Designing Radiotherapy Plans with Elastic Constraints and Interior Point Methods

Allen Holder[†]

May 29, 2000

Abstract

A new linear programming model used to aid in the design of radiotherapy plans is introduced. This model incorporates elastic constraints, and when solved with a path following interior point method, produces favorable plans. A sound mathematical analysis shows how to interpret the solution, and hence, the treatment planner receives meaningful knowledge about the radiotherapy plan being developed. Preliminary experiments are conducted.

Key words: analytic center, radiation oncology, elastic constraints
AMS subject classification: 90C05,90C51,90C90

[†] Department of Mathematics, Trinity University, San Antonio, TX, USA

1 Introduction

Because cancerous and displasiac cells are more sensitive to radiation than normal cells, external beam radiation is often used to treat such maladies. The idea is to focus radiation beams that are external to the body in such a way that they deposit radiation into a tumor but do not deposit an abundance of radiation into critical structures. The process of choosing a collection of beams, together with their energies, is called *treatment planning*. Treatment procedures fall into the two basic categories of radiosurgery and radiotherapy, the difference being that radiosurgery is completed all-at-once and radiotherapy is extended over several smaller treatments (usually fractions of the radiosurgery plan).

Linear programming models have been used extensively to find "good" radiotherapy treatment plans. In fact, the first model proposed in 1968 was a linear programming model [2], and many researchers have experimented with linear programming formulations [13, 18, 15, 16, 17, 23, 27, 29, 30, 29]. Because the integral dose to a cell is additive, the linear model seems natural and appropriate.

While linear models are prevalent in the literature, they have had their drawbacks. Subsequently, several nonlinear models have been proposed [16, 22, 27, 29]. The recent review article by Shepard, Ferris, Olivera, and Mackie [29] nicely characterizes many of the optimization models that have been used to design radiotherapy plans. One of the most significant drawbacks of using linear programming lies not in any specific model but rather in the solution technique. As discussed in [13], the fact that the simplex algorithm produces a basic optimal solution implies that some of the physician's bounds are achieved. Depending on the specific linear programming model, such achievement indicates that either portions of the critical structures are to receive their maximum allowable dose or that the tumor is to receive the lowest allowable dose. Neither of these outcomes is desirable. What would be nice, and is forthcoming, is a linear programming model together with a solution technique that adheres to the physician's prescribed limits as much as possible. Moreover, our model returns a treatment plan even when the treatment goals are impossible. So diagnosing infeasible demands, which is the other major criticism of the use of linear programming [22, 24, 27], is a non-issue.

The article is organized in the following manner. Section 2 describes the *dose deposition* matrix, which is used to calculate the amount of radiation small regions of tissue receive. In Section 3 we discuss the treatment planner's concerns when designing a specific plan. The new model is presented in Section 4, and the results of this section provide a meaningful solution analysis. In Section 5 we discuss why treatment plans designed from the solution provided by a path-following interior point algorithm are desirable. Experiments are found in Section 6, and the results are summarized in Section 7.

The symbol '$\equiv$' is used as a definitional assignment, and the vector of ones is denoted by e, where the length is decided by the context with which it is used. The matrix A^+ denotes the Moore-Penrose generalized inverse of the matrix A.

We use $\underline{\text{rs}}(A)$ to denote the minimum row sum of the matrix A –i.e. $\underline{\text{rs}}(A) \equiv \min_i\{A_{(i,\cdot)}e\}$. Other notation and terminology used is consistent with that found in the Mathematical Programming Glossary [12].

2 The Dose Deposition Matrix

In this section we show how to construct a linear operator that is used to model radiation deposition. Specifically, we show how to construct the *dose deposition matrix*, which is the transpose of the fully discretized radon transform [1, 7, 17]. Suppose we have a $N \times N$ pixel image, and that the available angles are θ_1, θ_2, θ_3, ..., θ_Θ. Furthermore, assume that each angle is comprised of η sub-beams. These sub-beams may be *elementary beams* or *pencils*, the difference being that pencils radiate from a point source and elementary beams are parallel to each other. Our development does not depend on whether pencils or elementary beams are chosen. Sub-beams are included in our model because modern treatment systems are capable of intricate *collimation*. This technology uses a multileaf collimator located in the gantry to shape the beam. The geometry of a model using elementary beams, where $n = 2$, $\rho = 4$, and $\eta = 4$, is indicated in Figure 1 (so as to clarify the geometry, the first angle is located at $\pi/4$, whereas, the initial angle is usually located at 0).

We set $x_{(a,i)}$ to be the dose along the i^{th} sub-beam of angle a, and $d_{(p,a,i)}$ to be the distance from where sub-beam $x_{(a,i)}$ enters the image to where it reaches pixel p. We further define $A_{(p,a,i)}$ to be the product of $e^{-\mu d_{(p,a,i)}}$ and the geometric area common to both the sub-beam $x_{(a,i)}$ and pixel p. For example, in Figure 1 the elementary beam corresponding to $x_{(1,2)}$ intersects one-half of pixel 3 and the distance to this pixel along this elementary beam is $3\sqrt{2}/2$ (assuming that each pixel has a width of one). Hence, $A_{(3,1,2)} = \frac{1}{2}e^{\frac{3\sqrt{2}}{2}\mu}$. The components of the dose deposition matrix, denoted by A, are $A_{(p,a,i)}$, where the rows of A are indexed by p and the columns are indexed by (a, i).

The factor $e^{-\mu d_{(p,a,i)}}$ is included to measure how a beam of radiation attenuates as it passes through the body, and values for the *attenuation coefficient*, μ, depend on the energy of the beam. The gamma-rays formed from Cobalt-60 have a relatively low energy of 1.25 MeV and correspond to a μ value of approximately 0.06. Most modern facilities use linear accelerators and can produce beams of greater intensities, like 8 MV and 15 MV. A Typical μ value for these higher energy beams is 0.02 [8]. For the geometry indicated in Figure 1, the dose deposition without attenuation is

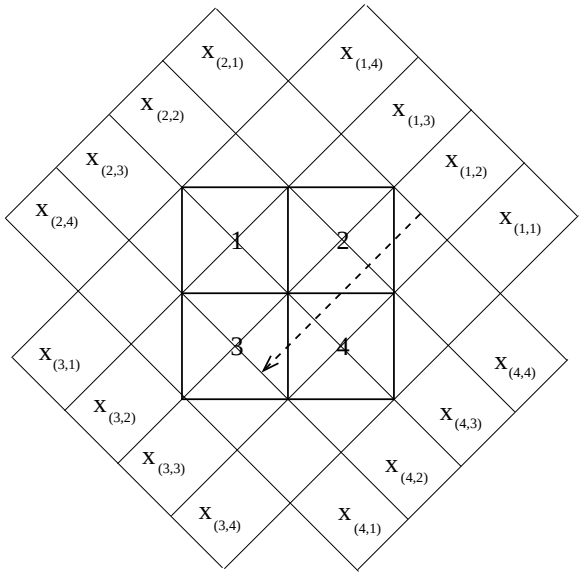

Figure 1: The geometry of a 2×2 pixel image with angles at $\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$, and $\frac{7\pi}{4}$.

$$\begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

We set x to be the transpose of

$$\begin{bmatrix} x_{(1,1)} & x_{(1,2)} & \dots & x_{(1,r)} & x_{(2,1)} & \dots & x_{(2,r)} & \dots & x_{(\rho,1)} & \dots & x_{(\rho,r)} \end{bmatrix},$$

so that the total, or *integral*, radiation dose for pixel p is the pth component of Ax.

3 Modeling Issues

While the dose deposition matrix allows one to easily model upper and lower bounds for any given pixel in an image, modeling the treatment planners desires is a non-trivial exercise. For example, a treatment planner might desire that the tumor receive no less than 80 Gy. Similarly, the planner may hope that some other critical structure receive no more than 40 Gy. Then, for each pixel p in the tumorous region

we have the linear inequality

$$\sum_{(a,i)} A_{(p,a,i)} x_{(a,i)} \geq 80,$$

and for every pixel p within the critical structure we have

$$\sum_{(a,i)} A_{(p,a,i)} x_{(a,i)} \leq 40.$$

Any element contained within the polyhedron described by these linear inequalities satisfies the planner's goals. Because of this, some researchers have chosen to pose this problem as a feasibility problem [6, 24]. However, the majority of research has been directed towards optimization models with linear constraints, with the most immediate and natural objective functions being either to maximize tumor dose or minimize critical structure dose. Because the maximization of a sum may lead to one term having an unusually high value, other researchers have maximized the minimum tumor dose or minimized the maximum critical structure dose [17]. Another linear objective that has been considered is that of maximizing the difference between the tumor dose and critical structure dose [23]. Each of these objectives functions captures a desirable quality of a treatment plan, but no single objective completely describes the planners desires. Some have tried to encompass several of these objectives with weighted sums [29, 19], but it is difficult to give precise meaning to the weights [9]. The goals listed below indicate that a treatment planner has a great deal to consider when deciding on a clinically desirable treatment plan.

Treatment Planners Desires
· Deliver a uniform, tumoricidal dose to the tumorous region.
· Deliver as little radiation as possible to the critical structures.
· Make the integral dose as small as possible.
· Reduce the frequency of unusually high doses outside the tumorous region.
· Control the number of beams in the treatment plan.

The goal of a uniform dose over the tumor is somewhat unnatural at first. Indeed, one would suspect that the goal is to deliver as much radiation to the tumor as possible. There are two main reasons for desiring a uniform tumoricidal dose. First, unusually high levels of radiation can lead to large amounts of necrosis, and the human body may experience difficulties dealing with a large volume of dead tissue. Second, tumor cells are interspersed among healthy tissue. Hence, a uniform dose to the tumorous region is crucial to a successful plan [30]. This follows because a dose that is too low allows cancer cells to survive, while a dose that is too high can have adverse affects to the surrounding tissue.

Another modeling issue that needs consideration is that different types of tissue react differently to radiation exposure. For some organs, known as *rope organs*, loosing functionality of a significant amount of tissue has little effect on the overall

ability of the organ to function properly. Moreover, rope organs can loose overall functionality if the entire organ is exposed to a relatively low level of radiation. The kidney, liver, and lung are rope organs. This type of organ lends itself well to what are known as dose-volume constraints [18, 15]. However, such constraints lead to large mixed integer programs, and current solution procedures have unrealistic time demands. Other organs, called *chain organs*, loose all functionality once a functional subunit is destroyed. This type of tissue can handle relatively high, uniform radiation levels, but a single hot spot could render the entire organ useless. The spinal cord and bowel are examples of chain organs. A more thorough discussion of rope and chain organs is found in [10, 25, 31, 32].

While there are many other obscure considerations taken into account when designing a specific patient's treatment plan, the desires just discussed are quantifiable. Optimization models found in the literature tend to focus on a single desirable characteristics, whether it be a uniform tumor dose or a dose-volume constraint. In the next section we present a new linear programming model and address how the planner's goals are dealt with.

4 A New Linear Model

Each pixel is assumed to represent either tumorous or non-tumorous tissue. Let m be the total number of pixels, m_T be the number of tumorous pixels, m_C be the number of critical structure pixels, and $m_G = m - m_T - m_C$ be the number of remaining pixels. Also let n be the number of sub-beams that intersect the target volume. A *prescription* is comprised of a physician's aspirations for the tumor, usually a tumoricidal dose, and upper bounds for the non-tumorous tissue. Specifically, a prescription is the 4-tuple (TUB, TLB, CUB, GUB), where

- TUB is a m_T vector of upper bounds for the tumor,
- TLB is a m_T vector of lower bounds for the tumor,
- CUB is a m_C vector of upper bounds for the the critical structures, and
- GUB is a m_G vector of upper bounds for the remaining good tissue.

We make the realistic assumptions that $0 < TLB \leq TUB$, $0 \leq CUB$, and $0 \leq GUB$. Because a uniform, tumoricidal dose is to be delivered to the tumor, the lower and upper bounds for the tumor pixels are taken to be a fixed percentage of the physician's goal for the tumor. Supposing that the physician's goal for a tumorous cell is TG, values for TUB_i and TLB_i are $(1 + tol)TG$ and $(1 - tol)TG$, respectively. Here, *tol* is a percentage of variation for the tumor dosage and is called the *tumor uniformity level*. Typical values of *tol* found in the literature range from 0.02 to 0.15. The vector GUB describes the highest amount of radiation that any single pixel is allowed, and in general no tissue should receive more than 10% of the tumor's desired dose. Hence, we set $GUB = TG(1 + 0.10)$.

The rows of the dose deposition matrix are partitioned and reordered into the rows that correspond to the tumor, the rows that correspond to the critical structures, and the rows that correspond to the remaining good tissue. This reordering is indicated by the sub-matrices A_T, A_C, and A_G, as indicated below,

$$A = \begin{bmatrix} A_T \\ A_C \\ A_G \end{bmatrix} \begin{array}{l} \longleftarrow \text{Tumor} \\ \longleftarrow \text{Critical Structures} \\ \longleftarrow \text{Remaining Good Tissue.} \end{array}$$

Sub-beams that do not intersect the tumor are removed from consideration by eliminating the columns of A that have a corresponding zero column in A_T. For notational brevity, we keep the A notation for the sub-matrix with these columns removed. So, in what follows $A \in \mathbb{R}^{m \times n}$, $A_T \in \mathbb{R}^{m_T \times n}$, $A_C \in \mathbb{R}^{m_C \times n}$, and $A_G \in \mathbb{R}^{m_G \times n}$.

A matrix $H \in \mathbb{R}^{p \times q}$ is *semimonotone* provided its More-Penrose generalized inverse, denoted H^+, is a nonnegative matrix. Nonnegative, semimonotone matrices are characterized by the fact that there exists a positive diagonal matrix for which $H^+ = DH^T$ [4]. For the remainder of this paper, the following semimonotone matrices are assumed to have full column rank: $l \in \mathbb{R}^{q_T}$, $u_C \in \mathbb{R}^{q_C}$, $u_G \in \mathbb{R}^{q_G}$, $L \in \mathbb{R}^{m_T \times q_T}$, $U_C \in \mathbb{R}^{m_C \times q_C}$, and $U_G \in \mathbb{R}^{m_G \times q_G}$. We further assume that l, u_C, and u_G are positive, and that L, U_C, and U_G are nonnegative with no row sum being zero –i.e. $Le > 0$, $U_C e > 0$, and $U_G e > 0$. Now, for any positive scalar ω, the *elastic model* is

$$EM(\omega) \begin{cases} \min \quad \omega \cdot l^T \alpha + u_C^T \beta + u_G^T \gamma \\ \\ \text{such that} \\ \\ \begin{array}{rcccl} TLB - L\alpha & \leq & A_T x & \leq & TUB \\ & & A_C x & \leq & CUB + U_C \beta \\ & & A_G x & \leq & GUB + U_G \gamma \\ 0 & \leq & L\alpha & \leq & TLB \\ -CUB & \leq & U_C \beta & & \\ 0 & \leq & U_G \gamma & & \\ 0 & \leq & x, & & \end{array} \end{cases}$$

where $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}^{q_T}$, $\beta \in \mathbb{R}^{q_C}$, and $\gamma \in \mathbb{R}^{q_G}$. We note that it is easy to show that because L and U_G are semimonotone, α, and γ are nonnegative.

The constraints $TLB - L\alpha \leq A_T x$, $A_C x \leq CUB + U_C \beta$, and $A_G x \leq GUB + U_G \gamma$ are called *elastic* because the bounds are allowed to vary with the vectors α, β, and γ, respectively. The matrices L, U_C, and U_G define how one measures the amount of elasticity, and l, u_C, u_G show how one either penalizes or rewards the amount of elasticity. With this in mind, we see that the assumption that l, u_C, and u_G are positive simply guarantees that any discrepancy is penalized, and that the

assumption that $Le > 0$, $U_C e > 0$, and $U_G e > 0$ makes sure that each constraint is elastic. Any collection of l, u_C, u_G, L, U_C, and U_G defines a set of *elastic functions*, and they are incorporated for the following reasons. First, as is shown in Lemma 2, the elastic constraints guarantee that for any collection of elastic functions, $EM(\omega)$ is always strictly feasible. Second, the different lower bounds on the elastic functions allow us to embody different treatment aspirations.

Because different elastic functions procure a different solution analysis, an interpretation of $EM(\omega)$ depends on the specific elastic functions. In particular, we make the following selections when discussing *average analysis* and *absolute analysis*.

Average Analysis

$$l^T\alpha : \mathbb{R}^{m_T} \to \mathbb{R} : \alpha \to (1/m_T)\, e^T\alpha \qquad L\alpha : \mathbb{R}^{m_T} \to \mathbb{R}^{m_T} : \alpha \to \alpha$$
$$u_C^T\beta : \mathbb{R}^{m_C} \to \mathbb{R} : \beta \to (1/m_C) e^T\beta \qquad U_C\beta : \mathbb{R}^{m_C} \to \mathbb{R}^{m_C} : \beta \to \beta$$
$$u_G^T\gamma : \mathbb{R}^{m_G} \to \mathbb{R} : \gamma \to (1/m_G) e^T\gamma \qquad U_G\gamma : \mathbb{R}^{m_G} \to \mathbb{R}^{m_G} : \gamma \to \gamma$$

Absolute Analysis

$$l^T\alpha : \mathbb{R} \to \mathbb{R} : \alpha \to \alpha \qquad L\alpha : \mathbb{R} \to \mathbb{R}^{m_T} : \alpha \to \alpha e$$
$$u_C^T\beta : \mathbb{R} \to \mathbb{R} : \beta \to \beta \qquad U_C\beta : \mathbb{R} \to \mathbb{R}^{m_C} : \beta \to \beta e$$
$$u_G^T\gamma : \mathbb{R} \to \mathbb{R} : \gamma \to \gamma \qquad U_G\gamma : \mathbb{R} \to \mathbb{R}^{m_G} : \gamma \to \gamma e$$

Suppose that average analysis is chosen. Then $(L\alpha)_p = \alpha_p$ tells us how deficient a plan is with regards to meeting the minimum tumor dose for pixel p, and $l^T\alpha = (1/m_T)e^T\alpha$ is the average amount of such deficiencies. The interpretation of $U_C\beta = \beta$ depends on the sign of the component. If $(U_C\beta)_p = \beta_p > 0$, pixel p, which is contained in some critical structure, is receiving more radiation than intended. However, if $\beta_p < 0$, pixel p is receiving less radiation than is allowed. We now see that the objective term $u_C^T\beta = (1/m_C)e^T\beta$ expresses the desire to decrease the average dose to the critical structures; in fact the desire is to have the critical structures receive no radiation. Similarly, $(U_G)\gamma_p = \gamma_p$ indicates how much pixel p is over its alloted upper bound, and $u_G^T\gamma = (1/m_G)e^T\gamma$ is the average amount of radiation the normal tissue is over its prescribed dose. The roles of β and γ differ because of the different lower bounds. Since $0 \leq \gamma$, any plan satisfying $A_G x \leq GUB$ contributes zero to the objective function. However, the lower bound of $-CUB$ on β means that plans with a low integral dose to the critical structures are preferred. So, for the average analysis case we see that the objective function is three tiered in its goals:

- minimize the average amount the tumor is under its prescribed dose,
- minimize the average amount of radiation that the critical structures receive, and
- minimize the average amount the remaining pixels are over their upper bounds.

If absolute analysis is chosen, the interpretation of $EM(\omega)$ is similar to that of average analysis. The difference is that the elastic functions are each controlled by a single parameter. So instead of minimizing an average discrepancy, the goal is to minimize the maximum amount of discrepancy. Hence, when absolute analysis is chosen, the three goals of the objective function are to

- minimize the maximum amount the tumor is under its prescribed dose,
- minimize the maximum amount of radiation that the critical structures receive, and
- minimize the maximum amount any remaining pixel is over its upper bound.

Of course any collection of elastic functions may be used, and there is really no need for them to even be linear. However, we continue with the two collections describing average and absolute analysis and show that a meaningful explanation of the solution is possible.

As just described, the objective function is a weighted sum of three goals, and while a common criticism of such objective functions is that the weights are difficult to understand, we show that choosing ω appropriately provides a meaningful interpretation. The positive scalar ω weights the importance of a plan achieving the minimum tumor dose –i.e. large values of ω encourage $l^T\alpha$ to be as small as possible. We would like to have the property that there exists a finite $\omega > 0$ such that the optimal value of $l^T\alpha$ is zero. This follows because the tumor is then guaranteed to receive its minimum radiation level. Such an ω would serve as a certificate of a tumoricidal dose. The bad news is that there are simple examples where the optimal value of $l^T\alpha$ is not zero for all $\omega > 0$. However, the good news is that we show how to easily find an ω that does certify that the discrepancy between the amount delivered to the tumor and the tumor's lower bound is sufficiently small. This result is stated in Theorem 1, and relies on the following two lemmas. In what follows, we use the standard big-O notation –i.e. $f(x) = O(g(x))$ if, for the non-negative functions f and g, there exists a positive constant κ, such that $f(x) \leq \kappa g(x)$.

Lemma 1 *Consider the functions $F_1 : \mathbb{R}^{p_v} \to \mathbb{R}^{q_1}$, $F_2 : \mathbb{R}^{p_y} \to \mathbb{R}^{q_1}$, $G_1 : \mathbb{R}^{p_v} \to \mathbb{R}^{q_2}$, $G_2 : \mathbb{R}^{p_z} \to \mathbb{R}^{q_2}$, $f : \mathbb{R}^{p_y} \to \mathbb{R}$, and $g : \mathbb{R}^{p_z} \to \mathbb{R}$. Let ω be an arbitrary positive scalar, $F_2(0) = 0$, $f(0) = 0$, and suppose the following mathematical programs are well-posed:*

$$\min\{\omega f(y) + g(z) : F_1(v) \leq a + F_2(y), G_1(v) \leq b + G_2(z)\} \tag{1}$$

$$\min\{g(z) : F_1(v) \leq a + F_2(y), G_1(v) \leq b + G_2(z)\} \tag{2}$$

$$\min\{g(z) : F_1(v) \leq a, G_1(v) \leq b + G_2(z)\}, \tag{3}$$

where we assume without loss in generality that each optimal value is positive. Then, denoting an optimal solution of (1) by $(v^*(\omega), y^*(\omega), z^*(\omega))$, we have that

$$f(y^*(\omega)) = O\left(\frac{1}{\omega}\right).$$

Proof: Let $(\bar{v}, \bar{z})$ be optimal to (3) and set M to be the optimal value of (2). Noticing that $(v, y, z) = (\bar{v}, 0, \bar{z})$ is feasible to (1), we have by the optimality of $(v^*(\omega), y^*(\omega), z^*(\omega))$ that

$$\omega f(y^*(\omega)) + g(z^*(\omega)) \leq \omega f(0) + g(\bar{z}) = g(\bar{z}).$$

Since $g(z^*(\omega)) \geq M$, we conclude that

$$\omega f(y^*(\omega)) \leq g(\bar{z}) - M. \tag{4}$$

The fact that both $g(\bar{z})$ and M are not reliant on ω completes the proof. ■

The condition that the mathematical programs in (1), (2), and (3) be well-posed manifests itself in different manners as related to $EM(\omega)$. First, $EM(\omega)$ being well posed is equivalent to the condition that both it and its dual are feasible. The elastic constraints actually allow the stronger statement that $EM(\omega)$ and its dual are *strictly feasible*, which means there exists a feasible element strictly satisfying the inequalities. We denote the feasible regions for $EM(\omega)$ and its dual by $\mathcal{P}$ and $\mathcal{D}$, respectively, and their strict interior by $\mathcal{P}^o$ and $\mathcal{D}^o$. Lemma 2 now shows that $EM(\omega)$ is not only well-posed, but that both the primal and dual feasible regions have non-empty strict interiors.

Lemma 2 *For any collection of elastic functions we have that*

$$\mathcal{P}^o \neq \emptyset \text{ and } \mathcal{D}^o \neq \emptyset.$$

Proof: From the fact that every tumor pixel intersects some elementary beam we have that $A_{_T}e > 0$. Hence, $\varepsilon A_{_T}e > 0$, for all $\varepsilon > 0$. The assumption that $TUB > 0$ implies that there exists $\varepsilon^1 > 0$ such that $\varepsilon^1 A_{_T}e < TUB$. Since $Le > 0$, $U_Ce > 0$, and $U_Ge > 0$, we may choose $\varepsilon^2 > 0$, $\varepsilon^3 > 0$, and $\varepsilon^4 > 0$ such that

$$\begin{aligned}
\varepsilon^2 Le &> \max\{0, TLB - \varepsilon^1 A_{_T}e\}, \\
\varepsilon^3 U_C e &> \max\left\{\varepsilon^1 A_{_C}e - CUB, 0\right\}, \text{ and} \\
\varepsilon^4 U_G e &> \max\left\{\varepsilon^1 A_{_G}e - GUB, 0\right\}.
\end{aligned}$$

From the assumption that L, U_C, and U_G are nonnegative matrices with no row sum of zero, we have that $(x, \alpha, \beta, \gamma) = (\varepsilon^1 e, \varepsilon^2 e, \varepsilon^3 e, \varepsilon^4 e) \in \mathcal{P}^o$.

The dual feasible region is defined by

$$\begin{aligned}
A_{_T}^T\pi^1 - A_{_T}^T\pi^2 - A_{_C}^T\pi^3 - A_{_G}^T\pi^4 &\leq 0 \\
L^T\pi^1 - L^T\pi^5 + L^T\pi^7 &= \omega \cdot l \\
U_C^T\pi^3 + U_C^T\pi^6 &= u_C \\
U_G\pi^4 + U_G\pi^8 &= u_G \\
\pi^1, \pi^2, \pi^3, \pi^4, \pi^5, \pi^6, \pi^7, \pi^8 &\geq 0.
\end{aligned}$$

Set $\bar{\pi}^5 = e$. Since L has full column rank, the system $L^T(\pi^1 + \pi^7) = \omega \cdot l + L^T e$ is consistent, and hence $(L^T)^+(\omega \cdot l + L^T e)$ is a solution. From Theorem 5.2 in [4] there exists a positive diagonal matrix D such that $L^+ = DL^T$. So, from the assumption that $l > 0$ and the fact that $(L^+)^T = (L^T)^+$ we have,

$$\begin{aligned}(L^T)^+(\omega \cdot l + L^T e) &= \omega(L^+)^T l + (L^+)^T L^T e \\ &= \omega(DL^T)^T l + (DL^T)^T L^T e \\ &= \omega LDl + LDL^T e \\ &> 0.\end{aligned}$$

Set $\bar{\pi}^1 = \bar{\pi}^7 = (1/2)(L^T)^+(\omega \cdot l + L^T e)$. In a like fashion, set $\bar{\pi}^3 = \bar{\pi}^6 = (1/2)(U_C^T)^+ u_C$ and $\bar{\pi}^4 = \bar{\pi}^8 = (1/2)(U_G^T)^+ u_G$. Since only sub-beams intersecting the tumor are included in the model, we have that $A_T^T e > 0$. Hence, there exist a $\delta > 0$ such that

$$\delta A_T^T e > A_T^T \bar{\pi}^1 - A_C \bar{\pi}^3 - A_G^T \bar{\pi}^4.$$

The proof is concluded since $(\bar{\pi}^1, \delta e, \bar{\pi}^3, \bar{\pi}^4, \bar{\pi}^5, \bar{\pi}^6, \bar{\pi}^7, \bar{\pi}^8) \in \mathcal{D}^o$ ■

Recall that the goal was to provide a certificate that guarantees the tumor receives at least the amount of radiation prescribed by TLB. We momentarily show how to construct such a certificate provided that the tumor's prescription is attainable with the geometry described by A. We formalize this concept in the next definition, and while the definition does not rely on the entire prescription, we include the entire prescription for linguistic simplicity.

Definition 1 *The prescription (TUB, TLB, CUB, GUB) allows tumor uniformity if there exists a plan, $x \geq 0$, such that $TLB \leq A_T x \leq TUB$. Any such plan is said to attain tumor uniformity.*

We note that every prescription allows tumor uniformity provided that the attenuation coefficient is zero –i.e. a sub-beams depth into the image is not a factor of the dose deposition coefficient $A_{(p,a,i)}$. This follows because each tumor pixel is covered by the sub-beams of any given angle. So, a dose deposition matrix formed without attenuation has the property that $A_T e = \Theta e$, where Θ is the number of possible angles. Hence, there exists ε such that $\varepsilon A_T e = TGe$, and tumor uniformity is guaranteed. Theorem 1 now shows that if the tumor's prescription is attainable, the tumor deficiency is uniformly bounded above by the inverse of ω.

Theorem 1 *Let $(x^*(\omega), \alpha^*(\omega), \beta^*(\omega), \gamma^*(\omega))$ be an optimal solution to $EM(\omega)$. For any collection of linear elastic functions we have that $l^T \alpha^*(\omega) = O\left(\frac{1}{\omega}\right)$, provided that the prescription allows tumor uniformity.*

Proof: Make the following assignments:

$$f(\alpha) = l^T\alpha, \qquad g(\beta,\gamma) = u_C^T\beta + u_G^T\gamma,$$

$$F_1(x) = \begin{bmatrix} -A_T x \\ 0 \\ 0 \end{bmatrix},\ F_2(\alpha) = \begin{bmatrix} L\alpha \\ L\alpha \\ -L\alpha \end{bmatrix},\ a = \begin{bmatrix} -TLB \\ 0 \\ TLB \end{bmatrix},$$

$$G_1(x) = \begin{bmatrix} A_T x \\ A_C x \\ A_G x \\ 0 \\ 0 \\ -x \end{bmatrix},\ G_2(\beta,\gamma) = \begin{bmatrix} 0 \\ U_C\beta \\ U_G\gamma \\ U_C\beta \\ U_G\gamma \\ 0 \end{bmatrix}, \text{ and } b = \begin{bmatrix} TUB \\ CUB \\ GUB \\ CUB \\ 0 \\ 0 \end{bmatrix}.$$

We now have that $EM(\omega)$ is

$$\min\{\omega f(\alpha) + g(\beta,\gamma) : F_1(x) \le a + F_2(\alpha), G_1(x) \le b + G_2(\beta,\gamma)\}. \tag{5}$$

From Lemma 2 we have that this linear program is well posed. To invoke Lemma 1 we need to establish that the following two linear programs are well-posed,

$$\begin{aligned} LP_1 \quad & \min\{g(\beta,\gamma) : F_1(x) \le a + F_2(\alpha), G_1(x) \le b + G_2(\beta,\gamma)\} \\ & = \min\{u_C^T\beta + u_G^T\gamma : (x,\alpha,\beta,\gamma) \in \mathcal{P}\}, \text{ and} \\ LP_2 \quad & \min\{g(\beta,\gamma) : F_1(x) \le a, G_1(x) \le b + G_2(\beta,\gamma)\} \\ & = \min\{u_C^T\beta + u_G^T\gamma : (x,\alpha,\beta,\gamma) \in \mathcal{P}, \alpha = 0\}. \end{aligned}$$

From Lemma 2, $\mathcal{P}^o \neq \emptyset$, and LP_1 is strictly feasible. Let $\bar{x}$ be a plan that attains tumor uniformity. Then, similar to the proof of Lemma 2 there exists $\varepsilon^1 > 0$ and $\varepsilon^2 > 0$ such that

$$\varepsilon^1 U_C e > \max\{A_C\bar{x} - CUB, 0\} \text{ and}$$
$$\varepsilon^2 U_G e > \max\{A_G\bar{x} - GUB, 0\}.$$

We now have that $(\bar{x}, \varepsilon^1 e, \varepsilon^2 e)$ is a feasible element of LP_2.

Since lower bounds for LP_1 and LP_2 are used below, instead of showing the duals of LP_1 and LP_2 are feasible (which is readily accomplished as in the proof of Lemma 2), we show that the objective functions in LP_1 and LP_2 are bounded below. In both of these linear programs we have

$$-CUB \le U_C\beta \text{ and } 0 \le U_G\gamma.$$

The assumption that both U_C and U_G have full column rank and are semimonotone implies that

$$\beta \ge -U_C^+ CUB \text{ and } \gamma \ge U_G^+ 0 = 0. \tag{6}$$

To see this suppose that H is a semimonotone matrix with full column rank, and consider the inequalities $Hv \geq h$. Then, $Hv - h = w \geq 0$, and the full column rank of H implies that $v = H^+w + H^+h$. Since H is semimonotone, $H^+ \geq 0$, and hence $v \geq H^+h$. From the inequalities in (6) we now have for both LP_1 and LP_2 that

$$u_C^T\beta + u_C^T\gamma \geq -u_C^TU_C^+CUB. \tag{7}$$

■

From Theorem 1 there is positive scalar κ such that $l^T\alpha \leq \kappa/\omega$, which is useful because an upper bound on κ is easily found. Let $(\hat{x}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ and $(\bar{x}, 0, \bar{\beta}, \bar{\gamma})$ be optimal solutions to LP_1 and LP_2, respectively. From (4) we have that

$$\omega \cdot l^T\alpha^*(\omega) \leq u_C^T\bar{\beta} + u_C^T\bar{\gamma} - \left(u_C^T\hat{\beta} + u_G^T\hat{\gamma}\right). \tag{8}$$

The inequality in (7) implies that

$$-\left(u_C^T\hat{\beta} + u_G^T\hat{\gamma}\right) \leq u_C^TU_C^+CUB.$$

So, all that remains is to provide an upper bound on $u_C^T\bar{\beta} + u_C^T\bar{\gamma}$. Notice that the maximum radiation allowed by a sub-beam is no more than

$$\imath \equiv \frac{\|TUB\|_\infty}{\min\limits_{(p,a,i)} \{A_{(p,a,i)} : A_{(p,a,i)} \neq 0, p \text{ is a tumor pixel}\}}.$$

Hence, the components of any plan attaining tumor uniformity are bounded above by $\imath$, from which we conclude that $\bar{x} \leq \imath e$. Set

$$\tilde{\beta} = \frac{\|A_C(\imath e) - CUB\|_\infty}{\underline{\text{rs}}(U_C)}e \text{ and } \tilde{\gamma} = \frac{\|A_G(\imath e) - CUB\|_\infty}{\underline{\text{rs}}(U_G)}e,$$

so that

$$\begin{aligned} A_C\bar{x} &\leq A_C(\imath e) \leq CUB + U_C\tilde{\beta} \text{ and} \\ A_G\bar{x} &\leq A_G(\imath e) \leq GUB + U_G\tilde{\gamma}. \end{aligned}$$

We now see that $(\imath e, \tilde{\beta}, \tilde{\gamma})$ is feasible to LP_2 and $u_C^T\tilde{\beta} + u_G^T\tilde{\gamma} \geq u_C^T\bar{\beta} + u_G^T\bar{\gamma}$. Using that $u_C^Te = \|u_C\|_1$ and $u_G^Te = \|u_G\|_1$, we have that the inequality in (8) now implies that $l^T\alpha^*(\omega)$ is no greater than

$$\frac{1}{\omega}\left(\frac{\|A_C(\imath e) - CUB\|_\infty\|u_C^T\|_1}{\underline{\text{rs}}(U_C)} + \frac{\|A_G(\imath e) - GUB\|_\infty\|u_G^T\|_1}{\underline{\text{rs}}(U_G)} + u_C^TU_C^+CUB\right) \equiv \frac{\kappa'}{\omega}.$$

Although κ' contains a Moore-Penrose generalized inverse, which is in general computationally expensive, calculating κ' for either average or absolute analysis

reduces to the much easier situation of finding the norms indicated. This follows because in the average analysis case U_C is the identity matrix, and in the absolute analysis case U_C is a column vectors of ones, which implies that $U_C^+ = \frac{1}{m_C} U_C^T$.

Recall that TG was the goal dose for the tumorous region and that we originally set $TLB = (1 - tol)TGe$. To utilize the upper bound provided by κ', we slightly increase this lower bound and instead set $TLB = (1-tol)TGe+\varepsilon e$. After calculating κ', we choose $\omega = \kappa'/\varepsilon$ and solve $EM(\omega)$. Theorem 1 now implies that the optimal value of $l^T\alpha$ is less than ε, and hence the sought after uniformity is guaranteed. So using only the optimal objective value, we have from Theorem 1 the analysis found in Figure 2. Of course a more detailed interpretation of the solution is possible by examining the individual components of $(\alpha^*(\omega), \beta^*(\omega), \gamma^*(\omega))$.

Before discussing our solution technique, we offer another interpretation of the objective function. The terms $\omega \cdot l^T\alpha$ and $(u_C^T\beta + u_G^T\gamma)$ compete in the following way. As $L\alpha \downarrow 0$ low dosage plans are removed from the feasible region, and as $(U_C\beta, U_G\gamma) \downarrow (-CUB, 0)$ high dosage plans are removed from the feasible region. There is the possibility of having feasible plans that simultaneously achieve $L\alpha = 0$, $U_C\beta = -CUB$, and $U_G\gamma = 0$. Such plans are called the *unencumbered plans* and include only those sub-beams intersecting the tumor and not intersecting a critical region. However, there are often not enough sub-beams striking only the tumor, and hence no unencumbered plans. Schematics showing how $L\alpha$ and $(U_C\beta, U_C\gamma)$ compete are found in Figures 3 and 4.

5 Solution Technique

As mentioned in Section 1, many different solution techniques have been used to design radiotherapy plans. Linear solvers have had the problem that physician limits are often attained [13], meaning that therapy plans narrowly adhere to the prescription. The problem here is that simplex algorithms produce an extreme point solution, and a therapy plan found using this approach is, in a very real sense, an extreme plan. However, with the number of sub-beams contained in a plan being no greater than the number of constraints, simplex solvers do have the advantage of producing plans with a controlled number of sub-beams [17]. Such control is accomplished by aggregating constraints to represent integral doses over a region, which reduces the number of constraints and subsequently the number of sub-beams. Such treatment plans are often clinically desirable because of technological limitations and time constraints. With many treatment systems the number of active beams is usually held to no more than 4 or 5. However, aggregating constraints has the disadvantage of allowing for inappropriately high levels of radiation.

Limiting the number of active beams is no longer as important as it once was because of a new treatment paradigm called *computer automated collimation*. With this type of treatment system the beam of radiation is continuously shaped as the gantry rotates around the patient. In fact, with computer automated collimation

Interpreting The Solution: Average Analysis
$(l = e, u_C = e, u_G = e)$

[**Case 1:** $l^T\alpha^*(\omega) > \varepsilon$] We conclude that on average the prescription does **not** allow tumor uniformity.

[**Case 2:** $l^T\alpha^*(\omega) \leq \varepsilon$] We conclude that on average the prescription does allow tumor uniformity. This situation contains two important sub-cases.

[**Case 2a:** $u_C^T\beta^*(\omega) + u_G^T\gamma^*(\omega) > 0$] The conclusion here is that an average tumor uniformity is achievable, but only at the expense of some of the non-tumorous tissue receiving more radiation than desired.

[**Case 2b:** $u_C^T\beta^*(\omega) + u_G^T\gamma^*(\omega) \leq 0$] The conclusion is that an average tumor uniformity is allowed, and at the same time the average amount of radiation over the non-tumorous tissue is at least as good as desired.

Interpreting The Solution: Absolute Analysis
$(l = u_C = u_G = 1)$

[**Case 1:** $\alpha^*(\omega) > \varepsilon$] We conclude that the prescription does **not** allow tumor uniformity.

[**Case 2:** $\alpha^*(\omega) \leq \varepsilon$] We conclude that the prescription does allow tumor uniformity. This situation contains two important sub-cases.

[**Case 2a:** $u_C^T\beta^*(\omega) + u_G^T\gamma^*(\omega) > 0$] The conclusion here is that tumor uniformity is achievable, but only at the expense of some of the non-tumorous tissue receiving more radiation than desired.

[**Case 2b:** $u_C^T\beta^*(\omega) + u_G^T\gamma^*(\omega) \leq 0$] The conclusion is that tumor uniformity is allowed, and at the same time the amount of radiation over the non-tumorous tissue is at least as good as desired.

Figure 2: Interpreting the solution for either average or absolute analysis

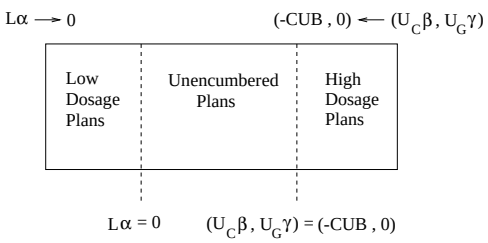

Figure 3: This schematic indicates that the tumor dose restrictions are attainable with zero dose to critical structures. Such plans are called the unencumbered plans.

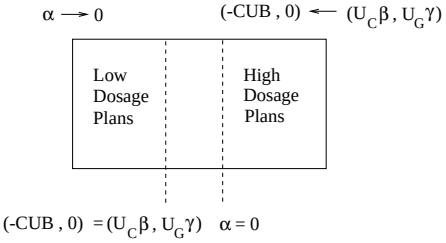

Figure 4: This diagram shows that when critical structures are required to receive radiation, there are no unencumbered plans.

there are no clinical limitations that need to be incorporated other than the fact that the gantry must rotate around the patient within a plane. However, we design plans based on co-planer images, and we "stack" these plans to achieve "thickness". For example, suppose a treatment plan is being designed for a spherical tumor with a diameter of 9mm. Each co-planer image represents a swath of the patient's anatomy that is about 3mm thick. For each of the 3 or 4 patient images that contain the tumor we design a plan, and the patient is treated with each of the 3 or 4 co-planer plans.

Because of the capability of administering intricate treatment plans, it now seems natural to find a plan without consideration of clinical limitations, and then fit this solution to the capabilities of the treatment facility. This is the perspective from which we approach the problem. Below we discuss why most interior point algorithms produce plans that are appropriate for the treatment procedure just described. The process of adapting intricate plans for use in older treatment facilities is not addressed and is the subject of current research efforts.

In 1979, Khachiyan showed that the class of linear programming problems is solvable in polynomial time, and the algorithm that Khachiyan used to demonstrate this property was an *interior point* method [14]. Since the middle 1980s, the study of interior point algorithms has been one of the most productive and intriguing subjects in mathematics. Interested readers are referred to the texts of Roos, Terlaky, and Vial [26], Wright [33], and Ye [34]. Most interior point algorithms converge to a unique solution known as the *analytic center solution*, and only when there is a unique optimal solution is the analytic center solution a basic optimal solution. Moreover, this solution has characteristics that are different from basic optimal solutions and more appropriate for radiotherapy planning.

The analytic center solution is the unique limit of a geometric structure called the *analytic central path*, which is contained in the strict interior of the feasible set.

Geometrically, path following interior point algorithms follow the analytic central path towards optimality, and hence, converge towards the analytic center solution. A fundamental result in the theory of interior point algorithms is that the analytic central path exists if, and only if, the strict interiors of both the primal and dual problems are non-empty. From Lemma 2 we have that path following interior point algorithms are amenable to $EM(\omega)$.

The analytic center solution is a *strictly complementary solution*, which means that each complementary pair of primal and dual variables contains a single zero. The fact that every linear program has a strictly complementary solution was first proven in 1956 [11], and such solutions induce what is known as the *optimal partition*. This partition is maximal [3] and divides the primal constraints into two categories, those that are allowed to be strictly satisfied at optimality (indexed by the set B) and those that must hold as equalities at optimality (indexed by the set N). Notice that this means the optimal set, denoted by $\mathcal{P}^*(\omega)$, is equal to the subset of $\mathcal{P}$ with the inequalities indexed by N replaced with equalities.

For $EM(\omega)$ we now have that a strictly complementary solution is provided by a path following interior point algorithm, and that such a solution indicates, relative to optimality, everywhere that strict adherence to the prescription is allowed. For example, suppose that the set of unencumbered plans is non-empty. Then, any plan satisfying

$$TLB \leq A_T x \leq TUB, \;\; A_C x = 0, \;\; A_G x \leq GUB, \text{ and } \;\; x \geq 0$$

is optimal. Suppose that x^* is the plan provided by the analytic center solution of $EM(\omega)$. We are now assured that if $x^*_{(a,i)} = 0$, the i^{th} sub-beam of angle a is never used in an unencumbered plan. For any pixel p we similarly have that if either $(A_T x^*)_p = TLB_p$, $(A_T x^*)_p = TUB_p$, or $(A_G x^*)_p = GUB_p$, the radiation level delivered to pixel p must attain the corresponding prescribed bound for every unencumbered plan. Alternatively, x^* is a plan that is strictly better than the prescription everywhere that is possible. We now discuss why the analytic center solution is preferred among all plans with these characteristics.

A result first shown by McLinden [20, 21] was that the analytic center solution, henceforth denoted as $(x^c(\omega), \alpha^c(\omega), \beta^c(\omega), \gamma^c(\omega))$, is the analytic center of $\mathcal{P}^*(\omega)$. Let (B_T, B_C, B_G, B_x) be a partition of B, where the subsets correspond to the constraints pertaining to the tumor, the critical structures, the remaining non-tumorous tissue, and the plan x, respectively. With this notation, the fact that $(x^c(\omega), \alpha^c(\omega), \beta^c(\omega), \gamma^c(\omega))$ is the analytic center of $\mathcal{P}^*(\omega)$ is equivalent to saying that it is the unique solution to

$$\begin{aligned}
\max \Big\{ & \sum_{i \in B_C} (\ln(CUB + U_C\beta - A_C x)_i + \ln(U_C\beta + CUB)_i) \\
& + \sum_{i \in B_T} (\ln(TUB - A_T x)_i + \ln(A_T x + L\alpha - TLB)_i + \ln(TLB - L\alpha)_i + \ln(L\alpha)_i) \\
& + \sum_{i \in B_G} (\ln(GUB + U_G\gamma - A_G x)_i + \ln(U_G\gamma)_i) + \sum_{i \in B_x} \ln x_i : (x, \alpha, \beta, \gamma) \in (\mathcal{P}^*)^o \Big\}.
\end{aligned}$$

Hence, the analytic center solution is a *maximum entropy* solution, relative to the restriction that the accumulative total of insufficient tumor dose and non-tumorous tissue over dose is already minimized. Similar maximum entropy radiotherapy plans were discussed by Sandham, Yuan, and Durrani [28]. However, they use the $\max\{f_i \ln(f_i)\}$ definition of maximum entropy and we use the $\max\{\ln(f_i)\}$ definition. More importantly, their objective is to find a maximum entropy plan relative to the non-elastic linear constraints. This means that they are assuming that the physician's treatment aspirations are attainable. Additionally, a solution to their model will always have the critical structures receiving radiation, even if this is not required. Our objective is to minimize undesirable deviations from the prescription, and the fact that we end up with a maximum entropy solution is a byproduct of the path-following interior point algorithm.

The most important observation to make is that relative to optimality, the analytic center solution provides a plan that strictly adheres to the prescription as much as possible. This is not to say that each prescription bound is satisfied as much as possible, but rather that the sum of the logarithms of the amounts of strict adherence is maximized. So, there is the possibility that some prescribed limits are nearly attained so that other prescribed limits are even more satisfied. In general, the analytic center solution is different from other centers, such as the Barry center and center of mass, and unfortunately does not rely on the geometry of the optimal set. This is because the analytic center solution depends on the manner in which the optimal set is represented. Recent investigations into this dependence are found in [5], where it is shown that the analytic center solution is invariant over all representations without redundant inequalities.

We end this section with a discussion of how rope and chain organs are handled. As noted in Section 3, chain organs lend themselves to deposition patterns that are uniform, and the idea is to spread the delivered energy evenly throughout the tissue. The just described nature of the analytic center solution tends to exhibit exactly this type of behavior. So, no modeling adjustments are warranted when dealing with chain organs. However, rope organs are susceptible to complete failure with a uniform dose. Instead, these organs permit rather large doses to fractions of their tissue, and we want to "clump" the delivered energy into a small band through the organ. When dealing with a rope organ we aggregate the constraints concerning this organ into a single constraint. For example, suppose that the liver, which is a rope organ, is designated as a critical structure, and that $A_{C'}$ is the submatrix of A_C whose rows correspond to the liver pixels. The physician prescribes an upper bound of 50 Gy on the liver cells and tags this structure as a rope organ. Instead of using the collection of constraints, $A_{C'}x \leq 50e$, we use the single constraint, $e^T A_{C'} x \leq 50 e^T e$. While we have no guarantee that this manipulation will have the desired effect, the idea is to allow greater levels of radiation for some cells in an attempt to have many other cells receive little or no radiation. This technique is not original and as already discussed has been used to control the number of beams in a plan.

6 Experimentation

A prototype treatment system called $\mathcal{R}$adiotherapy optim$\mathcal{A}$l $\mathcal{D}$esign, or $\mathcal{RAD}$, has been developed using MATLAB©. This system is available from `http://www.trinity.edu/aholder/research/oncology/`, and requires MATLAB's optimization toolbox. As with most software development, the step from theory to implementation was not trivial. We begin this section with a brief discussion of some of these issues.

First, the use of κ' as indicated in Section 4 leads to numerical instability because this value is often too large. We have instead found that setting $\kappa' = \|TLB\|_\infty$ works well. Second, $\mathcal{RAD}$ currently uses a 64×64 grid, and allows angles evenly spaced at every 15, 5, or 1 degree(s), with each beam being comprised of 10, 32, or 32 pencils, respectively. If each pixel were restricted, the problem is routinely too large for the linear programming solvers in MATLAB's optimization toolbox. Most of the constraints are from non-tumorous, non-critical structure pixels –i.e. the largest group of constraints is $A_{_G} x \leq GUB + U_G\gamma$. These constraints are included so that the otherwise unrestricted tissue does not receive an unusually high dose of radiation. We have found that MATLAB's solvers work well when instead of restricting each of the 4096 pixels, the user restricts only regions where hot spots are likely to occur. Moreover, most problems are solvable within a few minutes, and as such, an initial solve with $A_{_G}$ vacuous indicates whether hot spots are even of concern. If they are, the user simply restricts the appropriate regions and re-solves the problem.

In addition to allowing the user to choose from different angle geometries, $\mathcal{RAD}$ has the following features.

- Either absolute or average analysis is allowed.
- A prescription window allows the user to easily set the tissue type, the prescription levels, and the tumor uniformity level.
- A simplex based solver is available.
- After the optimization routine is complete, three figures are presented. The first and second figures are a contour plot and a 3-D image of the radiation levels delivered by the plan. The third figure provides an explanation of the solution that depends on whether absolute or average analysis was chosen.

In the examples that follow there were 360 beams equally spaced every degree, each beam containing 32 sub-beams. The amount by which TLB is increased is internally set at 10^{-4}. The problems were solved on a 500 MHz PC with 384 M of RAM. Because the simplex solver does not employ sparse matrix routines, we were not able to use the simplex based solver on the examples below.

The first example is of a tumor that has half-way grown around a critical structure. The tumoricidal dose was 80 Gy, and the critical structure pixels were restricted to 30 Gy. The tumor uniformity was set at 2%, and no other pixels were

restricted (A_G is empty). Using absolute analysis, we found the optimal plan depicted in Figures 5 and 6 in 62.50 seconds. This plan has a maximum and minimum tumor dose of 78.42Gy and 81.56Gy, respectively, and the critical structure receives no radiation. Hence, this is an unencumbered plan where every tumor constraint is strictly satisfied.

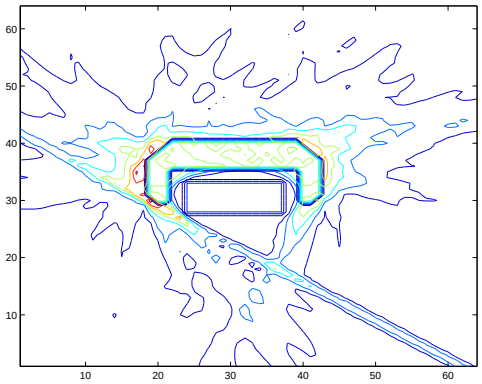

Figure 5: A contour plot showing how the deposition pattern of the plan "bends" around the critical structure.

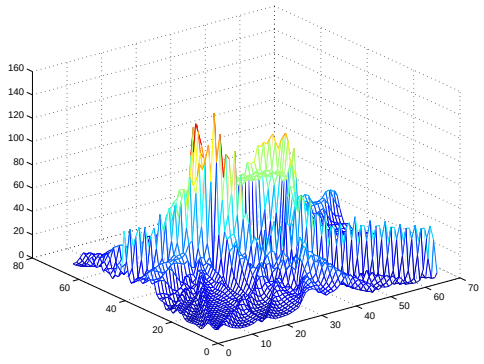

Figure 6: The vertical height is the level of radiation delivered by the plan over the image. The high peaks indicate hot spots.

In this example there are several small regions outside the critical structure and the tumor that receive more than 90 Gy, which is an unacceptably high amount. Notice that these hot spots are precisely where one would expect because the beams that appear most natural for this problem intersect over these regions. If we restrict the pixels surrounding the tumor and the critical structure to 85 Gy –i.e. include the pixels surrounding the original image in $A_G x \leq GUB + U_G\gamma$, we find that the hot spots are eliminated. The resulting plan is illustrated in Figures 7 and 8. The new problem was solved in 200.72 seconds, so the increase in overall time was nominal. The levels of radiation over the tumor are well within 80 Gy $\pm$2% with the maximum and minimum dose being 78.42 Gy and 81.57 Gy, respectively. Once again, the critical structure receives no radiation, and hence the plan is unencumbered.

The next example has a tumor surrounded by four critical structures. The tumoricidal dose is again set at 80 Gy, the critical structures on the the left and right are restricted to no more than 50 Gy, and the critical structures on the top and bottom are restricted to 30 Gy. The tumor uniformity was set at 10%. The absolute analysis plan had a solution time of 188.24 seconds, and has a minimum and maximum tumor dose of 72.04 Gy and 87.93 Gy. Each critical structure is guaranteed to be under its prescribed upper bound by at least 30 Gy, which means that no radiation is deposited into the critical structures on the top and bottom.

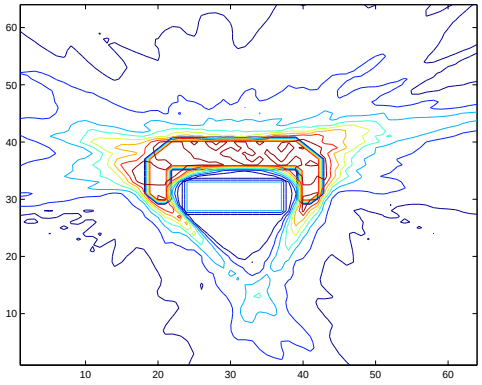

Figure 7: The contour plot of the plan found after restricting the regions where the hot spots occurred.

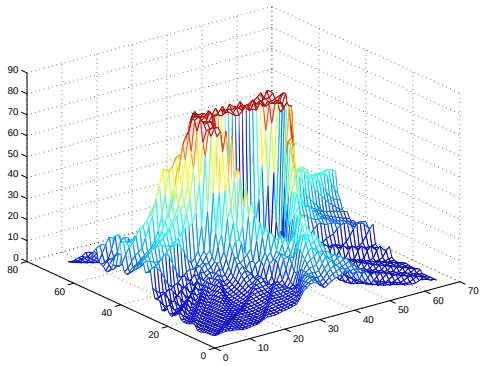

Figure 8: The radiation levels for the plan found after restricting the hot spot regions.

There are no hot spots, with every pixel in the image receiving less than $1.1 * \max_i\{TUB_i\}$ Gy $= 96.8$ Gy. This plan is shown in Figure 9.

Using the same situation just described, we developed a plan with average analysis that has an average tumor dose of 78.89 Gy. The critical structures are on average under their prescribed amount by 38.55 Gy. The plan took 220.28 seconds to design. From Figure 10, we see that the plan developed with average analysis tends to clump the energy into a few beams. This is intuitive because the average amount of radiation deposited in the critical structures is significantly less when only a few critical structure pixels receiving a moderate amount of radiation (the vast majority receive zero). This is in contrast to the absolute analysis case where reducing the maximum amount of radiation a critical structure receives tends to distribute the delivered energy among several beams.

7 Conclusions

We have shown that the model $EM(\omega)$ is theoretically sound and practical in use. From the analysis in Section 4 we know exactly how to make relevant statements about the solution returned by the optimization routine. The information provided is easily understandable, and hence treatment planners need not be mathematical programmers to understand the solution characteristics. The implementation discussed in Section 6 is easy to use, and problems are solved in a time frame that allows treatment planners to alter and re-solve problems.

In the future we wish to enlarge the grid size to the more realistic size of 1024×1024. This will most likely entail building a link to an interior point solver exterior to MATLAB. There is also the need to transform the plans provided by $\mathcal{RAD}$ into

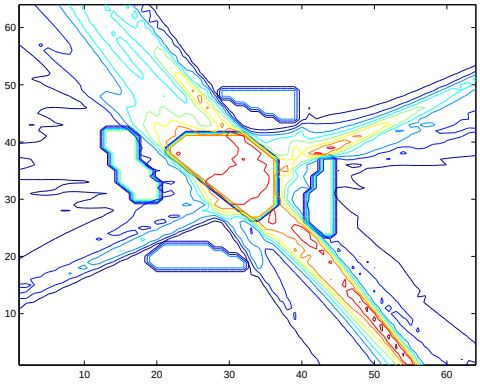

Figure 9: A plan found using absolute analysis.

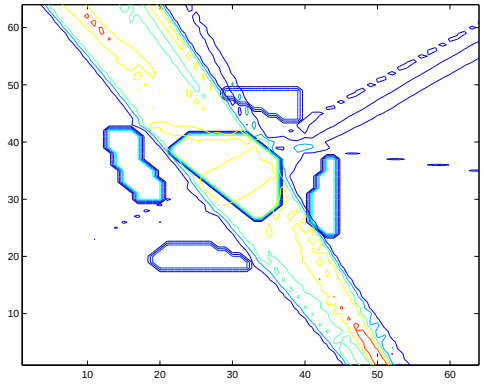

Figure 10: A plan found using average analysis.

plans that are implementable by older treatment facilities. This means that angles from a plan must be pruned, and that the sub-beam intensities must be altered to conform to paradigms of shaping the radiation beam other than collimation (most treatment facilities use variable degree wedges to shape the beam).

The author joins the sentiments of Shepard, Ferris, Olivera, and Mackie [29] when they say "It is our hope that the community of optimization experts will be able to offer further insights that will improve our ability to solve these difficult problems." Indeed, radiotherapy treatment planning is a research area where a sustained dialog between medical physicists and optimization experts could positively effect many lives.

Acknowledgments

The author would like the thank Francis Newman for the many long discussions about the clinical side of this problem. He would also like the thank the Math and Engineering Research Group (MERG) at Trinity University for being a supportive and attentive group during early presentations of this material.

References

[1] M. Altschuler, Y. Censor, and W. Powlis. Feasibility and optimization methods in teletherapy planning. In *Advances in Radiation Oncology Physics: Dosimetry, Treatment Planning, and Brachytherapy*, number 19 in Medical Physics Monograph, pages 1022–1057. American Institute of Physics, Woodbury, NY, 1992.

[2] G. K. Bahr, J. G. Kereiakes, H. Horwitz, R. Finney, J. Galvin, and K. Goode. The method of linear programming applied to radiation treatment planning. *Radiology*, 91:686–693, 1968.

[3] A. Berkelaar, C. Roos, and T. Terlaky. The optimal set and optimal partition approach to linear and quadratic programming. In T. Gal and H. Greenberg, editors, *Recent Advances in Sensitivity Analysis and Parametric Programming*, chapter 6. Kluwer Academis Publishers, Boston, MA, 1997.

[4] A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York, NY, 1979.

[5] R. Caron, H. Greenberg, and A. Holder. Analytic centers and repelling inequalities. Technical Report CCM 142, Center for Computational Mathematics, University of Colorado at Denver, Denver, CO, 1999.

[6] Y. Censor, M. Altschuler, and W. Powlis. A computational solution of the inverse problem in radiation-therapy treatment planning. *Applied Mathematics and Computation*, 25:57–87, 188.

[7] Y. Censor, M. Altschuler, and W. Powlis. On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning. *Inverse Problems*, 4:607–623, 1988.

[8] A. Cormack and E. Quinto. The mathematics and physics of radiation dose planning using x-rays. *Contemporary Mathematics*, 113:41–55, 1990.

[9] I. Das and J. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural Optimization*, 41(1):63–69, 1997.

[10] M. Goitein and A Niemierko. Biologically based models for scoring treatment plans. Scandanavian Symposium on Future Directions of Computer-Aided Radiotherapy, 1988.

[11] A. Goldman and A. Tucker. Theory of linear programming. In H. Kuhn and A. Tucker, editors, *Linear Inequalities and Related Systems*, volume 38, pages 53–97. Princeton University Press, Princeton, New Jersey, 1956.

[12] H.J. Greenberg. *Mathematical Programming Glossary*. World Wide Web, `http://www.cudenver.edu/~hgreenbe/glossary/glossary.html`, 1997-99.

[13] L. Hodes. Semiautomatic optimization of external beam radiation treatment planning. *Radiology*, 110:191–196, 1974.

[14] L. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244:1093–1096, 1979.

[15] M. Langer, R. Brown, M. Urie, J. Leong, M. Stracher, and J. Shapiro. Large scale optimization of beam weights under dose-volume restrictions. *International Journal of Radiation Oncology, Biology, Physics*, 18:887–893, 1990.

[16] J. Legras, B. Legras, and J. Lambert. Software for linear and non-linear optimization in external radiotherapy. *Computer Programs in Biomedicine*, 15:233–242, 1982.

[17] W. Lodwick, S. McCourt, F. Newman, and S. Humphries. Optimization methods for radiation therapy plans. In C. Borgers and F. Natterer, editors, *IMA Series in Applied Mathematics - Computational, Radiology and Imaging: Therapy and Diagnosis*. Springer-Verlag, 1998.

[18] J. Leong M. Langer. Optimization of beam weights under dose-volume restrictions. *International Journal of Radiation Oncology, Biology, Physics*, 13:1255–1260, 1987.

[19] S. McDonald and P. Rubin. Optimization of external beam radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, 2:307–317, 1977.

[20] L. McLinden. An analogue of Moreau's proximation theorem, with applications to the nonlinear complementary problem. *Pacific Journal of Mathematics*, 88(1):101–161, 1980.

[21] L. McLinden. The complementarity problem for maximal monotone multifunctions. In R. Cottle, R. Giannessi, and J. Lions, editors, *Variational Inequalities and Complementarity Probelems*, pages 251–270. John Wiley & Sons, 1980.

[22] S. Morrill, R. Lane, G. Jacobson, and I. Rosen. Treatment planning optimization using constrained simulated annealing. *Physics in Medicine & Biology*, 36(10):1341–1361, 1991.

[23] S. Morrill, I. Rosen, R. Lane, and J. Belli. The influence of dose constraint point placement on optimized radiation therapy treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, 19:129–141, 1990.

[24] W. Powlis, M. Altschuler, Y. Censor, and E. Buhle. Semi-automatic radiotherapy treatment plannnig with a mathematical model to satisfy treatment goals. *International Journal of Radiation Oncology, Biology, Physics*, 16:271–276, 1989.

[25] C. Raphael. Mathematical modeling of objectives in radiation therapy treatment planning. *Physics in Medicine & Biology*, 37(6):1293–1311, 1992.

[26] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization: An Interior Point Approach*. John Wiley & Sons, New York, NY, 1997.

[27] I. Rosen, R. Lane, S. Morrill, and J. Belli. Treatment plan optimization using linear programming. *Medical Physics*, 18(2):141–152, 1991.

[28] W. Sandham, Y. Yuan, and T. Durrani. Conformal therapy using maximum entropy optimization. *International Journal of Imaging Systems and Technology*, 1995:80–90, 1995.

[29] D. Shepard, M. Ferris, G. Olivera, and T. Mackie. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, 41(4):721–744, 1999.

[30] D. Sonderman and P. Abrahamson. Radiotherapy treatment design using mathematical programming models. *Operations Research*, 33(4):705–725, 1985.

[31] H. Withers, J. Taylor, and B. Maciejewski. Treatment volume and tissue tolerance. *International Journal of Radiation Oncology, Biology, Physics*, 14:751–759, 1987.

[32] A. Wolbarst. Optimization of radiation therapy II: The critical-voxel model. *International Journal of Radiation Oncology, Biology, Physics*, 10:741–745, 1984.

[33] S. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, PA, 1997.

[34] Y. Ye. *Interior Point Algorithms Theory and Analysis.* John Wiley & Sons, Inc., New York, NY, 1997.