Chapter 1

RADIOTHERAPY TREATMENT DESIGN AND LINEAR PROGRAMMING

Allen Holder

Trinity University

Department of Mathematics

715 Stadium Drive

San Antonio, TX 78212

http://www.trinity.edu/aholder/

aholder@trinity.edu

Abstract Intensity modulated radiotherapy treatment (IMRT) design is the process of choosing how beams of radiation will travel through a cancer patient to treat the disease, and although optimization techniques have been suggested since the 1960s, they are still not widely used. Instead, the vast majority of treatment plans are designed by clinicians through trial-and-error. Modern treatment facilities have the technology to treat patients with extremely complicated plans, and designing plans that take full advantage of the technology is tedious. The increased technology found in modern treatment facilities makes the use of optimization paramount in the design of successful treatment plans. The goals of this work are to 1) present a concise description of the linear models that are under current investigation, 2) develop the analysis certificates that these models allow, and 3) foreshadow future research avenues.

1.   Introduction

Fast proliferating cells, such as those found in cancerous and displasiac tissue, are more sensitive to radiation than healthy cells, and this fact has allowed tremendous strides in the fight against cancer. Chemotherapy uses this property by injecting radioactive substances into the blood

2

stream, with the goal being to administer enough radiation to kill the cancerous cells but not enough to kill the healthy cells. Because the substances are injected into the blood stream, chemotherapy treats the entire anatomy, and hence, all fast proliferating cells are attacked (such as hair cells). Intensity modulated radiotherapy treatment (IMRT) is a similar cancer treatment where external beams of radiation are focused on the cancerous regions. Since the radiation is not injected into the blood stream, IMRT is a local treatment. In fact, the radiation beams can be focused within sub-millimeter precision, giving a medical physicist precise control of how the radiation travels through the anatomy.

IMRT design is the process of deciding how the beams of radiation will travel through the patient so that they deliver a tumoricidal dose of radiation to the cancerous region. At the same time, the critical structures surrounding the cancer are to receive a limited dose of radiation so that they can survive the treatment. It is the struggle between the tumor, which we want to receive a high level of radiation, and the critical structures, which we do not want to receive a high level of radiation, that makes IMRT design a complicated process. Moreover, this process is further complicated because modern treatment facilities have the technology to deliver extremely intricate treatment plans. The amount of flexibility that is permitted makes optimizing treatment plans beyond the scope of human comprehension, and future technology will only increase the degree of complication. If our planning process does not advance with the technology, the benefits of the new technology will not be realized, and patients will not receive the added benefits that the advanced technology allows. So, the development of appropriate optimization models that can take full advantage of the new technology is critical.

There are three groups of specialists that are important to the success of improved treatment design: 1) the oncologists, who attend to the needs of the patients, 2) the medical physicists, who know how to model the deposition of radiation, and 3) the operational researchers, who are experts in the field of applied optimization. One of the difficulties of working in this field as an operational researcher is to find an oncologists and/or a medical physicist to work with, for a continued dialog between these three groups is important. Historically, the bulk of the research was accomplished by the oncologists and the medical physicists, and only in the last few years have operational researchers become interested in these problems. Because optimization is playing an ever increasing role in treatment design, it makes sense that operational researchers need to be included, as they are accustomed to investigating algorithms and performing solution analysis. The goal of this

chapter is to encapsulate the modern research directions so that operational researchers can quickly become familiar with the interdisciplinary field of designing IMRT plans. Before we continue, we mention two Internet resources. The *Operations Research & Radiation Oncology* web site at `http://www.trinity.edu/aholder/HealthApp/oncology/` has a depository of recent papers and a list of interested researchers. The other resource is *PubMed*, the index to the National Library of Medicine, located at `http://www.ncbi.nlm.nih.gov/`. A recent search on "optimization" and "radiotherapy" found 905 related citations, so the medical literature is immense. Most of these articles are case studies about specific types of cancer and are not directly related to operations research. The bibliography at the end of this chapter is designed to get those who are interested started in the field.

2. Modeling Dose Deposition

To understand how and why linear optimization models appropriately model the planning of IMRT design, one needs to have a basic understanding of how radiation is deposited into the anatomy. The basic question is how does a focused beam of radiation deposit energy as it travels through a patient. The question has two perspectives. A forward problem is one in which we know the amount of energy being transmitted along the beam, and we want to know how much energy is deposited at a point in the anatomy. An inverse problem is when we know how much energy is to be deposited in the anatomy, and we want to know what beam energies attain the desired amounts. IMRT planning is an inverse problem because we limit the amount of radiation received by certain tissues and find a collection of beam intensities that adhere to these bounds. While a complete discussion of the physics describing the dose deposition is beyond the scope of this article, we briefly explain a continuous model and its discrete counterpart (see [Censor, 1991] and [Cormack and Quinto, 1990] for more complete details).

We begin our model development with a description of the equipment found in a standard treatment facility (see Figure 1.1). The beams of radiation are formed by a linear accelerator, and once formed they travel through a gantry that is capable of rotating around the patient (the center of the rotation is called the *isocenter*). The fact that the gantry can rotate around the patient is important because this allows the beam of radiation to be directed at the patient from any angle. The head of the gantry is designed to accommodate one of several focusing apparatuses, with most modern facilities using a *multileaf collimator* (see Figure 1.2). This device can 'shape' the beam of energy by blocking por-

Figure 1.1. A gantry is capable of rotating around the patient as he or she lies on the couch. The head of the gantry contains a multileaf collimator, see Figure 1.2
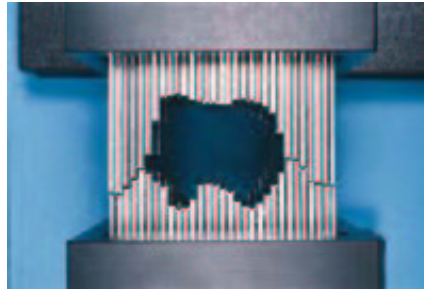
Figure 1.2. A multileaf collimator is used to 'shape' the beam of radiation so that surrounding tissues are shielded.

tions of the beam. Shaping the beam has tremendous benefits because it allows sensitive regions to receive relatively low levels of radiation, while nearby tumorous regions receive a higher amount of radiation. A planning model must take into consideration that the beams can be focused on the patient from any angle and in about any shape.

Consider the geometry depicted in Figure 1.3, where we want to calculate the dose at (r, θ). The gantry in this figure is located at angle a, and the sub-beam from angle a that passes through (r, θ) is i. The amount of energy that is to be transmitted along sub-beam (a, i) is $p(a, i)$. How far the cell is from the surface of the body, denoted by d in the diagram, affects the radiation dose received by the cell. This is because the beam *attenuates* as it travels through the body, meaning that it deposits more radiation when it first enters the body and 'decays' as it travels through the tissue. This attenuation is modeled as exponential decay –i.e. by $e^{-\mu d}$, where μ depends on the particular beam of energy formed by the linear accelerator. We now have that the radiation deposited at location (r, θ) by sub-beam (a, i) is $p(a, i)e^{-\mu d}$. To calculate the total, or *integral*, dose at point (r, θ) we need to accumulate the amount deposited from every possible sub-beam that passes through (r, θ). Allowing $L = \{(a, i) : \text{sub-beam } (a, i) \text{ passes through } (r, \theta)\}$, we find that the integral dose is

$$D(r, \theta) = \int_L p(a, i)e^{-\mu d} da.$$

Again, we point out that if we know $p(a, i)$ and want to calculate $D(r, \theta)$, then we are dealing with a forward problem. However, IMRT planning is
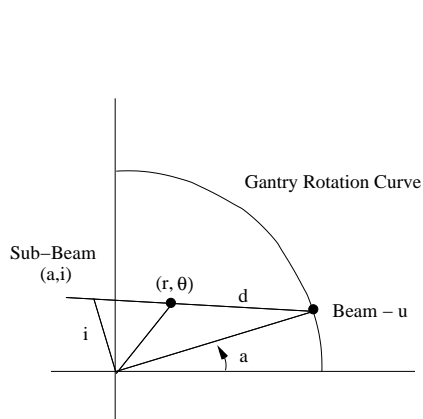
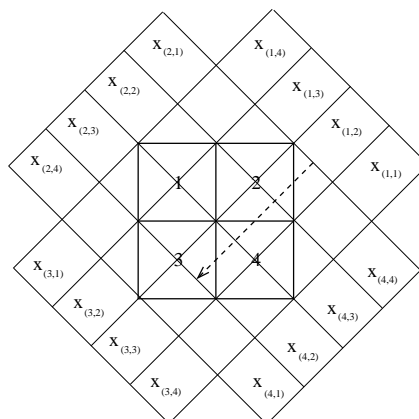*Figure 1.3.* The geometry of a continuous dose deposition calculation.

Figure 1.4. A discretized approximation to the continuous dose deposition calculation.

an inverse problem because we want to specify limits on the dose and use these limits to calculate the amount of energy to deliver along each sub-beam —i.e. we bound $D(r, \theta)$ and want to calculate $p(a, i)$ to satisfy the bound. So, for continuous IMRT planning we need to invert an integral transformation. There are a variety of techniques to accomplish this task, but the difficulty lies in the fact that the calculation of $p(a, i)$ must keep it non-negative, which is not guaranteed by these techniques. The hidden assumptions on the integrand are what often make inverse problems more difficult than forward problems.

In the continuous model we are integrating with respect to the angle a, so in the discrete model there are a fixed number of angles, denoted $a_1, a_2, \ldots, a_\Theta$. We assume that each angle is comprised of η sub-beams, which may be *elementary beams* or *pencils*, the difference being that pencils radiate from a point source and elementary beams run parallel to each other. Our development does not depend on whether pencils or elementary beams are chosen, only that there are a finite number of them. The patient image is divided into $N \times M$ pixels (or voxels in the 3D case), and we want to measure the amount of radiation that is deposited into each pixel. We let $x_{(a,i)}$ be the dose along the ith sub-beam of angle a, and $d_{(p,a,i)}$ be the distance from where sub-beam $x_{(a,i)}$ enters the image to where it reaches the center of 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 1.4 we have a 2×2 patient image surrounded by 4 angles, each with 4 sub-beams (in this

6

case they are elementary beams). The elementary beam corresponding to $x_{(1,2)}$ intersects one-half of pixel 3, and the distance to the center of this pixel along this elementary beam is $3\sqrt{2}/2$ (assuming that each pixel has a width of one). Hence, $A_{(3,1,2)} = \frac{1}{2}e^{3\sqrt{2}\mu/2}$. 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). Similarly, a *treatment plan*, or more succinctly a plan, is a non-negative vector x whose components are $x_{(a,i)}$, where the order corresponds to the columns of A. So, x is the vector of energies at the gantry, and the linear transformation $x \mapsto Ax$ deposits the radiation into the anatomy.

Let pixel p contain point (r, θ) from the continuous model. The dose calculation at (r, θ) is approximated by

$$D(r,\theta) = \int_L p(a,i)e^{-\mu d}da \approx \sum_{(a,i)} A_{(p,a,i)}x_{(a,i)} = [Ax]_p,$$

where the last notation indicates that the integral dose to pixel p is the pth component of Ax. We point out that both the continuous and the discrete models are linear in the energy transmitted along the sub-beams —i.e. the continuous model is linear in p and the discrete model is linear in x. Physical measurements show that the integral dose to a cell is a linear function of the amount of energy transmitted along the sub-beams. So, the linear models are not crude approximations, but rather they accurately measure how radiation is deposited into the anatomy. With that said, the linear operator $x \mapsto Ax$ that we use only approximates the dose deposition because it does not take into account the effects of scattering. The problem here is that some radiation 'bounces' off cells and scatters into areas where it was not intended. There are non-linear models that measure scattering, see [Bartolozzi et al., 2000] and [Cormack and Quinto, 1990], but once the scattering for a particular patient is understood, the dose to a cell is linear in the energy transmitted along the sub-beams. Because each patient is unique, the linear coefficients depend on the patient, and in a clinical setting these linear coefficients are decided during an initial planning appointment. For the purposes of this article, we use the technique discussed above to calculate the dose deposition matrix.

The rows of A are partitioned into the rows that represent the cancerous regions, the critical structures, and the remaining healthy tissue. This reordering is represented by the submatrices A_T, A_C, and A_G, as

indicated below,

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

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. 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 $A_G \in \mathbb{R}^{m_G \times n}$. Because radiation is measured in Grays (Gy), the right-hand sides of the constraints are given in units of Gy. Allowing e to be the vector of ones, where length is decided by the context of its use, we can guarantee that the tumor pixels receive $80Gy$, that the critical structures receive no more than $40Gy$, and that the remaining tissue receives no more than $90Gy$ by finding a non-negative x that satisfies

$$A_T x \geq 80e, \ A_C x \leq 40e, \text{ and } A_G x \leq 90e.$$

If these were the only treatment goals, the design process is a feasibility problem, meaning that any non-negative vector satisfying these constraints would be a suitable plan, see [Censor et al., 188] and [Powlis et al., 1989].

3. Treatment Concerns

The primary goal of IMRT design is to construct a treatment plan that delivers a tumoricidal dose to the cancerous region and at the same time delivers low enough radiation levels to the surrounding tissues so that they maintain functionality. However, there are several issues that make this overriding objective difficult to translate into an optimization model. When first presented with the problem, most operational researchers believe that the objective function should be to deliver as much radiation as possible to the tumor. This naively makes sense because killing the cancerous cells is the purpose of the treatment. There are two reasons why maximizing the amount of radiation deposited into the tumor is not an appropriate objective. First, healthy and cancerous cells are often interspersed, and there is a limited range of radiation that will kill a cancerous cell and allow a healthy cell to survive. So, it is important to deliver enough radiation to kill the cancerous cells, but not so much radiation that the dose also kills the healthy cells within the tumor. This is usually accomplished by the dosimetrist stating that he or she wants the cancerous regions to attain a specified amount of radiation plus or minus some percentage. For example, the tumor should

receive $80Gy \pm 2\%$ means that the tumor should receive between $78.4Gy$ and $81.6Gy$. Second, if any region of the anatomy receives an unreasonably high amount of radiation, the cells within this region are killed. If the area is large enough, the human physiology is disrupted, causing a condition known as neucrosis. For these two reasons, it is paramount that the tumor receives a uniform, tumoricidal dose of radiation and not simply as much as possible.

Treatment planning is further complicated by the fact that different organs react to radiation in different ways. For example, the liver can receive a large amount of radiation over a substantial portion of its tissue and maintain its functionality. However, if the entire liver receives a relatively low dose of radiation, the organ will fail. The colon is in stark contrast because it can handle a relatively high uniform dose, but the organ will fail if a small region receives a high dose. Organs like the liver that can successfully receive a high level of radiation over a portion of their tissue, but fail under a relatively low, uniform dose, are called *rope* organs. A *chain* organ is one that can handle a relatively high amount of radiation over its entirety but will fail when a small amount of the tissue is destroyed (see [Goitein and Niemierko, 1988; Raphael, 1992; Withers et al., 1987; Wolbarst, 1984] for more complete details on rope and chain organs). So, in addition to making sure that the tumor receives a uniform, tumoricidal dose, the dosimetrist must make sure that the treatment plan delivers radiation to the critical structures in a suitable manner.

The hope that every patient receiving IMRT is cured of cancer is unrealistic, and because of this, patients and physicians must routinely make difficult decisions about a course of treatment. The best of all situations is when the type and stage of cancer being treated has a high probability of cure with standard treatments. The "best" treatment plan in such a case is one that delivers a tumoricidal dose to the cancerous regions and as little radiation as possible over the critical structures. The treatment goals change if a patient's illness is terminal or the standard treatments are not promising. For terminally ill patients, destroying the cancer is not the primary objective, but rather it is often the case that the treatments are designed to increase the patient's quality-of-life. In some instances, this means that some nearby regions should receive no radiation. For example, in the case of a brain tumor it may be best to minimize the radiation deposited into adjoining regions that control speech and memory. However, minimizing the amount of radiation that is received by these regions can subsequently imply that the tumor is not treated with a uniform, tumoricidal dose, but because the patient's illness is terminal, this is not considered a detriment to the treatment

plan. In cases where the standard treatments do not provide a high probability of success, the question becomes to what degree are the patient and physician willing to risk the nearby regions to treat the tumor with higher amounts of radiation.

The point of highlighting these situations is that the objective of treatment is not the same for all patients and is decided by the ethics and values of the patient and physician. This ethical perspective of the objective is different from the modeling perspective of the objective, which is concerned with how we measure and penalize deviation from the physician's demands. The fact that ethical concerns often make it difficult to clearly state a primary objective means that the optimization model needs to be flexible enough to accommodate several scenarios.

Before continuing with the mathematical models, we take a moment to discuss a couple of treatment concerns that are not readily discussed in the literature, but that are beginning to receive some attention. First, radiation therapy is not delivered in a single session, but rather it is fractionated into several treatments (usually 20 to 30). So once a plan is developed, it is divided into a number of treatments and the patient receives these fractionated treatments daily. The idea here is to accumulate the radiation in a slow enough manner so that the healthy tissue has an increased chance of survival. This fractionization is the difference between radiotherapy and radiosurgery, the latter of which is delivered all at once and is often used to prevent strokes. A natural, but virtually unexplored, question is whether or not it is beneficial to deliver the overall dose in non-uniform increments. The optimization model associated with such a question turns out to be a challenging optimal-control problem, with the only work being the recent paper of Ferris and Voelker [Ferris and Voellker, 2002]. While the computational burden of solving their model makes it impossible to develop a patient specific course of treatment, their work clearly indicates that delivering a uniform, fractionated dose is not typically optimal. There are many related questions that are open for investigation, such as deciding the number of fractionated treatments that maximize the success of the treatment.

The second treatment question that is relatively new is how to move the gantry and adjust the multileaf collimator so that the plan is delivered in as little time as possible, see [Boland et al., 2002]. This is an extremely important question because typical treatment times are about 15 minutes, and if treatment plans can be delivered more efficiently, more complicated plans can be used. In general, plans with more than 5 to 7 angles are considered complicated because of the time it takes to administer them. Hence, the number of angles in a treatment plan is restricted, and the flexibility of the design is limited. The

restriction on the number of angles becomes less of a concern as we find more efficient ways to move the gantry and adjust the multileaf collimator, which follows because we can incorporate more angles and more beam shapes in the alloted 15 minutes. So, there is more flexibility in designing a plan, and hence, increasing the efficiency of how treatment plans are delivered translates into a benefit for the patient

4.   Optimization Models

In this section we develop a class of linear programs that are intended to aid a dosimetrist design IMRT plans, see [Holder, 2001 and Holder, 2000]. The first optimization model that was developed to aid IMRT design was linear and appeared in the literature in 1968 [Bahr et al., 1968]. Since then, many researchers have experimented with linear models [Hodes, 1974; M. Langer, 1987; Langer et al., 1990; Legras et al., 1982; Lodwick et al., 1998; Morrill et al., 1990; Rosen et al., 1991; Sonderman and Abrahamson, 1985].

While linear models are natural because dose deposition is experimentally linear, these models have been the focus of several complaints, and many other researchers have investigated nonlinear models [Legras et al., 1982; Morrill et al., 1991; Rosen et al., 1991]. The first complaint about linear models is that the physician's demands often produce an empty feasible region. For example, the physician may desire that the cancerous tissue receives $80\text{Gy} \pm 2\%$ and that the surrounding critical structures receive no more than 20Gy. This translates into the following constraints,

$$78.4 \leq A_T x \leq 81.6, \ A_C x \leq 20, \ x \geq 0. \tag{1.1}$$

and this system may not be consistent. If the physician's demands are not possible, the optimization routine simply states that the underlying optimization problem is infeasible and provides no information about how to adjust the physician's desires. Finding and explaining a source of infeasibility is a difficult question, and there is a substantial amount of literature that deals with this issue, see for example [Chinneck, 1997; Chinneck, 1995; Greenberg, 1996; Greenberg, 1993]. Since we can not ask the physicians or the physicists to become experts in mathematical programming, this is a problem that needs to be addressed. The linear models that we develop overcome this difficulty by using elastic constraints.

The second major complaint about linear models has nothing to do with the linearity of the problem but rather the solution technique. The problem here is that the simplex algorithm terminates with an extreme point solution, which means that some of the inequality constraints are

guaranteed to hold with equality at the solution. For example, in (1.1) we are guaranteed that either some of the cancerous regions are going to receive their upper bound of 81.6Gy, or that some of the cancerous regions are going to receive their lower bound of 78.4Gy, or that some of the critical structures are going to receive their upper bound of 20Gy. The problem here is that we are **guaranteeing** that some regions are going to attain the limits are placed on them, and this is alarming because these limits are rules-of-thumb. We address this issue in two ways. First, we use a path-following interior point algorithm to solve our problems, and this algorithm terminates with a solution that strictly satisfies as many inequalities as possible. So, we find an optimal plan that does not attain the limits placed on the regions, provided that such a plan is possible. Second, the elastic constraints that we use allows the physician's desires to 'float' during the optimization process, and the objective is to better them as much as possible.

From the dimensions of A, A_T, A_C, and A_G, we have that m is the total number of pixels, m_T is the number of tumorous pixels, m_C is the number of critical structure pixels, and $m_G = m - m_T - m_C$ is the number of remaining pixels. 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 a fixed percentage of the physician's goal for the tumor. So, if 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 the percentage of variation permitted over the cancerous region 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 $GUB = (1.1)TG$.

The model that we use allows many ways for us to measure and penalize any deviation from the physician's goals. This generality is permitted through the use of *semimonotone* matrices, which are matrix whose Moore-Penrose generalized inverse is non-negative. (for more information see [Berman and Plemmons, 1979]. 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 $U_G e > 0$. Any collection of l, u_C, u_G, L, U_C, and U_G satisfying these assumptions defines a set of *elastic functions*. The feasible region, denoted by $\mathcal{F}$, is the collection of $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}^{q_T}$, $\beta \in \mathbb{R}^{q_C}$, and $\gamma \in \mathbb{R}^{q_G}$ that satisfy

$$\left. \begin{array}{rcccl} TLB - L\alpha & \leq & A_T x & \leq & TUB \\ & & A_C x & \leq & CUB + U_C \beta \\ & & A_G x & \leq & GUB + U_G \gamma \\ 0 & \leq & L\alpha & \leq & TLB \\ -CUB & \leq & U_C \beta & & \\ 0 & \leq & U_G \gamma & & \\ 0 & \leq & x. & & \end{array} \right\} \tag{1.2}$$

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 we measure the amount of elasticity, and with this in mind, we see that the assumption that $Le > 0$, $U_C e > 0$, and $U_G e > 0$ makes sure that each constraint is elastic. The elastic constraints are incorporated for two reasons. First, Lemma 1 shows that $\mathcal{F}$ is not empty for any collection of L, U_C, and U_G. Hence, the complaint that linear models are often infeasible does not apply to this model. Second, the different lower bounds on the elastic functions allow us to embody different treatment aspirations.

Each of L, U_C, and U_G correspond with a vector, denoted by l, u_C, and u_G, that decides how discrepancies are penalized. For example, $L\alpha$ measures how deficient a plan is with regards to meeting the minimum tumor dose, and $l^T \alpha$ is the total penalty assigned to these discrepancies. Similarly, $U_C \beta$ and $U_G \gamma$ measure a plans deviation from CUB and GUB, and $u_C^T x$ and $u_G^T x$ are the aggregated penalties assigned to these deviations. The separation of how we measure and penalize deviation is convenient because it allows us to consider one set of constraints, de-

cided by L, U_C, and U_G, and at the same time we can manipulate the objective function to address different situations. So we can design a patient-specific objective function that takes into account their ethical desires.

The objective functions that we consider are comprised of the three penalty functions $l^T\alpha$, $u_C\beta$, and $u_G\gamma$, and we consider variants of the following three optimization problems,

$$
\begin{aligned}
LP_\omega \quad &: \quad \min\{\omega \cdot l^T\alpha + u_C^T\beta + u_G^T\gamma : (x,\alpha,\beta,\gamma) \in \mathcal{F}\}, \\
MOLP \quad &: \quad \min\{(l^T\alpha, \ u_C^T\beta + u_G^T\gamma)^T : (x,\alpha,\beta,\gamma) \in \mathcal{F}\}, \text{ and} \\
MOLP'_{(T,C,G)} \quad &: \quad \min\{(l^T\alpha, \ u_C^T\beta, \ u_G^T\gamma)^T : (x,\alpha,\beta,\gamma) \in \mathcal{F}\}.
\end{aligned}
$$

The first math program, LP_ω, is a linear program where the three penalization functions are accumulated, with the weight ω deciding the importance of the tumor uniformity. If ω is small, we are indicating that finding a treatment plan that attains the lower bound on the tumor is not that important. As ω increases, we increase the emphases of finding a plan that achieves a uniform, tumoricidal dose. The second math program is a multiple objective linear program, where the two objectives are 1) to attain a uniform, tumoricidal dose and 2) to minimize the radiation received by all other structures. The third optimization problem is another multiple objective linear program, where the three objectives are to 1) minimize any deficiencies in the cancerous regions, 2) make sure that the critical structures receive as little radiation as possible, and 3) eliminate hot spots by minimizing the amount the remaining tissue is over CUB.

We point out that each of the mathematical programs is capable of addressing different ethical situations. In LP_ω, we adjust the relative importance of the tumor receiving its desired amount of radiation by adjusting the value of ω. In Section 5.1 we show that the minimum amount of tumor deficiency is uniformly bounded by the inverse of ω, and we use this result to construct an ω that guarantees that the tumor receives a uniform, tumoricidal dose.

Because the other two mathematical programs have multiple objectives, we need to define the sense of optimization that we are going to use. For *MOLP* we are interested in the set of *pareto* optimal, or *efficient*, solutions. We say that (x,α,β,γ) is pareto optimal if there exists a θ strictly between 0 and 1 such that (x,α,β,γ) is optimal to

$$
\min\{(1-\theta)l^T\alpha + \theta \cdot (u_C^T\beta + u_G^T\gamma) : (x,\alpha,\beta,\gamma) \in \mathcal{F}\}. \qquad (1.3)
$$

Since θ is positive, we have that dividing the objective function by θ transforms this problem into LP_ω, where $\omega = (1-\theta)/\theta$. So the set of

pareto optimal solutions to *MOLP* is the same as the collection of all optimal solutions to LP_ω, where ω is positive. While this means that LP_ω and *MOLP* are variants of each other, in the results that follow we use LP_ω to show that we can chose ω large enough to guarantee that there is no tumor deficiency and $MOLP^2$ to find a collection of beams that are not used in any pareto optimal solution. Because these two goals are different, we consider these as separate, but related, problems.

We use *lexicographic* optimization, instead of pareto optimization, for the third optimization problem. For example, for $MOLP'_{(T,C,G)}$ we minimize $l^T\alpha$ first, and then minimize $u_C^T\beta$ over $\mathrm{argmin}\{l^T\alpha : (x,\alpha,\beta,\gamma) \in \mathcal{F}\}$. Similarly, the third objective of minimizing $u_G^T\gamma$ is undertaken with only the solutions of the second problem. Because the three objectives are treated individually, we can easily alter their importance. In the case of a terminally ill patient, we may decide that attaining a uniform, tumoricidal dose is the least important objective, and hence we might order the objectives by 1) guaranteeing that the critical structures are underneath their bounds, 2) making sure that there are no unusually high depositions of radiation, and 3) attempt to deliver a uniform, tumoricidal dose. In such a case, we use the subscript (C,G,T) to indicate that the critical structures have the highest importance, that the normal (good) tissue has the second highest priority, and that the tumor has the lowest priority. So, the first objective is to minimize $u_C^T\beta$, the second objective is to minimize $u_G^T\gamma$, and the third objective is to minimize $l^T\alpha$. To make sure that the consequences of lexicographic optimization are understood, suppose for $MOLP_{(T,C,G)}$ that there are treatment plans that achieve a uniform, tumoricidal dose. This means that the set of optimal solutions found by minimizing $l^T\alpha$ are exactly those plans that achieve a uniform, tumoricidal dose. When the second objective is minimized, we are only going to consider those treatment plans that achieve a uniform, tumoricidal dose, and it is possible that none of these plans adhere to the bounds placed on the critical structures. However, there may be a plan that delivers sufficiently low levels to the critical structures and only has the slightest tumor deficiency, but we would not find such a plan because it would not be optimal to the first problem. This is the nature of lexicographic optimization, and this type of optimization is appropriate only if a hierarchy of the objectives is clear.

Different elastic functions lead to different interpretations of the solution, and the following two collections are of particular interest.

Average Analysis

$$l = \frac{1}{m_T}e \quad u_C = \frac{1}{m_C}e \quad u_G = \frac{1}{m_G}e \quad L = I \quad U_C = I \quad U_G = I$$

Absolute Analysis

$$l = 1 \quad u_C = 1 \quad u_G = 1 \quad L = e \quad U_C = e \quad U_G = e$$

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 the physician 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 that the tumor is under its prescribed dose,

- minimize the average amount of radiation that the critical structures receive, and

- minimize the average amount that the remaining pixels are over their upper bounds.

If absolute analysis is chosen, the interpretation is similar to that of average analysis, with the difference being 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 that the tumor is under its prescribed dose,

Citations	Models Investigated
Rosen et al., 1991; Shepard et al., 1999	reviews linear and nonlinear models
M. Langer, 1987; Langer et al., 1990	mixed integer models
Morrill et al., 1991; Raphael, 1992	probabilistic models

Table 1.1. Research papers that investigate and review optimization models that are used to aid IMRT design.

- minimize the maximum amount of radiation that the critical structures receive, and

- minimize the maximum amount any remaining pixel is over its upper bound.

There are several models in the literature, ranging from linear to mixed integer to nonlinear models. While our concentration is on the linear models developed above, it is important for operational researchers that want to work in the field to have an awareness of these models. We direct researchers to the works listed in Table 1.1.

5.   Mathematical and Computational Results

As mentioned in the introduction, the medical literature associated with IMRT design and optimization is immense. The history of the medical research is to design a specific treatment plan for a specific type of cancer, and then show that it is appropriate through several examples. So in the medical literature, the methodology of treatment is verified through examples —i.e. this technique works because it has favorable properties on these examples. While this verification approach is important, and indeed the great strides we have made in medicine are a direct result of such work, this type of research is foreign to a mathematician. For the field of mathematics is concerned with statements that can be universally proved and not ones that can simply be shown to hold for a few examples. This means that an applied mathematician's perspective of a problem is different from the perspective held by a practitioner. An applied mathematician, such as an operational researcher, approaches a problem by finding the mathematical language needed to describe the essence of the problem and then proceeds to prove statements about the situation at hand. The proofs provide a theoretical certificate for what can and can not be stated about the problem. So instead of stating that a technique works because we can show that it does on a few examples, we can guarantee that a technique does or does not work because the proofs establish that they will, or will not. The benefit of this theoreti-

cal approach is that it is not example dependent, so we do not have to wonder if there are examples where the technique fails.

In this section we are interested in developing theoretical statements about the linear models presented in Section 4. Unfortunately, mathematical rigor is scarce in the literature. Consequently, this field of research has not benefited from a sound mathematical development, and this author hopes that the operational researchers working in this field feel a sense of responsibility to provide the needed theoretical basis (areas in telecommunications and geosciences have benefited greatly from a similar mathematical foundation). The results of this section are divided into three subsections, each related to one of the three models presented in Section 4. The proofs of the mathematical results are excluded for brevity, but a citation is provided where a proof can be found.

The numerical results in each of the following sections rely on the *optimal partition* of a linear program. Consider the standard form linear program, $\min\{c^T x : Ax = b, x \geq 0\}$. Allowing $\mathcal{P}^*$ to be the optimality set, we have that the optimal partition $(B|N)$ is defined by

$$N = \{i : x_i = 0 \text{ for all } x \in \mathcal{P}^*\}, \quad \text{and} \quad B = \{1, 2, \ldots, n\}\backslash N.$$

The reason that the optimal partition is important is that it provides an algebraic characterization of the optimal set. We do not have the space in this article to rigorously develop this representation, but it is well known that $\mathcal{P}^* = \{x : Ax = b, x \geq 0, x_i = 0, i \in N\}$, see [Roos et al., 1997].

The optimal partition was not easily computed until path-following interior point algorithms became viable alternatives to the simplex method. Path-following interior point algorithms terminate with a solution that induces the optimal partition, which means that if x^* is an optimal solution that is found by such an algorithm, then $B = \{i : x_i^* > 0\}$ and $N = \{i : x_i^* = 0\}$. Having this type of solution is important for us because it strictly satisfies as many inequalities as possible —i.e. the B set indexes the **entire** collection of inequalities that can be strictly satisfied by an optimal solution. So for IMRT design, we have that if we use a path-following interior point algorithm, the plan that we find strictly satisfies as much of the prescription as possible.

A path-following interior point algorithm is appropriate only if the strict interior of the feasibility set is non-empty. For us this means that there must be an $(x, \alpha, \beta, \gamma)$ that **strictly** satisfies the inequalities in (1.2). Fortunatly, Lemma 1 states that we are guaranteed that a path-following interior point algorithm is applicable.

Lemma 1 (Holder, 2001) *We have for any collection of elastic functions that the strict interior of $\mathcal{F}$ is non-empty.*

5.1   Analysis Certificates for LP_ω

The objective function of LP_ω 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 can calculate an ω that certifies that the discrepancy between the amount delivered to the tumor and the tumor's lower bound is sufficiently small.

We say that a prescription *allows tumor uniformity* if there is a treatment plan x such that $TLB \leq A_\tau x \leq TUB$. Moreover, a prescription is *attainable* if there is an $(x, \alpha, \beta, \gamma)$ in $\mathcal{F}$ such that $L\alpha = 0$, $U_C\beta \leq 0$, and $U_G\gamma = 0$. Obviously every attainable prescription allows tumor uniformity, but it is not the case that every prescription that allows tumor uniformity is attainable. Theorem 1 shows that if the prescription allows tumor uniformity, then the tumor deficiency is uniformly bounded above by the inverse of ω. In what follows, we let $\underline{rs}(M)$ be the minimum row sum of the matrix M, and we use the standard big-O order 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 $f(x) \leq \kappa g(x)$. Also, we use the standard notations for the 1-norm, $\|x\|_1 = \sum_i^n |x_i|$, and the infinity-norm, $\|x\|_\infty = \max\{|x_i| : i = 1, 2, \ldots, n\}$.

Theorem 1 (Holder, 2000) *Let $(x^*(\omega), \alpha^*(\omega), \beta^*(\omega), \gamma^*(\omega))$ be an optimal solution to LP_ω. For any collection of elastic functions we have that $l^T\alpha^*(\omega) = O\left(\frac{1}{\omega}\right)$, provided that the prescription allows tumor uniformity.*

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 if either average or absolute analysis is used. In particular, setting

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

we have from [Holder, 2000] that κ is no greater than

$$\frac{\|A_C(ue) - CUB\|_\infty \|u_C^T\|_1}{\underline{\mathrm{rs}}(U_C)} + \frac{\|A_G(ue) - GUB\|_\infty \|u_G^T\|_1}{\underline{\mathrm{rs}}(U_G)} + u_C^T CUB,$$

if average use is used, and

$$\frac{\|A_C(ue) - CUB\|_\infty \|u_C^T\|_1}{\underline{\mathrm{rs}}(U_C)} + \frac{\|A_G(ue) - GUB\|_\infty \|u_G^T\|_1}{\underline{\mathrm{rs}}(U_G)} + \frac{u_C^T U_C^T CUB}{m_C},$$

if absolute analysis is used. We let κ' be the greater of these two bounds so that regardless of the type of analysis, we have that $l^T \alpha^*(\omega) \leq \kappa'/\omega$.

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 each component of TLB by $\varepsilon > 0$ —i.e we instead let $TLB = (1 - tol)TGe + \varepsilon e$. After calculating κ', we choose $\omega = \kappa'/\varepsilon$ and solve LP_ω. 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 1.5. Of course a more detailed interpretation of the solution is possible by examining the individual components of $(\alpha^*(\omega), \beta^*(\omega), \gamma^*(\omega))$.

A prototype treatment system called $\mathcal{R}$adiotherapy optim$\mathcal{A}$l $\mathcal{D}$esign, or $\mathcal{RAD}$, has been developed using MATLAB$^\copyright$. This system is available from `http://www.trinity.edu/aholder/research/oncology/`, and requires MATLAB's optimization toolbox. $\mathcal{RAD}$ 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. In addition to allowing the user to choose from different angle geometries, $\mathcal{RAD}$ has the following features.

- Either absolute or average analysis can be used.

- 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 equally spaced beams, and each beam contained 32 sub-beams. The amount by which TLB is

---

Interpreting The Solution: Average Analysis

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

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

[**Case 2a:** $e^T \beta^*(\omega) + e^T \gamma^*(\omega) > 0$] 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:** $e^T \beta^*(\omega) + e^T \gamma^*(\omega) \leq 0$] 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

[**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:** $\beta^*(\omega) + \gamma^*(\omega) > 0$] 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:** $\beta^*(\omega) + \gamma^*(\omega) \leq 0$] 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 1.5.   Interpreting the solution for either average or absolute analysis
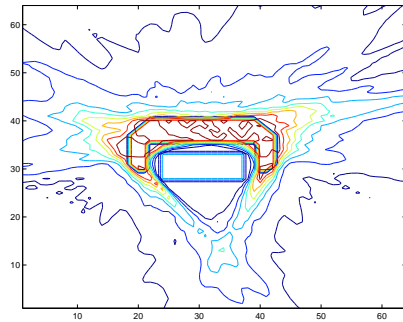
Figure 1.6. A contour plot showing how the deposition pattern 'bends' around the critical structure.
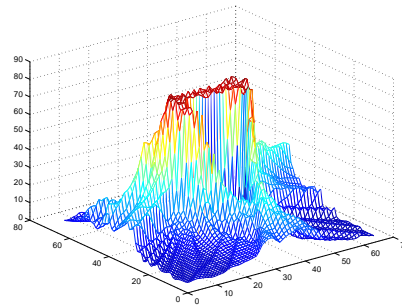

Figure 1.7. The vertical height is the amount of radiation delivered by the plan over the image.

increased is internally set at 10^{-4}, so $TLB = (1 - tol)TG + 10^{-4}e$. The problems were solved on a 1.5 GHz PC with 1G of RAM.

The first example is found in Figures 1.6 and 1.7. In this example a tumor has grown half-way around a critical structure. The tumoricidal dose was 80Gy, and the critical structure was restricted to no more than 30Gy. The tumor uniformity level was 2%, and an absolute analysis was used. The value of $l^T\alpha^*$ was less than 10^{-4}, from which we conclude that the prescription allows tumor uniformity. Indeed, the maximum and minimum doses were 78.42Gy and 81.57Gy, which are within the 80Gy $\pm$ 2%. Not only does this plan strictly satisfy the tumor uniformity bounds, but it also does not deliver any radiation to the critical structure. So we have designed a plan that delivers a uniform, tumoricidal dose to the tumor and does not deposit any radiation in the critical structure.

The example in Figure 1.8 is significantly more complicated because the tumor is nearly surrounded by critical structures. The tumoricidal dose was 78Gy, with a uniformity level of 4%. The plan depicted in Figures 1.9 and 1.10 attained the tumor uniformity with the minimum and maximum doses inside the cancerous region being 75.17Gy and 81.1Gy. However, there is no plan that achieves a uniform, tumoricidal dose that does not violate the bounds placed on the critical structure. We know this because the value of $l^T\alpha^*$, which is simply α^* for the absolute analysis used, is less than 10^{-4} and the value of $u_C^T\beta^*$, which is simply β^*, is 6.62. So we know that some critical structure tissue must receive 6.62Gy over its prescribed bound to attain a uniform, tumoricidal dose. Depending on the type of critical structure, this may or may not be acceptable, and if not, the planners need to reconsider their desires for the tumor.
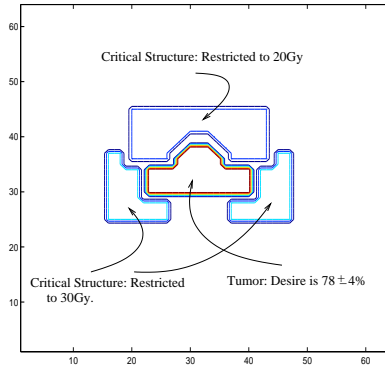
Figure 1.8. A difficult geometry to plan because the tumor is almost surrounded by low-dose critical structures.
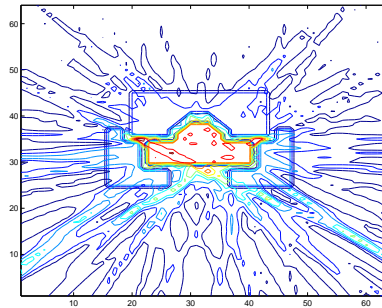

Figure 1.9. A contour plot of an optimal plan. Notice that each angle is used in an attempt to accumulate the required tumoricidal dose and satisfy the critical structure bounds.
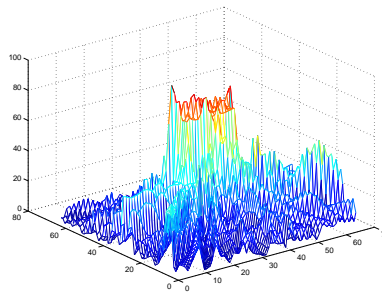

Figure 1.10. The amount of radiation over the image. The amount over the tumor is fairly uniform, but there is a spike between the tumor and one of the critical structures.

5.2   Evaluating an Angles Value with *MOLP*

We see from the examples in Section 5.1 that the plans developed by a path-following interior point algorithm tend to design plans that use many angles. In fact, these plans use so many angels that they are not practical —i.e. the time that it would take to deliver such a plan is well beyond the 15 minutes of a typical treatment. The problem here is that a path-following interior point algorithm terminates with a solution that strictly satisfies as many inequalities as possible, and as already mentioned, this is favorable because we design a plan that strictly satisfies as much of the prescription as possible. However, this

is bad news because we also design a plan that uses as many angles and sub-beams as possible. What is needed is a technique to prune the collection of possible angles so that the 'best' angles remain in the pruned collection. This and the following section develop ways to measure how important an angle is when the priorities of the treatment are uncertain.

The set of pareto optimal solutions of *MOLP*, called the *efficient frontier*, induces an optimal partition for the multiple objective program that is similar to the optimal partition, see [Holder, 2001].

Definition 1 *Let $\mathcal{E}$ be the efficient frontier of MOLP. The MOLP optimal partition, denoted $(\overset{molp}{B} \mid \overset{molp}{N})$, is defined by*

$$\overset{molp}{N} = \{i : x_i = 0 \text{ for all } x \in \mathcal{E}\} \text{ and}$$
$$\overset{molp}{B} = \{1, 2, 3, \ldots, n\}\backslash N.$$

The definition of the MOLP optimal partition retains the quality that an index being in N indicates that the component is zero in every pareto optimal solution. Likewise, an index in B demonstrates that the component is allowed to be positive on the efficient frontier. A property that is unfortunately lost is that the MOLP optimal partition is not capable of characterizing the efficient frontier —i.e. $\mathcal{E} \neq \{(x, \alpha, \beta, \gamma) \in \mathcal{F} : (x, \alpha, \beta, \gamma)_i = 0, i \in \overset{molp}{N}\}$. However, we do have that $\mathcal{E} \subseteq \{(x, \alpha, \beta, \gamma) \in \mathcal{F} : (x, \alpha, \beta, \gamma)_i = 0, i \in \overset{molp}{N}\}$.

An algorithm to compute the MOLP optimal partition is found in [Holder, 2001]. This algorithm uses the parameterization in 1.3 by finding the linear programming optimal partition for every value of θ between 0 and 1. The set $\overset{molp}{B}$ is the union of all the linear programming B sets, and $\overset{molp}{N}$ is $\{1, 2, \ldots, n\}\backslash \overset{molp}{B}$. So $\overset{molp}{N}$ indexes the sub-beams that are not used for any $\theta \in (0, 1)$. Recall that the value of θ is a measure of how important it is to deliver a uniform, tumoricidal dose, with small values of θ giving the cancerous regions a high importance and values near 1 giving little importance to the cancerous regions. This means that $\overset{molp}{N}$ contains the sub-beams that are not used in any weighting of the objectives, and hence, these sub-beams should never be used in a treatment plan. The other side of this is that $\overset{molp}{B}$ indexes the sub-beams that are used for some collection of weights, and hence, there is at least one circumstance where each of these sub-beams is used.

Because of the way the algorithm works, we actually acquire more information than just described. Consider the situation depicted in Figure 1.11, where a tumor is surrounded by three critical structures. This

experiment was run with 72 equally spaced angles, each with 32 sub-beams. The tumor uniformity level was $\pm 4\%$ and an average analysis was used. Figures 1.12 and 1.13 provide information about which angles are used, and not used, as θ traverses the interval $(0, 1)$. We saved the optimal partition for each θ and used this information to calculate how often an angle is used. We say that an angle is *used at level k* if there are k sub-beams from that angle with positive amounts of radiation in the optimal plan. Furthermore, we say that an angle is *on* provided its level of use is at least 1. In Figure 1.12 we calculated each angle's level of use, and then added these together for each of the 399 different optimal partitions. These values are recorded above the circle around the image. The highest peak is at 90° and has a value of $1,424$, which means that $1,424$ sub-beams were used from this angle as θ traversed the interval $(0, 1)$. Figure 1.13 is similar, but instead of accumulating sub-beams from each angle, the percentage of times an angle is on is displayed over the circle. Angle 85° was on in 100% of the optimal partitions, and while 90° had the highest amount of sub-beam usage, it was not on in each optimal partition (it was used in 99.75% of the optimal partitions).

The point of Figures 1.12 and 1.13 is that a dosimetrist can easily decide which angles are, and are not, important. The most definitive information lies in the angles that are never used, which follows because there is no situation where these angles are used. Similarly, any angle whose use is 100%, which is only 85° for this example, is used in every situation. The graphs provide a measure of an angles usefulness in other situations. For example, if a dosimetrist wants a 3 beam plan, then he or she might decide to use angles 85°, 40°, and 205°, all of which have higher peaks in Figure 1.13.

5.3 Using *MOLP'* to Reduce the Number of Angles

Recall that we use lexicographic optimization for *MOLP'* and that the order in which the objectives are considered is indicated by the subscript. Lexicographic optimization has its own optimal partition, which is easily calculated for our problem, see [Holder, 2001]. As an example, consider $MOLP'_{(C,T,G)}$, where we calculate the lexicographic optimization as follows.

Step 1 Solve $\min\{u_C^T \beta : (x, \alpha, \beta, \gamma) \in \mathcal{F}\}$ and let $(B^1 | N^1)$ be the optimal partition.

Step 2 Solve $\min\{l^T \alpha : (x, \alpha, \beta, \gamma) \in \mathcal{F}, (x, \alpha, \beta, \gamma)_i = 0, i \in N^1\}$ and let $(B^2 | N^2)$ be the optimal partition.
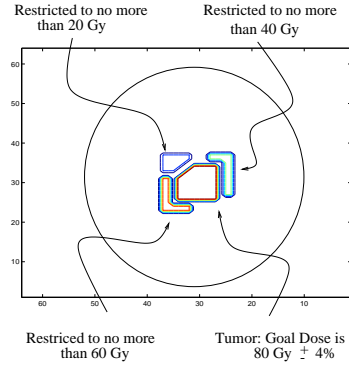
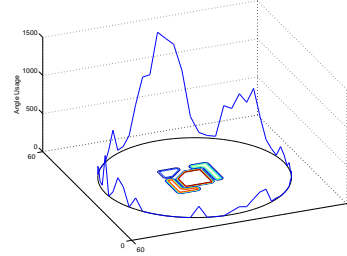*Figure 1.11.* A tumor surrounded by 3 critical structures.

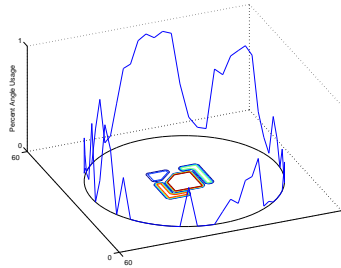*Figure 1.12.* Accumulative totals of sub-beam usage along each angle.

Figure 1.13. Percent totals for each angle.

Step 3 Solve $\min\{u_G^T\gamma : (x,\alpha,\beta,\gamma) \in \mathcal{F}, (x,\alpha,\beta,\gamma)_i = 0, i \in N^2\}$ and let $(B_L|N_L)$ be the optimal partition.

The last partition, $(B_L|N_L)$, is called the *lexicographic optimal partition*. If we rearranged the priorities, to say (G, C, T), the only difference would be that the objective function in step 1 would be $u_C^T\beta$, in step 2 it would be $u_C^T\gamma$, and in step 3 it would be $l^T\alpha$.

Similar to the *MOLP* optimal partition, the lexicographic optimal partition provides an insight into the usefulness of an angle. As an example, consider the problem in Figure 1.14, where a tumor has gown around a critical structure. There were 72 equally spaced angles, each containing 32 sub-beams. The tumor uniformity level was ±4%, and an average analysis was used. We calculated the lexicographic optimal partition for each of the 6 possible orderings of the objectives, and Figures 1.15 through 1.20 show each angles level of use in corresponding optimal plan.
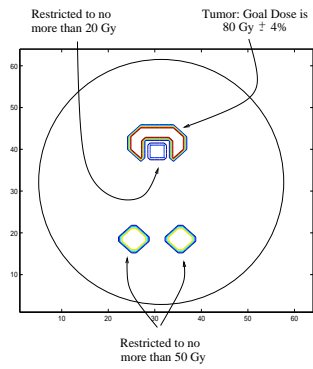
Restricted to no
more than 20 Gy

Tumor: Goal Dose is
80 Gy ± 4%

Restricted to no
more than 50 Gy

Figure 1.14. A tumor that has grown around a critical structure.
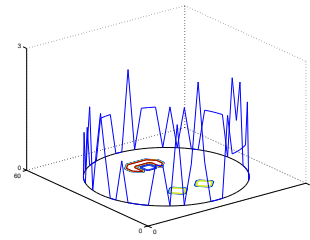

Figure 1.15. Each angles use for priority list (T, C, G).
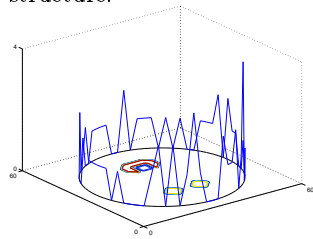

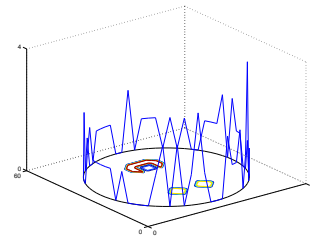*Figure 1.16.* Each angles use for priority list (T, G, C).

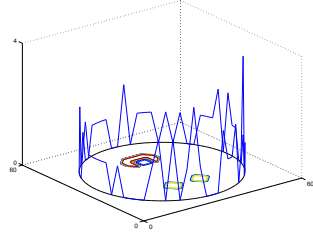*Figure 1.17.* Each angles use for priority list (G, T, C).

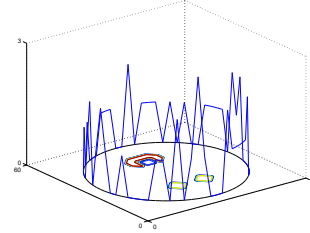*Figure 1.18.* Each angles use for priority list (G, T, C).

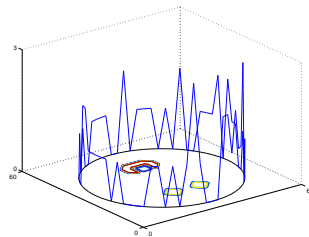*Figure 1.19.* Each angles use for priority list (C, T, G).

Figure 1.20. Each angles use for priority list (C, G, T).

Again, the most definitive information comes from the angles that have a zero level of use, for these angles are not to be considered for that priority list. As an example, angle 170° has a zero level of use for the priority list (T, C, G), but has a level of use of 2 for the priority list (T, G, C). There were 9 angles whose level of use was zero in all 6 priority list: 55°, 75°, 120°, 125°, 205°, 225°, 250°, 310°, and 355°. These are the angles that can be excluded from consideration regardless of how the dosimetrist orders the priorities (which means we have removed 12.5% of the angles from consideration).

6. Conclusion

We conclude this article with a plea to the operations research community. Operations Research has successfully been used in many disciplines, but one of the few areas that has not witnessed the benefits of the field is clinical medicine. Of course, the statistical training that most operational researchers posses is useful in drug trials and other data intensive medical applications. However, many of the medical procedures that are used in practice have not been mathematically modeled and optimized, which means that any improvement in the treatment is found by trial and error. The area of IMRT design is starting to benefit from the optimization process, and we are now at a point where the operational researchers can make a substantial improvement in a patient's treatment. The author urges those who work in the field of operations researcher to consider working on a problem that involves some clinical treatment, for all of humankind benefits from this work.

References

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

[Bartolozzi et al., 2000] Bartolozzi, F. et al. (2000). Operational research techniques in medical treatment and diagnosis. a review. *European Journal of Operations Research*, 121(3):435–466.

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

[Boland et al., 2002] Boland, N., Hamacher, H., and Lenzen, F. (2002). Minimizing beam-on time in cancer radiation treatment using multileaf collimators. Technical Report KLUEDO:2002-02-10, Universittsbibliothek Kaiserslautern.

[Censor, 1991] Censor, Y. (1991). Mathematical aspects of radiation therapy treatment planning: Continuous inversion versus full discretization and optimization versus feasibility. In C. Börgers and F. Natterer, editor, *Computational Radiology and Imaging: Therapy and Diagnostics*, volume 110 of *The IMA Volumes in Mathematics and its Applications*, pages 101–112. Springer-Verlag, New York, NY.

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

[Chinneck, 1995] Chinneck, J. (1995). An effective polynomial-time heuristic for the minimum-cardinality iis set-covering problem. *Annals of Mathematics and Artificial Intelligence*, 17:127–144.

[Chinneck, 1997] Chinneck, J. (1997). Finding a useful subset of constraints for analysis in an infeasible linear program. *INFORMS Journal on Computing*, 9(2):??

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

[Ferris and Voellker, 2002] Ferris, M. and Voellker, M. (2002). Neurodynamic programming for radiation treatment planning. Technical Report NA-02/06, Numerical Analysis Group, Computing Laboratory, Oxford University.

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

[Greenberg, 1993] Greenberg, H. (1993). How to analyze results of linear programs, part 3: Infeasibility diagnosis. *Interfaces*, 23(6):120–139.

[Greenberg, 1996] Greenberg, H. (1996). Consistency, redundancy and implied equalities in linear systems. *Annals of Mathematics and Artificial Intelligence*, 17:37–83.

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

[Holder, 2000] Holder, A. (2000). Desinging radiotherapy plans with elastic constraints and interior point methods. Technical Report No. 49, Trinity University Mathematics, San Antonio, TX. To appear in Health Care and Management Science.

[Holder, 2001] Holder, A. (2001). Partitioning multiple objective solutions with applications in radiotherapy design. Technical Report 54, Trinity University Mathematics.

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

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

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

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

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

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

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

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

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

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

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

30

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

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

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