates and eventually to the irregular human brain structure. The most important fact here is that most finite element formulation in literature tend to be in cartesian coordinates in lower dimension, which really cannot be applied to real world which is three–dimensional. We believe that biomedical teams working on aspects like image processing of human brain could use this work as a starting point to design a simulation where one can predict the optimal dosage of the drug in–silico. Because of the optimization problem the optimal drug dosage would ensure that the sole focus is not on reduction of the tumor burden but also to enhance the quality of life of the patient in terms of minimal side effects. We do however acknowledge that this challenging task would involve a big team comprising of bioengineers, oncologist, computer scientists, statisticians and mathematicians.

8.1. Future directions

One issue that is worth studying is to relate our data to the ones obtained from MRI/CT scan as outlined in Clatz et al. [11] or using translated /wrapped version given by Gee et al. [15]. Some of these data can be obtained from nice computerized brain atlas like the one by Greitz et al. [17]. Following Davis et al. [12] the mathematical representation of mapping elastic deformation can be done between Ω_R (our results) with a target domain Ω^*_{MC} as *defined* by the MRI/CT scan. A couple of other different approaches to the problem include:

- Application to general curvilinear coordinates for general brain geometries;
- Application to heterogeneous brain structures such as spinal fluid cavities, variable brain matter, vascular system and the blood brain barrier.

Acknowledgments

The authors gratefully acknowledge Dr. Herbert H. Engelhard, Chief of Neuro-Oncology Section of the University of Illinois Medical Center for suggesting the problem and Professor William D. O'Neill of Bioengineering at the University of Illinois at Chicago for putting us in contact. The authors are grateful to two reviewers whose comments helped improve both the biological and mathematical quality of the paper. The first author did part of the work during his tenure as a post doctoral Visiting Fellow at the Tata Institute of Fundamental research (TIFR) Centre in Bangalore. The work was supported by the National Science Foundation under Grant DMS-02-07081. The content of this material is that of the authors and does not necessarily reflect the views of the National Science Foundation.

References

[1] D. Acquafredda, Brain Tumor Treatments: BCNU (Carmustine), Musella Foundation, `http://virtualtrials.com/bcnu2.cfm`

[2] N. U. Ahmed and K. L. Teo, Optimal Control of Distributed Parameter Systems, Elsevier Science Ltd., New York, 1981.

[3] R. P. Araujo, D. L. S. McElwain, A history of the study of solid tumour growth: The contribution of mathematical modeling, Bull. Math. Biol. 66 (2004) 1039-1091.

[4] H. T. Banks and K. Kunish, Estimation Techniques for Distributed Parameter Systems, Birkhauser, Boston 1989.

[5] L. T. Baxter, R. K. Jain, Transport of fluid and macromolecules in tumors I. Role of interstitial pressure and convection, Microvascular Research 37 (1989) 77-104.

[6] M. Berggren, Numerical solution of a flow-control problem: Vorticity reduction by dynamic boundary action, SIAM J. Sci. Comp. 19 (1998) , 829860.

[7] S. P. Chakrabarty, Optimal Control of Drug Delivey to Brain Tumors Using a Distributed Parameters Deterministic Model, PhD Thesis, University of Illinois at Chicago, 2006.

[8] S. P. Chakrabarty, F. B. Hanson, Optimal control of drug delivery to brain tumors for a distributed parameters model, in Proc. American Control Conf. (2005) 973-978.

[9] S. P. Chakrabarty, F. B. Hanson, Optimal control of drug delivery to brain tumors for a test of PDE driven models using the Galerkin finite element method, in Proc. Conf. Decision and Control and European Control Conf. (2005) 1613-1618.

[10] S. P. Chakrabarty, F. B. Hanson, Cancer drug delivery in three dimensions for a distributed parameter control model using finite elements, in Proc. Conf. Decision and Control (2006) 2088-2093.

[11] O. Clatz, M. Sermesant, P-Y. Bondiau, H. Delingette, S. K. Warfield, G. Malandain, N. Ayache, Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation, IEEE Trans. Medical Imaging 24(10) (2005) 1334-1346.

[12] M. H. Davis, A. Khotanzad, D. P. Flamig, S. E. Harms, A physics-based coordinate transformation for 3-D image matching, IEEE Trans. Medical Imaging 16(3) (1997) 317-328.

[13] H. H. Engelhard, Brain tumors and the blood-brain barrier, in Neuro-Oncology, The Essentials, Thieme Medical Publishers, New York, NY, 2000, pp. 49-53.

[14] R. A. Gatenby, E. T. Gawlinski, A reaction–diffusion model of cancer invasion, Cancer Research 56 (1996) 5745-5753.

[15] J. C. Gee, D. R. Haynor, M.Reivich, R. Bajcsy, Finite element approach to warping of brain images, in Medical Imaging 1994: Image Processing, 1994, 11 pages.

[16] J. H. Goldie, A. J. Coldman, Drug Resistance in Cancer: Mechanisms and Models, Cambridge University Press, Cambridge, UK, 1998.

[17] T. Greitz, C. Bohm. S. Holte, L. Eriksson, A computerized brain atlas: Construction, anatomical content, and some applications, J. Computer Assisted Tomography 15(1) (1991) 26-38.

[18] M. D. Gunzburger, Perspectives in Flow Control and Optimization, Series on Advances in Design and Control, vol. DC05, SIAM Books, Philadelphia, PA, 2003.

[19] R. Haberman, Elementary Applied Partial Differential Equations with Fourier Series and Boundary Value Problems, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[20] W. Hackbusch, A numerical method for solving parabolic equations with opposite orientations, Computing 20 (1978) 229-240.

[21] F. B. Hanson, Computational dynamic programming on a vector processor, IEEE Trans. Automatic Control 36(4) (1991) 507-511.

[22] F. B. Hanson, Techniques in computational stochastic dynamic programming, in Control and Dynamic Systems: Advances in Theory and Applications, vol. 76, Academic Press, New York, NY, 1996, pp. 103-162.

[23] F. B. Hanson, Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis, and Computation, Series on Advances in Design and Control, vol. DC13, SIAM Books, Philadelphia, PA, 2007.

[24] F. B. Hanson, K. Naimipour, Convergence of numerical method for multistate stochastic dynamic progamming, in Proc. Int. Fed. of Auto. Control 12th World Congress 9 (1993) 501-504.

[25] M. Heinkenschloss, Numerical solution of implicitly constrained optimization problems, Tech. Rep. TR0805, Department of Computational and Applied Mathematics, Rice University, Houston, TX 770051892, 2008.

[26] M. Hinze, K. Kunisch, Second order methods for optimal control of timedependent fluid flow, SIAM J. Control and Optimization 40 (2001) 925946.

[27] H. R. Joshi,"Optimal Control Problems in PDE and ODE Systems", PhD Thesis, The University of Tennessee, Knoxville, 2002.

[28] C. T. Kelley, Iterative Methods for Optimization, SIAM, Philadelphia, 1999.

[29] F. A. Khodja, C. Dupaix and K. Hamdache, "Optimal Control Problem for a Nonlinear Parabolic System Modeling the Drug Delivery to Brain Tumor," V Workshop on Partial Differential Equations, Campus of the Federal University of Rio de Janeiro Rio de Janeiro, 2006.

[30] F. A. Khodja, A. Benabdallah and C. Dupaixa, "Null-controllability of some reaction–diffusion systems with one control force" Journal of Mathematical Analysis and Applications, 320(2), (2006) pp. 928–943.

[31] D. E. Kirk, Optimal Control Theory: An Introduction, Prentice-Hall, Englewood Cliffs, NJ, 1970 (reprinted by Dover Publications, Mineola, NY, 2004.).

[32] J.-L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer Verlag, Berlin, Heidelberg, New York, 1971.

[33] J. Mansuri, The Modelling of Tumor Growth Using Reaction-Diffusion Equations, MSc Thesis, Oxford University, 2002.

[34] J. D. Murray, Mathematical Biology I: An Introduction, Springer, New York, NY, 2002.

[35] J. D. Murray, Mathematical Biology II: Spatial Models and Biomedical Applications, Springer, New York, NY, 2003.

[36] G. Sewell, The Numerical Solution of Ordinary and Partial Differential Equations, Academic Press, New York, NY, 1988.

[37] K. R. Swanson, Mathematical Modeling of the Growth and Control of Tumors, PhD Thesis, University of Washington, 1999.

[38] C. H. Wang, J. Li, Three dimensional simulation of IgG delivery to tumors, Chemical Engineering Science 53 (1998) 3579-3600.

[39] C. H. Wang, J. Li, C. S. Teo, T. Lee, The delivery of BCNU to brain tumors, J. Controlled Release 61 (1999) 21-41.

[40] J. J. Westman, B. R. Fabijonas, D. L. Kern, F. B. Hanson, Probabilistic rate compartment cancer model: Alternate versus traditional chemotherapy scheduling, in Stochastic Theory and Control, in Proc. Workshop held in Lawrence, Kansas, October 18-20, 2001, Lecture Notes in Control and Information Sciences, vol. 280, B. Pasik-Duncan (Editor), Springer-Verlag, New York, NY, 2002, pp. 491-506.

[41] Wikipedia, Vascular Endothelial Growth Factor (VEGF) `http://en.wikipedia.org/wiki/Vascular_endothelial_growth_factor`

[42] D. E. Woodward, J. Cook, P. Tracqui, G. C. Cruywagen, J. D. Murray, E. C. Alvord, Jr., A mathematical model of glioma growth: The effect of extent of surgical resection, Cell Proliferation 29 (1996) 269-288.

Appendix A. Spherical degeneracy removed

(1) *Origin*: $r = 0$, $k_e = 1$, $(x, y, z) = (0, 0, 0)$ for $i_e = 1 : M_\theta$, $j_e = 1 : M_\psi$, $i = 1{:}4$ (see Fig.1 to see why only local nodes $i = 1{:}4$ are involved), so

$$Y_{i_s,i}^{(i_e,j_e,1)}(t) = Y_{i_s,1}^{(1,1,1)}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t),$$

where the initial count is $\hat{\jmath} = 1$ of an independent set of unknowns; then the ODEs must be combined corresponding to the combined aliased unknowns, but this is only illustrated on the central derivative terms as

$$\mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv Y_{i_s,1}^{(1,1,1)\,\prime}(t) \cdot \sum_{i_e=1}^{M_\theta} \sum_{j_e=1}^{M_\psi} \sum_{i=1}^{4} \sum_{j=1}^{4} \mathcal{M}_{i,j}^{(i_e,j_e,1)},$$

where $\hat{\jmath}$ is 1.

(2) *Higher Pole*: $\psi = 0$, $(x, y, z) = r(0, 0, +1)$ for elements $i_e = 1{:}M_\theta$, $j_e = 1$, $k_e = 2{:}M_r - 1$ for $i = 1{:}2$ (see Fig.1),

$$Y_{i_s,i}^{(i_e,1,k_e)}(t) = Y_{i_s,1}^{(1,1,k_e)}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t)$$

where $\hat{\jmath} = 2 + (k_e - 2)(M_\theta(M_\psi - 1) + 2)$, but also in the $k_e - 1$ neighboring

element $Y^{(i_e,1,k_e-1)}_{i_s,i}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t)$ for $i = 5{:}6$, then

$$\mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv Y^{(1,1,k_e)\,\prime}_{i_s,1}(t) \cdot \sum_{i_e=1}^{M_\theta} \sum_{k=0}^{1} \sum_{i=\bar{i}_k}^{\bar{i}_k+1} \sum_{j=\bar{i}_k}^{\bar{i}_k+1} \mathcal{M}^{(i_e,1,k_e-k)}_{i,j},$$

where $\bar{i}_0 = 1$ and $\bar{i}_1 = 5$. If $k_e = M_r$ when $j_e = 1$, then the no-flux BC (5.43) holds so these terms must be added,

$$\mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv Y^{(1,1,M_r)\,\prime}_{i_s,1}(t) \cdot \sum_{i_e=1}^{M_\theta} \left(\sum_{k=0}^{1} \sum_{i=\bar{i}_k}^{\bar{i}_k+1} \sum_{j=\bar{i}_k}^{\bar{i}_k+1} \mathcal{M}^{(i_e,1,M_r-k)}_{i,j} + \sum_{i=5}^{6} \sum_{j=5}^{6} \mathcal{M}^{(i_e,1,M_r)}_{i,j} \right),$$

where $\hat{\jmath} = 2 + (M_r - 2)(M_\theta(M_\psi - 1) + 2)$.

(3) *Lower Pole*: $\psi = \pi$, $(x, y, z) = r(0, 0, -1)$ for elements $i_e = 1 : M_\theta$, $j_e = M_\psi$, $k_e = 2{:}M_r - 1$ for $i = 3{:}4$ (see 1),

$$Y^{(i_e,M_\psi,k_e)}_{i_s,i}(t) = Y^{(1,M_\psi,k_e)}_{i_s,3}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t)$$

where $\hat{\jmath} = 1 + (k_e - 1)(M_\theta(M_\psi - 1) + 2)$, but also in the $k_e - 1$ neighboring element $Y^{(i_e,M_\psi,k_e-1)}_{i_s,i}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t)$ for $i = 7{:}8$, then

$$\mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv Y^{(1,M_\psi,k_e)\,\prime}_{i_s,3}(t) \cdot \sum_{i_e=1}^{M_\theta} \sum_{k=0}^{1} \sum_{i=\bar{i}_k}^{\bar{i}_k+1} \sum_{j=\bar{i}_k}^{\bar{i}_k+1} \mathcal{M}^{(i_e,M_\psi,k_e-k)}_{i,j},$$

where $\bar{i}_0 = 3$ and $\bar{i}_1 = 7$. If $k_e = M_r$ when $j_e = M_\psi$, then the no-flux BC (5.43) holds so these terms must be added,

$$\mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv Y^{(1,M_\psi,M_r)\,\prime}_{i_s,3}(t) \cdot \sum_{i_e=1}^{M_\theta} \left(\sum_{k=0}^{1} \sum_{i=\bar{i}_k}^{\bar{i}_k+1} \sum_{j=\bar{i}_k}^{\bar{i}_k+1} \mathcal{M}^{(i_e,M_\psi,M_r-k)}_{i,j} + \sum_{i=7}^{8} \sum_{j=7}^{8} \mathcal{M}^{(i_e,M_\psi,M_r)}_{i,j} \right),$$

where $\hat{\jmath} = 1 + (M_r - 1)(M_\theta(M_\psi - 1) + 2)$.

(4) *Periodic BC*: $\theta = 2\pi$, $i_e = M_\theta$, $(x, y, z) = r(\sin(\psi), 0, \cos(\psi))$, the same if $\theta = 0$ or $\theta = 2\pi$, for element $i_e = M_\theta$, $j_e = 2{:}M_\psi - 1$, $k_e = 2{:}M_r - 1$ for local node $i = 4$ (see 1), so

$$Y^{(M_\theta,j_e,k_e)}_{i_s,4}(t) = Y^{(1,j_e,k_e)}_{i_s,3}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t)$$

where $\hat{\jmath} = 3 + (k_e - 2)(M_\theta(M_\psi - 1) + 2) + (j_e - 1)M_\theta$, but also

$$Y_{i_s,2}^{(M_\theta, j_e+\bar{j}_m, k_e-\bar{k}_m)}(t) = \widehat{Y}_{i_s,\hat{\jmath}}(t)$$

when $\bar{j}_m = 1, 0, 1, 0$ and $\bar{k}_m = 0, 0, 1, 1$, respectively, for $m = 1\!:\!4$, then

$$\mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv \left(\mathcal{M}_{3,3}^{(1,j_e,k_e)} + \sum_{m=1}^{4} \mathcal{M}_{2m,2m}^{(M_\theta, j_e+\bar{j}_m, k_e-\bar{k}_m)} \right) \cdot Y_{i_s,3}^{(1,j_e,k_e)\,\prime}(t).$$

If $k_e = M_r$ when $i_e = M_\theta$ and $j_e = 2\!:\!M_\psi - 1$, then the no-flux BC must be added, so

$$\begin{aligned} \mathcal{M}_{\hat{\jmath},\hat{\jmath}} \cdot \widehat{Y}'_{i_s,\hat{\jmath}}(t) \equiv \Bigg(& \mathcal{M}_{3,3}^{(1,j_e,M_r)} + \sum_{m=1}^{4} \mathcal{M}_{2m,2m}^{(M_\theta, j_e+\bar{j}_m, M_r-\bar{k}_m)} \\ & + \mathcal{M}_{6,6}^{(M_\theta, j_e+1, M_r)} + \mathcal{M}_{8,8}^{(M_\theta, j_e, M_r)} \Bigg) \cdot Y_{i_s,3}^{(1,j_e,M_r)\,\prime}(t), \end{aligned}$$

where $\hat{\jmath} = 3 + (M_r - 2)(M_\theta(M_\psi - 1) + 2) + (j_e - 1)M_\theta$.

The actual implementation uses subscripted subscripts to gather the aliased unknowns and equation into their unique locations.

Appendix B. De-aliased spherical node count

The total number of de-aliased points used for the purpose of plotting the results is given by

$$1 + M_r + M_r \cdot (M_\psi - 1) \cdot M_\theta + M_r,$$

and can be seen by the following argument.

(1) $1 + M_r$ points come from the degeneracies at $\psi = 0$, $\theta = 2\pi$ and for $0 < r < R_r$.
(2) $0 < \psi < \pi$, $0 \leq \theta < 2\pi$ and $0 < r \leq R_r$ degeneracies provide for $M_r \cdot (M_\psi - 1) \cdot M_\theta$ points.
(3) Finally we get M_r points for the case where $\psi = \pi$, $\theta = 2\pi$ and $0 < r < R_r$.

We use the three steps above to keep track of the values of the radial, azimuthal and polar values given by r_{k_e}, ψ_{j_e} and θ_{i_e} and made the spherical conversion as usual,

$$x_h = (r_{k_e})_h \cdot \cos(\theta_{i_e})_h \cdot \sin(\psi_{j_e})_h,$$
$$y_h = (r_{k_e})_h \cdot \sin(\theta_{i_e})_h \cdot \sin(\psi_{j_e})_h,$$
$$z_h = (r_{k_e})_h \cdot \cos(\psi_{j_e})_h.$$

Here h is the index used in steps (1)-(3) outlined above. It is important to point out at this stage that the plots obtained were for $r \in [0, 5]$ and not centered as in the case of the two dimensional plots. To obtain plots in three dimension the key thing was the definition of k_{plot} which is an $(M_r) \times (M_\theta + 1) \times (M_\psi + 1)$ array. The k_{plot} is an indexing to convert the $1 + 2M_r + M_r \cdot (M_\psi - 1) \cdot M_\theta$ back to indices for the three variables. The reconversions are summarized below.

(1) For the non-boundary case $k_e = 2 : M_r$,
   (a) $j_e = 1$,

$$k_{plot,k_e,i_e,j_e} = \begin{cases} 2 + (k_e - 2)[M_\theta(M_\psi - 1) + 2], i_e = 1 : M_\theta \\ k_{plot,k_e,1,1}, i_e = M_\theta + 1. \end{cases}$$

   (b) $j_e = 2 : M_\psi$,

$$k_{plot,k_e,i_e,j_e} = \begin{cases} 2 + (k_e - 2)[M_\theta(M_\psi - 1)] + M_\theta(j_e - 2) + i_e \\ \text{for} \quad i_e = 1 : M_\theta \\ k_{plot,k_e,1,M_\psi+1}, i_e = M_\theta + 1. \end{cases}$$

   (c) $j_e = M_\psi$,

$$k_{plot,k_e,i_e,j_e} = \begin{cases} 1 + (k_e - 1)[M_\theta(M_\psi - 1)], i_e = 1 : M_\theta \\ k_{plot,k_e,1,M_\psi+1}, i_e = M_\theta + 1. \end{cases}$$

(2) For the boundary case $k_e = M_r + 1$,
   (a) $j_e = 1$,

$$k_{plot,k_e,i_e,j_e} = \begin{cases} 2 + (M_r - 2)[M_\theta(M_\psi - 1) + 2], i_e = 1 : M_\theta \\ k_{plot,k_e,1,1}, i_e = M_\theta + 1. \end{cases}$$

   (b) $j_e = 2 : M_\psi$,

$$k_{plot,k_e,i_e,j_e} = \begin{cases} 2 + (k_e - 2)[M_\theta(M_\psi - 1)] + M_\theta(j_e - 2) + i_e, \\ \text{for} \quad i_e = 1 : M_\theta \\ k_{plot,k_e,1,M_\psi+1}, i_e = M_\theta + 1. \end{cases}$$

   (c) $j_e = M_\psi$,

$$k_{plot,k_e,i_e,j_e} = \begin{cases} 1 + (M_r - 1)[M_\theta(M_\psi - 1)], i_e = 1 : M_\theta \\ k_{plot,k_e,1,M_\psi+1}, i_e = M_\theta + 1. \end{cases}$$

Maple™ was used to exactly evaluate the integrals of the element matrices off-line. The the integrals and their exact values evaluated by Maple™ .

Appendix C. A constrained control formulation

In this Appendix, we briefly present an alternative numerical approach, as suggest by one of the reviewers, from the perspective of constrained and flow control optimization as envisioned in [6,25,26,28]. The state equations in the compact form can be written as,

$$\mathcal{M}_1 \widehat{\mathbf{Y}}' = \mathcal{A}(\widehat{\mathbf{Y}})\widehat{\mathbf{Y}} - \mathcal{M}_2 \widehat{\mathbf{U}}(t),$$

where $\mathcal{M}_i$ has a dimension of 3×3. The corresponding objective functional in a terse form is as follows,

$$\begin{aligned} J(\widehat{\mathbf{Y}}, \widehat{\mathbf{U}}) = & \frac{1}{2} \int_0^{t_f} dt \int_\Omega d\mathbf{x} \left(\widehat{\mathbf{Y}}^\top R \widehat{\mathbf{Y}} + \left(\widehat{\mathbf{U}} - \widehat{\mathbf{U}}_0 \right)^\top S \left(\widehat{\mathbf{U}} - \widehat{\mathbf{U}}_0 \right) \right) \\ & + \frac{1}{2} \int_\Omega d\mathbf{x} \left(\widehat{\mathbf{Y}}^\top Q \widehat{\mathbf{Y}} \right) (\mathbf{x}, t_f). \end{aligned}$$

The discretized version of this control problem would be,

$$\begin{aligned} & \min \frac{\Delta t}{2} \left[\left[\sum_{i=1}^{t_{\max}} \left(0.5 \left(\widehat{\mathbf{Y}}_i^\top R \widehat{\mathbf{Y}}_i + \widehat{\mathbf{Y}}_{i+1}^\top R \widehat{\mathbf{Y}}_{i+1} \right) \right. \right. \right. \\ & \left. \left. \left. + 0.5 \left(\left(\widehat{\mathbf{U}}_i - \widehat{\mathbf{U}}_{0,i} \right)^\top S \left(\widehat{\mathbf{U}}_i - \widehat{\mathbf{U}}_{0,i} \right) + \left(\widehat{\mathbf{U}}_{i+1} - \widehat{\mathbf{U}}_{0,i+1} \right)^\top S \left(\widehat{\mathbf{U}}_{i+1} - \widehat{\mathbf{U}}_{0,i+1} \right) \right) \right) \right] \\ & + \left(0.5 \left(\widehat{\mathbf{Y}}_i^\top Q \widehat{\mathbf{Y}}_i + \widehat{\mathbf{Y}}_{i+1}^\top Q \widehat{\mathbf{Y}}_{i+1} \right) \right) \Big]. \end{aligned}$$

subject to,

$$\mathcal{M}_1 \left(\widehat{\mathbf{Y}}_{i+1} - \widehat{\mathbf{Y}}_i \right) = \frac{\Delta t}{2} \left[0.5 \left(\mathcal{A}(\widehat{\mathbf{Y}}_i)\widehat{\mathbf{Y}}_i + \mathcal{A}(\widehat{\mathbf{Y}}_{i+1})\widehat{\mathbf{Y}}_{i+1} \right) - \mathcal{M}_2 0.5 \left(\widehat{\mathbf{U}}_i + \widehat{\mathbf{U}}_{i+1} \right) \right].$$

4. Parameter values for numerical implementation

Symbol	Parameter	Value
a_C	Reabsorption rate for the drug [39]	11.3 per day
a_N	Intrinsic growth rate for normal tissue	8.64e-7 per day
a_T	Intrinsic growth rate for tumor cells [37]	1.20e-2 per day
D_C	Diffusion coefficient for drug concentration [39]	2.16e-1 cm^2 per day
D_N	Diffusion coefficient for normal tissue	1.0e-15 cm^2 per day
D_T	Diffusion coefficient for tumor cells [37]	4.2e-3 cm^2 per day
k_N	Normal tissue carrying capacity	1.0
k_T	Tumor cell carrying capacity	1.0
q_C	Drug delivery final cost coefficient	0.1
q_T	Tumor burden final cost coefficient	0.1
r_T	Tumor burden running cost coefficient	0.2
s_U	Drug delivery running cost coefficient	0.05
R_T	Initial tumor radius	1.25 cm
R_D	Initial drug radius	1.25 cm
wU_C	Weight of U_0 drug control distributions	8.0
$\alpha_{T,N}$	Death rate of tumor cells due to competition	1.0e-4 per day
$\alpha_{N,T}$	Death rate of normal tissue due to competition	1.0e-4 per day
$\kappa_{T,C}$	Death rate of tumor cells due to treatment	8.0 per day
$\kappa_{N,C}$	Death rate of normal tissue due to treatment	1.0e-4 per day