Minimum–Support Solutions for Radiotherapy Planning

Stephen C. Billups [a] Janine M. Kennedy [b]

[a] *Department of Mathematics, University of Colorado, Denver, Colorado 80217-3364*
E-mail: sbillups@carbon.cudenver.edu
[b] *Department of Mathematics, University of Colorado, Denver, Colorado 80217-3364*
E-mail: jmkenned@math.cudenver.edu

Computer-generated plans for radiation treatment sometimes involve an unnecessarily large number of gantry angles. Such plans are time consuming and expensive to administer because each distinct angle causes a delay while the gantry is repositioned. To address this issue, we consider an optimization model that generates minimum-support solutions to the treatment planning problem–that is, solutions involving a minimum number of gantry angles. This model is a polyhedral concave program, which is solved using a successive linearization algorithm based on work by Mangasarian [9].

Keywords: Radiotherapy planning, minimum–support solutions, polyhedral concave program, linear programming

1. Introduction

External beam radiotherapy is used to treat various types of cancer. The equipment used consists of a linear accelerator or betatron, which is mounted on a gantry that is able to rotate around the patient. Typical treatment sessions consist of administering radiation from a variety of different gantry angles in order to minimize the damage done to healthy tissue while delivering a prescribed dose of radiation to the tumor. For each gantry angle, the radiation beam must be *shaped* in some way so that radiation is delivered only where it is desired. This shaping consists of blocking out parts of the radiation beam. Until recently, this was accomplished by inserting lead wedges in front of the betatron. This resulted in a significant setup time for each distinct gantry angle. A more recent technology is called *multileaf collimation*. Here, an array of small metal leaves

is used to shape the beam. For each gantry angle, numerous small doses of radiation are administered, each with a different subset of the collimator leaves obstructing part of the beam. This enables very precise shaping of the radiation delivered from each gantry angle. Because this process is computer automated, it is considerably faster than the use of lead wedges. However, a nontrivial amount of time is still required for each gantry angle.

Before a patient undergoes external beam radiotherapy, a radiotherapy plan (RTP) must be created. The RTP specifies the gantry angles to be used as well as the shape and intensity of the radiation beam delivered from each gantry angle. The goals of the RTP are to deliver a prescribed dose of radiation to the tumor while minimizing the damage done to surrounding organs and other healthy tissue. Numerous optimization algorithms have been proposed for automating the generation of such treatment plans [1,3–8,11,12,14–16]. Unfortunately, the resulting computer-generated plans may involve an unnecessarily large number of gantry angles. Because each distinct gantry angle represents a time cost in administering the plan, it is desirable to generate plans that involve as few gantry angles as possible.

This paper is concerned with finding solutions to the Radiotherapy Planning Problem (RTPP) that involve the fewest number of gantry angles possible. Such solutions are called *minimum–support solutions*. To address this issue we will modify a linear programming model from [8] by introducing a term into the objective function that penalizes each gantry angle used in the treatment plan by some fixed cost β. The resulting mathematical program has a discontinuous concave objective function and a polyhedral feasible region. Solutions to the resulting formulation will be referred to as *β–minimum–support solutions*. Because this problem is NP-hard, we do not attempt to find an exact solution–instead we will employ a successive linearization algorithm based on [9] to find a stationary point of this problem.

We begin in Section 2 by describing the radiotherapy planning problem and presenting a detailed explanation of a specific linear programming formulation. Section 3 then discusses our adaptations to the algorithm described in [9] for finding β–minimum support solutions. Implementation details are given in Section 4 and an illustrative example problem is solved in Section 5. Section 6 concludes.

1.1. Notation

Let $\Re^n_+$ denote the positive n-dimensional real space. All vectors will be column vectors unless transposed to a row vector by a prime superscript $'$. For a vector x in $\Re^n_+$, $|x|_*$ is defined by

$$(|x|_*)_i = \begin{cases} 0 \text{ if } x_i = 0 \\ 1 \text{ otherwise.} \end{cases}$$

For a linear program $\min_{x \in X} c'x$ the notation {arg vertex $\min_{x \in X} c'x$} will denote the set of vertex solutions of the linear program. The vector e will denote a column vector of ones, and ε will denote the base of the natural logarithm. For a scalar c and a vector $z \in \Re^n$, we define the vector $c^z := [c^{z_1}, c^{z_2}, \ldots, c^{z_n}]'$.

2. Background

2.1. The Radiotherapy Planning Problem

The first step in radiotherapy planning is to collect computed tomography (CT) images of the patient's body surrounding the tumor. Each CT image represents a two-dimensional slice of the patient's body. Several CT images, taken along parallel slices of the body can be stacked together to form a three-dimensional picture of the body surrounding the tumor. For simplicity, for the remainder of this paper, we will deal only with the two-dimensional planning problem where we have only a single CT image. However, the model can be extended to handle the three-dimensional problem.

For planning purposes, the CT image is discretized into a two-dimensional grid of pixels. This is illustrated in Figure 1 with a 4×4 grid. In this figure, the gray oval represents the patient's body. In practical applications, grid resolutions of 64×64 to 512×512 are typically used. Using image processing software, a technician examines the CT image and identifies organ and tumor boundaries. This results in classifying each pixel of the image into one of the following subsets: **critical**, the set of healthy tissue pixels for sensitive structures, such as the spinal column or kidneys; **body**, the set of less sensitive healthy pixels such as bones or muscles; **tumor**, the set of cancerous pixels.

Recall that radiation can be delivered from a variety of gantry angles. Indeed, the gantry typically can be positioned at any angle in a complete circle around the body in 1 or 2 degree increments. The position of the gantry is repre-

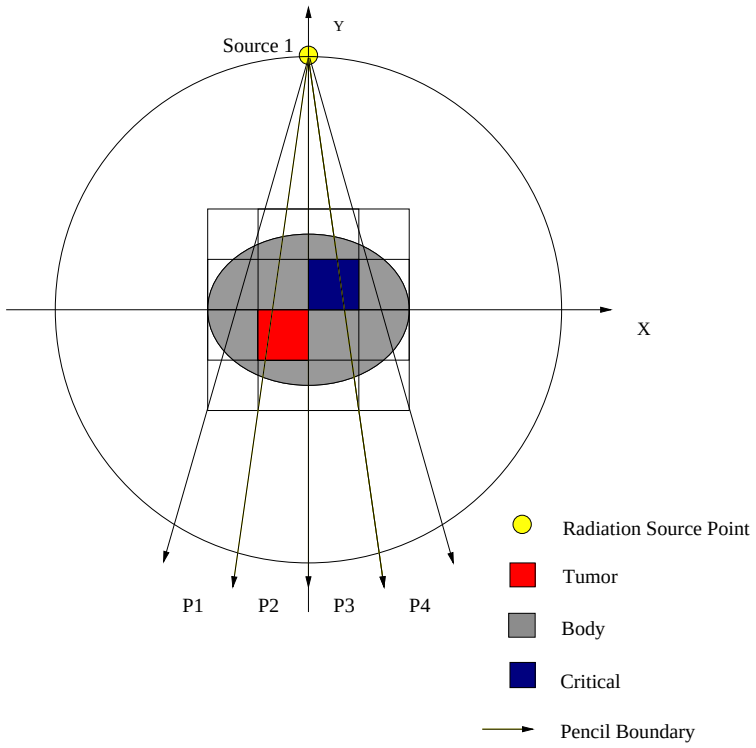

Figure 1. Radiation Beams with Pencils

sented by the angle the gantry makes with the x-axis. In Figure 1, the gantry is at 90 degrees. The radiation beam is always aimed directly at the center of the image, (called the *isocenter*), but the beam spreads out as it moves away from the source point, much like light emitted from a flashlight. The beam passes through the image and deposits radiation to all the pixels in its path. In order to deliver radiation exactly where it is desired, the beam must be shaped using lead wedges or a multi-leaf collimator. To model the beam shape, we divide the beam up into a collection of evenly spaced strips called *pencils*. In Figure 1, the beam is divided up into four pencils. In practice, there will typically be 40 or more pencils for each gantry angle.

Using the above definitions, the treatment plan is specified by a collection of gantry angles, henceforth referred to as *beams*, together with the amount of radiation to be delivered along each pencil within the beam. This information is stored in a two-dimensional array x, whose entries specify the radiation levels to be administered along each pencil within each beam. For example, if we number the gantry angles in 2 degree increments, then $x(p, b)$ represents the amount of

radiation to be delivered along the pth pencil from the bth beam, which makes an angle of $2b$ degrees from the x-axis. If $x(p,b) = 0$ then the pth pencil of the bth beam is off. If an entire column $x(\cdot, b)$ is zero, then the bth beam is not used. Generally, the total treatment time depends on the number of beams but not on the number of pencils used within each beam. Thus, a minimum–support solution to the radiotherapy planning problem is one that uses a minimum number of beams, whereas the actual number of pencils used is irrelevant.

The last component needed for modeling the treatment planning problem is the dose deposition operator. This is a three dimensional matrix D whose entries specify the fractional amount of radiation delivered to each pixel of the image from each pencil. That is, $D(i,p,b)$ represents the fraction of $x(p,b)$ that will be delivered to pixel i. Thus if $x(p,b) = d$ and $D(i,p,b) = \delta$, then the radiation dose delivered to pixel i by the pth pencil for the bth gantry angle will be $D(i,p,b)x(p,b) = \delta d$.

2.2. Linear Programming Formulation

There are many linear programming formulations for automating treatment planning. We shall work with the following formulation from [8].

minimize	maximum dose to critical structures
subject to	required tumor dose $\leq$ tumor dose $\leq$ maximum tumor dose
	normal tissue dose $\leq$ dose bound for normal tissue
	dose ≥ 0.

This mathematical programming model produces uniform treatment plans that satisfy all radiation bounds. The objective function minimizes the maximum dose deposited to any pixel in any critical structure. This is superior to minimizing the total radiation dose for all critical structures because it eliminates **hot spots**, which are areas that receive very high doses of radiation. We avoid hot spots, even in the tumor, for two reasons: first, hot spots can be necrotic and therefore difficult for the patient to excrete; second, if the tumor is very small and the patient is not set up perfectly, the hot spot could be shifted to normal tissue, which can be a very serious problem.

The above formulation can easily be recast as a linear program using standard techniques [2]. But first, we examine each of the constraints in more detail. To begin, consider the tumor requirements. Clearly, the tumor must receive a therapeutic dose of radiation. This dose can be specified by setting a lower bound

T_l and an upper bound T_u on the amount of radiation delivered to each tumor pixel. This yields the following constraint:

$$T_l \leq \sum_{p \in P} \sum_{b \in B} D(t,p,b)x(p,b) \leq T_u, \quad \text{for all } t \in \text{tumor}, \tag{1}$$

where the sets P and B index the pencils and gantry angles, respectively. Equation (1) states that the total sum of radiation distributed to a tumor pixel by all the pencils in the plan must be more than the minimum bound T_l and less than the maximum bound T_u prescribed by the oncologist. We use the same method to put an upper bound on the body pixels:

$$\sum_{p \in P} \sum_{b \in B} D(o,p,b)x(p,b) \leq O_u, \quad \text{for all } o \in \text{body}. \tag{2}$$

Finally, we want to minimize the maximum dose delivered to any pixel in a critical structure. This can be accomplished by minimizing a dummy variable γ subject to the constraint

$$\gamma - \sum_{p \in P} \sum_{b \in B} D(c,p,b)x(p,b) \geq 0, \quad \text{for all } c \in \text{critical}.$$

Observe that since γ is minimized, it will, at the solution, be exactly equal to the maximum dose to critical structures. The complete LP formulation of the RTPP is as follows:

$$\begin{array}{rl} \min_{\gamma,x} & \gamma \\ \text{subject to} & \gamma - \sum_{p \in P} \sum_{b \in B} D(c,p,b)x(p,b) \geq 0, \quad c \in \text{critical} \\ & T_l \leq \sum_{p \in P} \sum_{b \in B} D(t,p,b)x(p,b) \leq T_u, \; t \in \text{tumor} \\ & \sum_{p \in P} \sum_{b \in B} D(o,p,b)x(p,b) \leq O_u, o \in \text{body}. \end{array} \tag{3}$$

Defining S to be the feasible region of (3), the linear program can be written succinctly in the form

$$\min_{(\gamma,x) \in S} \gamma.$$

In the next section, we discuss modifications to this linear program aimed at producing minimum support solutions.

3. Finding Minimum–Support Solutions

3.1. Penalty Problem

In order to reduce the number of beams used in the RTP, we modify the objective function in (3) to penalize each beam that is turned on by some fixed penalty β. The resulting penalized RTPP is

$$\min_{(\gamma,x,z)\in T} \gamma + \beta e'|z|_*, \beta > 0, \tag{4}$$

where

$$T = \{(\gamma, x, z)|(\gamma, x) \in S, z = \sum_p x(p, \cdot)\},$$

S is the feasible region of (3). In this model, the components of z represent the total radiation level of each beam. The penalty parameter β is chosen to balance the conflicting goals of minimizing critical structure dose, while simultaneously keeping the number of beams as small as possible. If $(\gamma^\beta, x^\beta, z^\beta)$ is a solution to (4), we shall call (γ^β, x^β) a β–minimum–support solution to the RTPP (3).

We now examine how solutions of (4) compare to solutions of (3). When we solve a feasible instance of (3) we get an optimal solution $(\gamma^\star, x^\star)$; when we solve (4) for a particular β we get a solution $(\gamma^\beta, x^\beta, z^\beta)$ for which γ^β can be no better than $\gamma^\star$, i.e. $\gamma^\beta \geq \gamma^\star$. The next result gives a bound on the difference between γ^β and $\gamma^\star$.

Theorem 1. (Error Bounds for the Objective Value of the Penalized Linear Program) If the RTPP (3) has solution $(\gamma^\star, x^\star)$ and the Penalized RTPP (4) has solution $(\gamma^\beta, x^\beta, z^\beta)$, then

$$\gamma^\beta - \gamma^\star < \beta e'|z^\star|_* \leq \beta n,$$

where

$$z^\star := \sum_{p\in P} x^\star(p, \cdot),$$

and n is the total number of beams in the model.

Proof. By construction, $(\gamma^\star, x^\star, z^\star)$ is feasible for (4); hence, since $(\gamma^\beta, x^\beta, z^\beta)$ is optimal for (4),

$$\gamma^\beta + \beta e'|z^\beta|_* \leq \gamma^\star + \beta e'|z^\star|_*.$$

Rearranging terms yields the inequality

$$\gamma^\beta - \gamma^\star \leq \beta(e'|z^\star|_* - e'|z^\beta|_*).$$

Furthermore, in order to satisfy the tumor dosage bounds set forth in the RTPP, $e'|z^\beta|_* \neq 0$. Hence $\gamma^\beta - \gamma^\star \leq \beta(e'|z^\star|_* - e'|z^\beta|_*) < \beta e'|z^\star|_* \leq \beta n$. □

3.2. An Exponential Approximation

In [9], Mangasarian discusses how to find minimum–support solutions to polyhedral concave programs of the form

$$\min_{\xi \in \mathcal{S}} f(\xi),$$

where $\mathcal{S}$ is a polyhedral set in $\Re^n$ and $f : \Re^n \to \Re$ is a concave function bounded below on $\mathcal{S}$. The approach is to replace the penalized problem

$$\min_{\xi \in \mathcal{S}} f(\xi) + \beta e'|\xi|_*,$$

by the approximation

$$\min_{\xi \in \mathcal{S}} f(\xi) + \beta e'(e - \varepsilon^{-\alpha\xi}), \tag{5}$$

where α is a positive scalar parameter. Figure 2 illustrates how the the exponential function $e - \varepsilon^{-\alpha z}$ approximates $|z|_*$ for $z \in \Re_+$.

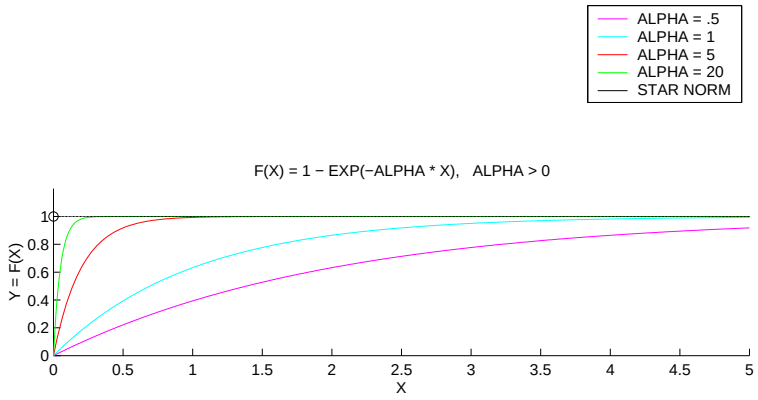

Figure 2. Comparison of Exponential Approximation $1 - \varepsilon^{-\alpha z}$ to $|z|_*$.

To solve (5), a successive linearization algorithm from [10] is used that finds a stationary point of (5) in a finite number of iterations. In this section, we make

some small modifications to Mangasarian's approach and associated theory to accomodate the RTP problem. The main task is to generalize the framework to find solutions that have minimum–support with respect to a *subset* of the variables. That is, we will consider polyhedral concave programs of the form

$$\min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta), \tag{6}$$

where $\mathcal{T}$ is a polyhedral set in $\Re^p \times \Re^m$ and $f : \Re^p \times \Re^m \to \Re$ is a concave function bounded below on $\mathcal{T}$. To find solutions (ξ,η) that have the fewest number of nonzero components of η, we replace the problem

$$\min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta|\eta|_* \tag{7}$$

by the approximation

$$\min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta'(e - \varepsilon^{-\alpha\eta}). \tag{8}$$

The remainder of this section adapts some theoretical results from [9] to our framework and presents a successive linearization algorithm for finding stationary points of (8).

The following Lemma is similar to [9, Lemma 2.2]. However, note that our result is true for *all* $\alpha \geq \alpha_0(\beta)$, whereas Mangasarian shows only that the result holds for *some* $\alpha \geq \alpha_0(\beta)$. This distinction is crucial in the proof of Theorem 3. We also simplify the statement of the Lemma by restricting $\mathcal{T}$ to be a subset of the positive orthant, thereby ensuring that there are no straight lines contained in $\mathcal{T}$.

Lemma 2. (Vertex Solution of Penalty Problem) Let $f : \Re^m \times \Re^p \to \Re$ be a concave function that is bounded below on the polyhedral set $\mathcal{T} \subset \Re^m_+ \times \Re^p_+$. Then for each $\beta > 0$, there exists $\alpha_0(\beta) > 0$ such that for all $\alpha \geq \alpha_0(\beta)$, the problem (8) has a vertex solution, and every vertex solution of (8) is also a vertex solution of (6).

Proof. For $\alpha > 0$, and $(\xi,\eta) \in \mathcal{T}$

$$(e - \varepsilon^{-\alpha\eta}) \leq |\eta|_*, \text{ since } \eta \geq 0. \tag{9}$$

It follows that

$$\min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta e' \left(e - \varepsilon^{-\alpha\eta}\right) \leq \inf_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta e'|\eta|_*. \tag{10}$$

For any $\alpha > 0$, (8) has a vertex solution. This follows from the fact that a concave function bounded below on a polyhedral set not containing lines going to infinity in both directions attains a minimum at a vertex of the polyhedral set [13, Corollaries 32.3.3 & 32.3.4]. Let V be the set or vertices (ξ, η) of $\mathcal{T}$ that solve (8) for arbitrarily large values of α. Since $\mathcal{T}$ has a finite number of vertices, V is nonempty. It follows that there exists some finite number $\alpha_0(\beta)$ such that for $\alpha \geq \alpha_0(\beta)$ the only vertex solutions of (8) are in V.

Let $\alpha \geq \alpha_0(\beta)$, and let $(\bar{\xi}, \bar{\eta})$ be a vertex solution of (8). Since $(\bar{\xi}, \bar{\eta})$ is in V, it repeatedly solves

$$\min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta e'(e - \varepsilon^{-\alpha_i \eta}), \tag{11}$$

for some sequence $\{\alpha_0, \alpha_1, \cdots\} \uparrow \infty$. Hence for each α_i in the sequence,

$$\begin{aligned} f(\bar{\xi},\bar{\eta}) + \beta e'\left(e - \varepsilon^{-\alpha_i \bar{\eta}}\right) &= \min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta e'\left(e - \varepsilon^{-\alpha_i \eta}\right) \\ &\leq \inf_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta e'|\eta|_* \qquad \text{by (10).} \end{aligned}$$

Letting $i \to \infty$ gives

$$f(\bar{\xi},\bar{\eta}) + \beta e'|\bar{\eta}|_* \leq \inf_{(\xi,\eta)\in\mathcal{T}} + \beta e'|\eta|_*. \tag{12}$$

Since $(\bar{\xi}, \bar{\eta})$ is a vertex of $\mathcal{T}$, it follows that $(\bar{\xi}, \bar{\eta})$ is a vertex solution of (6). □

The following theorem, which is based on [9, Theorem 2.3], relates the solution of the unpenalized problem (6) to the solution of the exponential approximation (8).

Theorem 3. Let $f : \Re^m \times \Re^p \to \Re$ be a concave function that is bounded below on the polyhedral set $\mathcal{T} \subseteq \Re^m_+ \times \Re^p_+$. Then there exists $\beta_0 > 0$ such that for each $0 < \beta \leq \beta_0$ there exists $\alpha(\beta)$ such that for all $\alpha \geq \alpha(\beta)$, (8) has a vertex solution, and each such solution is also a vertex solution of (6).

Proof. For each $\beta > 0$, (7) has a vertex solution. This follows from the fact that a concave function bounded below on a polyhedral set not containing lines going to infinity in both directions attains a minimum at a vertex of the polyhedral set [13, Corollary 32.3.3 & 32.3.4].

Let W be the set of vertices of $\mathcal{T}$ that solve (7) for arbitrarily small $\beta > 0$. Since $\mathcal{T}$ has a finite number of vertices, W is nonempty. It follows that there

exists $\beta_0 > 0$ such that for $0 < \beta \leq \beta_0$, the only vertex solutions of (7) are in W. Let $0 < \beta \leq \beta_0$, then (7) has a solution $(\bar{\xi}, \bar{\eta})$ that repeatedly solves

$$\min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta_i e'|\eta|_*, \tag{13}$$

for some sequence $\{\beta_1, \beta_2, \cdots\} \downarrow 0$. Hence for each β_i in the sequence,

$$\begin{aligned}(\bar{\xi}, \bar{\eta}) &\in \{\text{arg vertex min}_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) + \beta_i e'|\eta|_*\} \\ &= \{\text{arg vertex min}_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta) - f(\tilde{\xi},\tilde{\eta}) + \beta_i e'|\eta|_*\},\end{aligned} \tag{14}$$

where $(\tilde{\xi}, \tilde{\eta})$ is defined as a solution of (6), that is

$$(\tilde{\xi}, \tilde{\eta}) \in \mathcal{T} \text{ and } f(\tilde{\xi}, \tilde{\eta}) = \min_{(\xi,\eta)\in\mathcal{T}} f(\xi,\eta). \tag{15}$$

First we show by contradiction that

$$f(\bar{\xi}, \bar{\eta}) - f(\tilde{\xi}, \tilde{\eta}) \leq 0 \text{ (or equivalently } f(\bar{\xi}, \bar{\eta}) - f(\tilde{\xi}, \tilde{\eta}) = 0). \tag{16}$$

If not, then for sufficiently small β_i,

$$\frac{1}{\beta_i} > \max\left\{ \frac{e'|\tilde{\eta}|_* - e'|\bar{\eta}|_*}{f(\bar{\xi}, \bar{\eta}) - f(\tilde{\xi}, \tilde{\eta})}, \frac{1}{\beta_0} \right\}. \tag{17}$$

Thus, we get the contradiction

$$\begin{aligned} e'|\tilde{\eta}|_* &= e'|\tilde{\eta}|_* + \frac{1}{\beta_i}(f(\tilde{\xi}, \tilde{\eta}) - f(\tilde{\xi}, \tilde{\eta})) \\ &\geq e'|\bar{\eta}|_* + \frac{1}{\beta_i}(f(\bar{\xi}, \bar{\eta}) - f(\tilde{\xi}, \tilde{\eta})) \\ &> e'|\bar{\eta}|_* + e'|\tilde{\eta}|_* - e'|\bar{\eta}|_*, \end{aligned}$$

where the first inequality follows from (14) and the last inequality from (17). Hence (16) holds. We now show that $e'|\bar{\eta}|_*$ is a minimum over the set of solutions of (6), that is $(\bar{\xi}, \bar{\eta})$ has a minimal number of zeros besides being optimal for problem (6). We have for $\beta_i \leq \beta_0$ and all optimal (ξ, η),

$$\begin{aligned} e'|\bar{\eta}|_* &= e'|\bar{\eta}|_* + \frac{1}{\beta_i}(f(\bar{\xi}, \bar{\eta}) - f(\tilde{\xi}, \tilde{\eta})) \\ &\leq e'|\eta|_* + \frac{1}{\beta_i}(f(\xi, \eta) - f(\tilde{\xi}, \tilde{\eta})) \\ &\leq e'|\eta|_*, \end{aligned}$$

where the first inequality follows from (14) and the second inequality from $f(\xi,\eta) - f(\tilde{\xi},\tilde{\eta}) \leq 0$. Hence $(\bar{\xi}, \bar{\eta})$ is an optimal vertex solution. Finally, by Lemma 2, for each $\beta \leq \beta_0$, there exists $\alpha(\beta)$ such that for all $\alpha \geq \alpha(\beta)$, (8) has

a vertex solution, and each such solution is also a vertex solution of (7), which since $\beta \leq \beta_0$ is a vertex solution of (6). □

This theorem states that for sufficiently small β and sufficiently large α, a β–minimum–support solution to (8) will also be a solution to (6). The above proof is very similar to the proof of [9, Theorem 2.3]. However, we show that the result holds for *all* sufficiently small α, whereas the proof of [9, Theorem 2.3] shows only that the result holds for some large α.

The remainder of this paper works primarily with Lemma 2 rather than Theorem 3. We search for solutions to (4) rather than a solution to (3) with fewer beams because we can not guarantee that the beam reduction for the latter will be sufficient. Therefore we seek to exert some control over the number of beams in a solution to (4) by controlling β, where β does not necessarily satisfy $\beta \leq \beta_0$.

3.3. The Successive Linearization Algorithm

The preceding section established that to solve (4), it is sufficient, by Lemma 2, to solve

$$\min_{(\gamma,x,z)\in T} \gamma + \beta e'\left(e - \varepsilon^{-\alpha_i z}\right) \tag{18}$$

for α sufficiently large.

Instead of solving this NP-hard problem exactly, we will apply the successive linearization algorithm (SLA) from [10], which will find a stationary point for this problem in a finite number of iterations.

Successive Linearization Algorithm (SLA)
[9, Algorithm 3.1] Choose $\alpha > 0$ and $\beta > 0$ and starting point $(\gamma^0, x^0, z^0) \in \Re_+ \times \Re^n_+ \times \Re^p_+$ and a stopping tolerance ϵ. For $i = 0, 1, \ldots$, choose

$$(\gamma^{i+1}, x^{i+1}, z^{i+1}) \in \{\text{arg vertex } \min_{(\gamma,x,z)\in T} (\gamma - \gamma^i) + \alpha\beta(\varepsilon^{-\alpha z^i})'(z - z^i)\}. \tag{19}$$

Stop if

$$(\gamma^{i+1} - \gamma^i) + \alpha\beta(\varepsilon^{-\alpha z^i})'(z^{i+1} - z^i) = 0.$$

Note that the right hand side of (19) is the set of vertex solutions to a linear programming problem. To ensure that only vertex solutions are found, we use

a simplex algorithm rather than an interior–point method to solve these linear programs. Note also that considerable computation time can be saved by warm-starting each linear program using as a starting point the solution (γ^i, x^i, z^i) from the previous iteration.

While the successive linearization algorithm does not necessarily find a global minimum of (18), it does tend to converge to a local minimum, and always finds at least a stationary point. This is formalized in the following theorem which follows directly from [10, Theorem 3].

Theorem 4. [10, Theorem 3] The SLA generates a finite sequence of feasible iterates $\{(\gamma^1, x^1, z^1), (\gamma^2, x^2, z^2), \ldots, (\gamma^{\bar{i}}, x^{\bar{i}}, z^{\bar{i}})\}$ with strictly decreasing objective function values: $f(\gamma^1, x^1, z^1) > f(\gamma^2, x^2, z^2) > \cdots > f(\gamma^{\bar{i}}, x^{\bar{i}}, z^{\bar{i}})$, such that $(\gamma^{\bar{i}}, x^{\bar{i}}, z^{\bar{i}})$ satisfies the minimum principle necessary optimality condition:

$$(\gamma - \gamma^{\bar{i}}) + \alpha\beta(\varepsilon^{-\alpha z^{\bar{i}}})'(z - z^{\bar{i}}) = 0 \quad \forall (\gamma, x, z) \in T.$$

4. Implementation

In the previous sections we defined the dose deposition operator D and the three pixel subsets: critical, body, and tumor. In this section we begin by describing the processes used to calculate D and to create the critical, tumor and body sets. Finally, we discuss how to choose the parameters α and β.

4.1. Creating the Pixel Sets

The inequality constraints of the RTPP (3) require that we be able to classify pixels from the CT image into one of three different pixel subsets: critical, body, and tumor. We created a pixel selection routine with the library functions from MATLAB's Image Processing Toolbox. The user begins with a datafile that contains the patient image. The MATLAB routine reads the image file and stores the image. The user then uses a mouse to outline the body, critcal structures and tumor. MATLAB then creates a file in AMPL format that lists each pixel in either the body, critical, or tumor sets. This file is part of the data that is needed to define a particular instance of the problem within an AMPL model.

4.2. Calculating the Dose Deposition Operator

In Section 2.1 we briefly discussed the dose deposition operator D. This section describes how D is calculated. The center of the patient image, or isocenter, is 100 cm from the source point in the gantry. The gantry's position or angle describes the placement of the source point in relation to the fixed isocenter of the patient image. This ensures that the center of each beam passes through the center of the patient. In this model, we take into account the natural divergence of the radiation waves as they move away from the source point in the gantry. The outside boundaries of the beam make an angle of .19 radians from the center of the beam. Each beam is subdivided into a fixed number of pencils. For simplicity, we ignore the width and divergence of the pencils, and instead assume that all radiation is concentrated along the center line of each pencil. As radiation travels along this line, it deposits energy into each pixel that it intersects. The radiation beam attenuates at different rates, which depend on the distance traveled, the material passed through, and the starting energy level. To calculate how much radiation is deposited into each pixel, we order the pixels intersected by the center line starting from the pixel closest to the source point. The intensity I_{k+1} of the pencil as it exits the kth pixel is calculated using the formula

$$I_{k+1} = I_k e^{-\mu_k d_k}. \tag{20}$$

where I_k is the intensity of the pencil as it enters the kth pixel (or equivalently as it exits the $(k-1)$st pixel), d_k is the distance traveled within the kth pixel, and the constant μ_k is the attenuation factor for the pixel, which varies depending on whether the pixel represents tissue, bone, water, or air. I_0 is the intensity of the pencil at the source point, d_0 is the distance the pencil travels until it reaches the first pixel, and μ_0 is the attenuation factor for air.

The rate of energy deposition for the kth pixel is given by the equation

$$\Delta_k = I_k - I_{k+1} = (1 - e^{-\mu_k d_k}) I_k. \tag{21}$$

Thus, we assign $D(i,p,b) = \frac{\Delta_k}{I_0}$, where i is the index into the dose deposition array of the kth pixel in the path.

To assign the attenuation factors for each pixel, it is necessary to specify which pixels are bone, tissue, water, or air. This is accomplished using a MATLAB interface similar to that described in Section 4.1. The MATLAB routine

allows the user to look at a CT image and to classify pixels by outlining various structures in the image.

4.3. Choosing α and β

The successive linearization algorithm requires that two parameters, α and β, be chosen. Recall that β represents how heavily each active beam is penalized in (4). Intuitively, we need to choose β large enough to make the total number of beams acceptable while still satisfying all the therapeutic requirements. Theorem 1 provides some guidance for accomplishing this goal. In particular, using this theorem, a conservative choice of β is available via the formula

$$\beta := \frac{\Gamma - \gamma^\star}{|z^\star|_*} \tag{22}$$

where $\gamma^\star$ and $z^\star$ are determined by the solution to the unpenalized linear program (3) and are defined in Theorem 1 and Γ is the highest allowable dose to critical structures. This value of β is the largest choice that guarantees by Theorem 1 that the maximum critical structure dose in a solution to (4) does not exceed Γ. However, this choice of β is often too conservative, producing solutions with an inadequate reduction in the number of beams and very little change in the maximum critical structure dose γ. In such cases, a larger value of β can be tried.

Care must also be taken in choosing the smoothing parameter α. Recall that this parameter governs how closely the exponential penalty term in (18) approximates $|z|_*$. To gain some insight into how to choose α, consider the linear programs (19) solved at each iteration of the successive linearization algorithm. The coefficient of the jth beam z_j in the $(k+1)$st linear program is given by

$$g(\alpha) := \alpha\beta\varepsilon^{-\alpha z_j^k}, \tag{23}$$

where z_j^k is the beam intensity of the the jth beam in the solution of the previous LP. Figure 3 plots $g(\alpha)$ for various choices of z_j^k. Observe, that for fixed values of β and z_j^k, $g(\alpha)$ can be small both for very large and very small values of α. In either of these two extremes, the jth beam would not be sufficiently penalized to affect the solution. To maximize the penalty associated with beam j, we would use the formula $\alpha = \frac{1}{z_j^k}$, which maximizes $g(\alpha)$ for that particular beam. This

suggests the following heuristic for determining α: determine in advance some target minimum $\bar{z}$ for the strength of each nonzero beam. Then set

$$\alpha = \frac{1}{\bar{z}}. \tag{24}$$

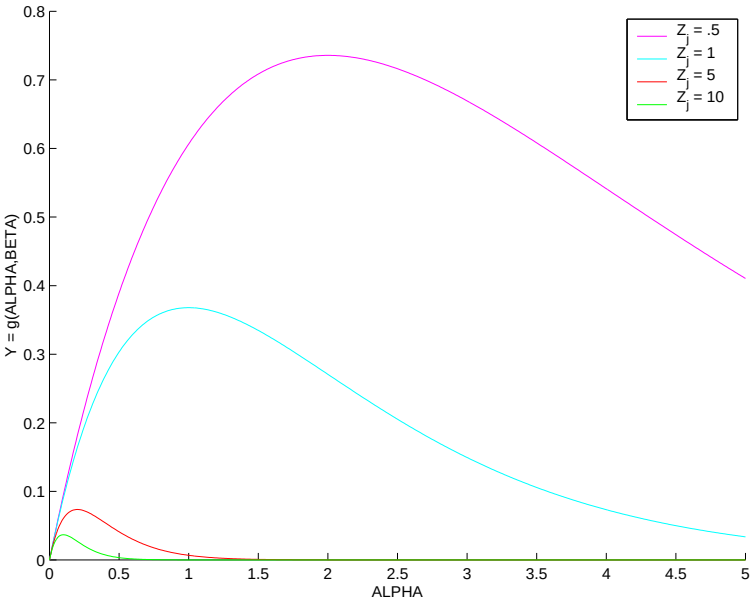

Figure 3. $g(\alpha, \beta) := \alpha\beta\varepsilon^{-\alpha z_j}$, for fixed β and $\alpha > 0$.

This choice of α results in the maximum possible penalty for beams whose intensities in the solution to the previous LP are near $\bar{z}$. Thus, beams of this intensity or lower will be encouraged to go to zero. We caution, however, that if $\bar{z}$ is chosen too large, then none of the penalties will be sufficient to affect the solution.

5. Results

Figure 4 illustrates a sample RTP problem that we used to test our algorithm. In this figure, the body and critical tissue are outlined in yellow, whereas the tumor is outlined in cyan. We used a 64×64 grid, with 180 beams and 64 pencils. The following bounds were specified: $T_l = 9, T_u = 10, O_u = 7$. Table 5 gives the attenuation factors used in the problem. The solution to the unpenalized RTP (3) used 72 different beams. The radiation doses deposited throughout

the body by this solution are shown in Figure 5. In this solution, the maximum dose to critical structures was $\gamma = 0.48$.

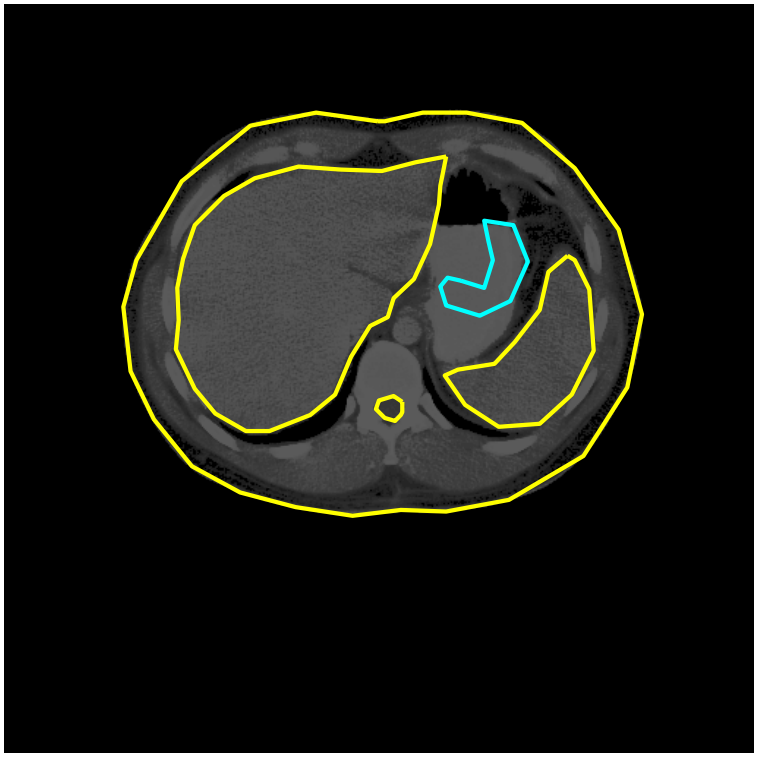

Figure 4. CT image of patient with critical structures and tumor outlined

Table 1
Attenuation Factors

tissue	.0306
bone	.0295
water	.0309
air	.0278

To calculate a minimum–support solution, we used the successive linearization algorithm with $\beta = 1$ and $\alpha = 0.05$, which corresponds to a target minimum beam intensity of $\bar{z} = 20$ (see (24)). The successive linearization algorithm required 17 iterations and produced a solution with 11 beams. This solution is

shown in Figure 6. In this solution, the maximum dose to critical structures was $\gamma = 2.11$.

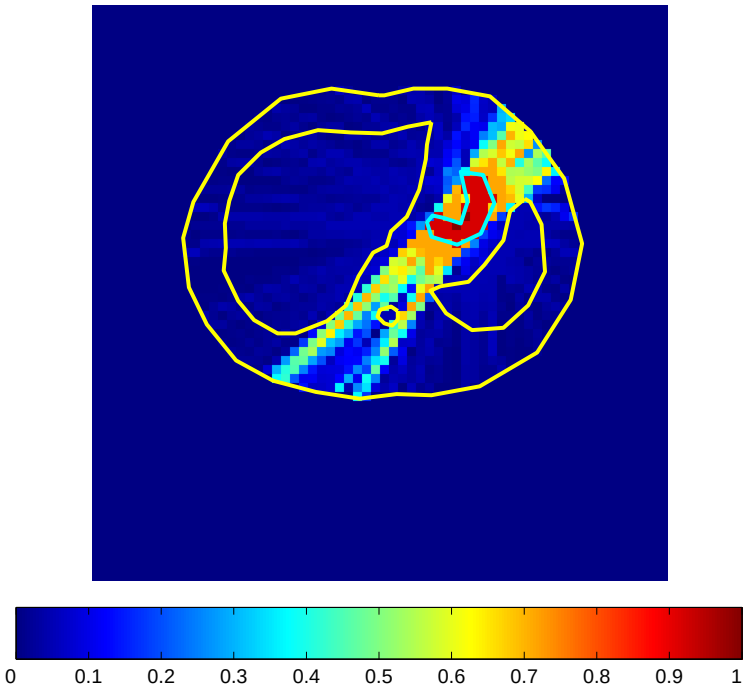

Figure 5. Energy deposition from solution to LP formulation

6. Conclusions

The results shown illustrate that our method can significantly reduce the number of beams in a radiotherapy plan without unacceptable degradation in the quality of the plan. The algorithm is very efficient in that it requires only a small number of iterations to converge. Moreover, while each iteration involves solving a linear program, the work involved can be minimized by warm–starting each successive linear program from the solution to the previous LP. Thus, the total time to find a minimum support solution is generally only a few times the cost of solving the unpenalized RTP.

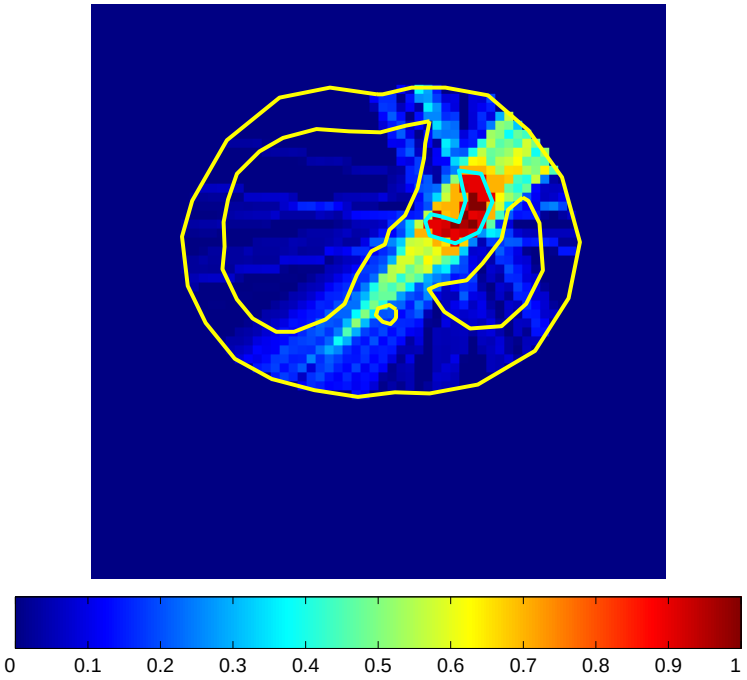

Figure 6. Energy deposition from minimu–support solution

References

[1] 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.

[2] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Belmont, Massachusetts, 1997.

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

[4] A. Holder. Designing radiotherapy plans with elastic constraints and interior point methods. Mathematics Technical Report 49, Trinity University, San Antonio, TX, 2000.

[5] 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.

[6] 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.

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

[8] W. Lodwick, S. McCourt, F. Newman, and S. Humphries. Optimization methods for radiation therapy plans. In *IMA Series in Applied Mathematics–Computational, Radiology*

and Imaging: Therapy and Diagnosis. Springer-Verlag, 1998.

[9] O.L. Magasarian. Minimum-support solutions of polyhedral concave programs. *Optimization.*, 45:149–162, 1999.

[10] O. L. Mangasarian. Solution of general linear complementarity problems via nondifferentiable concave minimization. *Acta Mathematica Vietnamica*, 22:199–205, 1997.

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

[12] 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.

[13] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

[14] I. Rosen, R. Lane, S. Morrill, and J. Belli. Treatment plan optimization using linear programming. *Medical Physics*, 18(2):141–152, 1990.

[15] D. M. Shepard, M. C. Ferris, G. Olivera, and T. R. Mackie. Optimizing the delivery of radiation to cancer patients. *SIAM Review*, 41:721–744, 1999.

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