An optimization framework for conformal radiation treatment planning

Jinho Lim* Michael C. Ferris† Stephen J. Wright†

David M. Shepard‡ Matthew A. Earl‡

December 2002

Abstract

An optimization framework for three-dimensional conformal radiation therapy is presented. In this type of therapy, beams of radiation are applied to a patient from different directions, where the aperture through which the beam is delivered from each direction is chosen to match the shape of the tumor, as viewed from that direction. Given a set of equispaced beam angles, a mixed-integer linear program can be solved to determine the most effective angles to be used in a treatment plan, and the weight (exposure time) to be used for each beam. This model can be enhanced to account for the use of wedge filters, which may be placed in front of a beam to produce a gradient in beam intensity across the aperture. Several techniques for strengthening the formulation (and therefore reducing the solution time) are described, and methods to control the dose-volume histogram implicitly for various parts of the treatment region using hot- and cold-spot control parameters are presented. The paper concludes with computational results that show the effectiveness of the proposed approach on two practical data sets.

1 Introduction

The optimization of radiation therapy for cancer has become an active research topic in recent years [1, 2, 3, 6, 17, 18, 29, 34, 35]. Many types of cancer are treated by applying radiation from external sources, aiming beams at the patient from a number of different angles. The general goal is to apply a significant total dose of radiation to the cancerous region (the tumor) while sparing the

*Industrial Engineering Department, 1513 University Ave. University of Wisconsin, Madison, WI 53706, U.S.A.

†Computer Sciences Department, 1210, W. Dayton Street, University of Wisconsin, Madison, WI 53706, U.S.A.

‡Department of Radiation Oncology, 22 South Greene St. University of Maryland School of Medicine, Baltimore, MD 21201, U.S.A.

surrounding normal tissues (especially sensitive structures near the tumor) from excessive radiation. The increasing sophistication of treatment devices allows delivery of complex and sophisticated treatment plans. Besides choosing from a wide range of different directions, a treatment planner can choose the shape of the aperture through which the beam is delivered, possibly applying a number of beams with different shapes from a single direction. In addition, wedges can be used to vary the intensity of the radiation across the aperture.

Radiation treatments are typically delivered using a linear accelerator (see Figure 1) with a multileaf collimator (see Figure 3(b)) housed in the head of the treatment unit. The shape of the aperture through which the beam passes can be varied by moving the computer-controlled leaves of the collimator. In *conformal* radiation therapy, the subject of this paper, three-dimensional anatomical information is used to shape the beam of radiation at each angle to match the shape of the tumor, as viewed from that angle. We refer to this approach to selecting the beam shape as the *beam's-eye view (BEV)* technique.

The goal in conformal radiation therapy is to provide a high probability of tumor control while minimizing radiation damage to surrounding normal tissue. This goal can accomplished by cross-firing beams from a number of beam directions. In practice, a dosimetrist usually uses a trial-and-error approach to determine how many beams of radiation should be used, which beam angles are most effective, and what weight should be assigned to each beam.

Often, additional flexibility is available to the dosimetrist, in the form of wedge filters that can be placed in front of the aperture to induce a gradient in the radiation field from one side of the aperture to the other. Wedge filters are particularly useful in treating cancers that lie near a curved patient surface, as is common in breast cancer. In addition to selecting beam directions and weights, the dosimetrist must decide whether it is appropriate to use a wedge, and if so, which orientation to choose for the wedge. It may be appropriate to use a combination of wedged and non-wedged beams from a single direction.

As we show in this paper, optimization techniques can be used to design these treatment plans automatically. Although the conformal techniques described above are the current standard of care in radiation therapy, used in the treatment of the vast majority of patients today, the benefits of automated treatment planning have gone largely unrealized. We focus on the conformal approach because requires little alteration to current clinical practices, and therefore has a good chance of rapid adoption. A more sophisticated treatment planning approach known as intensity modulated radiation therapy (IMRT), which is currently receiving a good deal of attention from optimization experts, allows a number of differently shaped beams to be delivered from each direction, thereby allowing a high degree of flexibility in modulating the intensity of the radiation delivered from each beam angle. Although this approach is undoubtedly interesting, its often-nonintuitive choice of aperture shapes represents a significant departure from current clinical practice, and therefore will require more time to be adopted widely.

We now given some further specifics of our proposed modeling and optimization methodology, and outline the remainder of the paper. The overall process

Figure 1: A Linear Accelerator

consists of determination of the beam's-eye view from each given angle; generation of the corresponding dose matrices; development of optimization models for the beam angles, beam weights, and wedge orientations; techniques to improve the optimization formulation and reduce the solution time; and techniques to control the dose-volume histogram (DVH[1]) on organs. By automating the treatment planning process, we aim to improve the quality of treatment plans while reducing the time required for planning each patient case.

Data for the model consists of the dose distribution matrices for beams from each angle and the dose requirements for different regions of the treatment space. The dose matrix for a given radiation beam consists of the radiation deposited by the beam into each of the small three-dimension regions ("voxels") into which the treatment area is divided. We divide the beam from each direction into a rectangular array of *pencil beams*, or *beamlets*, calculating the dose matrix independently for each, as described in Section 2.1. The beamlet dose matrices are used to identify the BEV, and the aggregate dose matrix for the BEV aperture is obtained by simply adding the contributions from the does matrices for the beamlets that make up the BEV.

The second important component of the data is specification of the tumor region and critical structures. Three-dimensional organ geometries are outlined by a physician on a set of CT or MRI images. The physician labels some of the voxels as PTV (for "Planning Target Volume," the tumor region) and others as OAR (for "Organ At Risk," also known as "sensitive structure" or "critical structure"). The desired or required dose information for each region is also specified.

In Section 3, we present several formulations of the treatment planning prob-

[1]A DVH is a plot that shows what fraction of volume of a structure receives dosage in a given range.

lem using linear programming (LP), quadratic programming (QP), and mixed-integer programming (MIP) approaches. In these optimizations, each "voxel" within the target volume typically requires at least a specified minimum amount of radiation to be delivered (a lower bound), while an upper bound is used for voxels in the sensitive structures and in the normal tissue. Since sensitive structures often are located close to target volumes, it is sometimes difficult or impossible to determine a treatment plan that satisfies the required bounds at every voxel. Instead, penalty terms can be included in the objective of the optimization problem that penalize violations of these bounds, with more significant violations incurring larger penalties.

Section 3.1 describes the problem in which the gantry angles for the treatment plan are fixed, and the task is merely to determine the beam weights for each angle. Several problem LP and QP formulations are presented; we explore the characteristics of each. In Section 3.2, we discuss the "angle selection" problem, in which the most effective angles (and their beam weights) are determined from among a set of candidate angles. A MIP model is used here, with binary variables indicating whether or not a particular angle used in the treatment. Treatments with fewer beams can be delivered more rapidly, and hence are generally preferred. We consider treatment plans using wedges in Section 3.3, using an extension of the MIP formulation for the angle selection problem. In Section 4, we describe several techniques for improving the formulation and reducing the solution time without degrading the solution quality for this model.

The quality of a treatment plan is typically specified and evaluated using a dose-volume histogram (DVH). Using the DVH as a guide, a planner may choose to allow a certain portion of voxels in each sensitive structure to exceed a specified dose, or require a large fraction of the volume to receive at least a certain dose. Due to the need to incorporate many binary variables into the optimization [13], formulation of a constraint of this type is not easy to handle using conventional optimization techniques. In Section 5, we show how the MIP formulations can be modified to account for the DVH constraints by using several control parameters.

In Section 6, we present computational results for the models described above on clinical data. We demonstrate in particular the usefulness of wedges in devising good treatment plans, and the effectiveness of our techniques for enforcing DVH constraints. We conclude by outlining our expected future research in Section 7.

2 Model Data Generation

2.1 Dose Matrices and Beam's-Eye View

A multileaf collimator located inside the head of the linear accelerator is used to shape the beam of radiation generated by the linear accelerator [14, 33]. To calculate the radiation dosages that can be delivered by a beam applied from

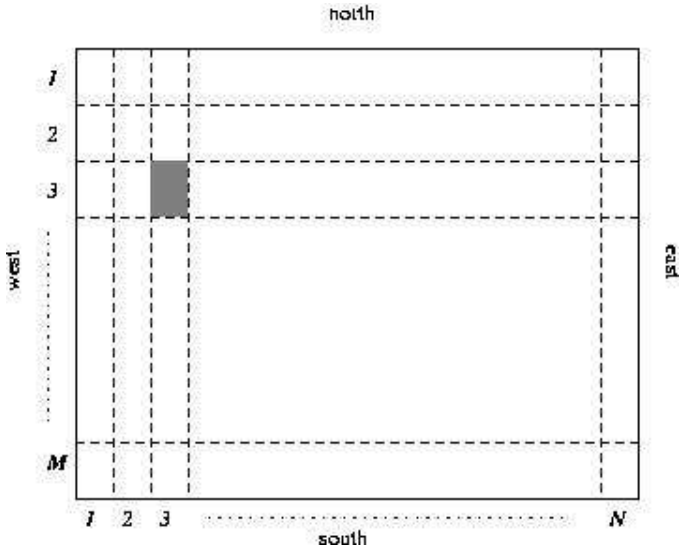
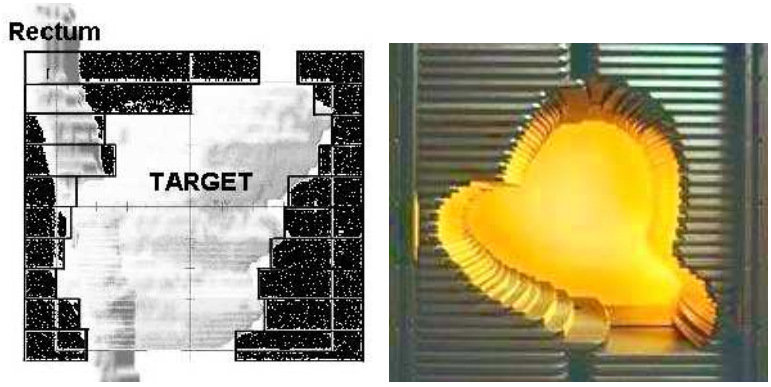


Figure 2: Division of Aperture into Pencil Beams (Shaded Area Represents One Beamlet)

a given angle, the rectangular aperture obtained by opening the collimator as widely as possible is divided into rectangular subfields arranged in a regular $M \times N$ rectangular pattern, as shown in Figure 2. Each of the subfields is called a *pencil beam* or *beamlet*. M represents the number of leaf pairs in the multileaf collimator, while N represents the number of possible settings we allow for each leaf. We identify each beamlet by the index pair (i, j), where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. In our work, the leaves of the multileaf collimator are 1 cm wide, and a pencil beam is assigned a length of 0.5 cm. Thus, for a 10 cm by 10 cm field, we would use $M = 10$ and $N = 20$, giving a total of 200 beamlets.

The dose distribution matrix for each pencil beam from each angle is calculated using a Monte Carlo technique, which simulates the track of individual radiation particles, for a large number of particles. A unit-intensity, non-wedged beam is assumed for the purposes of these calculations.

In conformal radiotherapy, the shape of each beam is set to match the beam's-eye view (BEV) of the tumor volume, which is essentially the projection of the three-dimensional shape of the tumor onto the plane of the multileaf collimator. One technique for determining the BEV is to employ a ray-tracing algorithm from the radiation source to the tumor volume, setting the beam's-eye view to include all of the rays that pass through the tumor volume. We use an alternative approach based on the dose matrices of the pencil beams. We include in the BEV all pencil beams whose field of significant dose intersects with the target region. To be specific, given a threshold value T, we include include a pencil beam in the BEV if its dose delivered to at least one voxel within the



(a) Beams-eye-view in a given angle

(b) Beams-eye-view can be produced using a multileaf collimator

Figure 3: Beams-eye-view

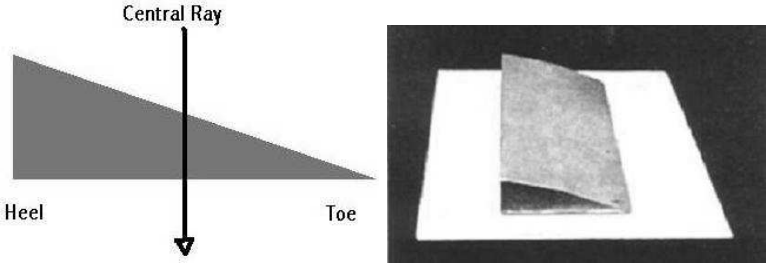
target region is at least $T\%$ of the dose delivered by that pencil beam to *any* voxel. Figure 3(a) shows an example of a BEV. Implementation of a BEV by a multileaf collimator is shown in Figure 3(b).

Once the BEV from a particular angle has been chosen, we can construct the dose matrix for the BEV aperture by simply summing the dose matrices of all the pencil beams that make up the BEV.

The choice of threshold parameter T is critical. If the value of T used in the determining the BEV is too small, the BEV overestimates the target, producing an aperture that irradiates not only the target but also nearby normal tissue and organs at risk. On the other hand, if the value of T is too large, the BEV underestimates the target, and the optimizer might not be able to find a solution that adequately delivers radiation dose within the required range to all parts of the target. The best value of T to use depends somewhat on the shape of the tumor. We choose T as the minimum value such that the resulting BEVs provide a complete 2D coverage of the target from all beam angles considered in the problem. Based on our experiments, a value of T of between 10% and 15% appears to be appropriate.

2.2 Use of Wedges

Wedges (also called wedge filters) are used to produce a gradient in the radiation field across the aperture. As shown in Figure 4, a wedge is a tapered metallic block with a the thick side (the *heel*) and a thin edge (the *toe*). Less radiation is transmitted through the heel of the wedge than through the toe. Figure 4(b) shows an external 45° wedge, so named because it produces isodose lines that are



(a) A wedge filter (b) An external wedge

Figure 4: Wedges

oriented at approximately 45°, as illustrated in Figure 5. Figure 5(a) shows the dose attenuation pattern produced when no wedge are used, while Figure 5(b) is the dose contour map resulting from use of a wedge. (In this example, the wedge is oriented with its heel on the right side of the figure.) As well as tilting the isodose lines, the wedge produces a general attenuation of the dose as compared with the open beam. We assume that the wedge can be oriented in four ways: with its heel aligned with each of the four sides of the rectangular aperture obtained by opening all the leaves of the collimator. We refer to these orientations as "north," "south," "east," and "west," according to the edge of the aperture in question, as indicated in Figure 2.

We include a wedge transmission factor τ in the model to account for the effect of the wedge on the dose delivered to the voxels in the treatment region. Wedges are characterized by two constants τ_0 and τ_1, with $0 \leq \tau_0 < \tau_1 \leq 1$ that indicate the smallest and largest transmission factors for the wedge among all pencil beams in the field. Specifically, τ_0 indicates the factor by which the dose is decreased for the pencil beams along the edge of the aperture with which the heel of the wedge is aligned. Correspondingly, τ_1 indicates the transmission factor along the opposite (thin) edge. When the heel lies along the west edge, the transmission factor for beamlet (i, j) is calculated as follows:

$$\tau_{ij}^{\text{west}} = \tau_0 + \frac{j - 0.5}{N}(\tau_1 - \tau_0), \quad i = 1, 2, \ldots, M, \;\; j = 1, 2, \ldots, N. \qquad (1)$$

When the wedge is oriented with its heel at the top (north) of the field, the weight applied to the (i, j) beamlet is

$$\tau_{ij}^{\text{north}} = \tau_0 + \frac{i - 0.5}{M}(\tau_1 - \tau_0), \quad i = 1, 2, \ldots, M, \;\; j = 1, 2, \ldots, N. \qquad (2)$$

The shift of 0.5 is introduced in both formulae to capture the transmission factor at the *center* of each beamlet.

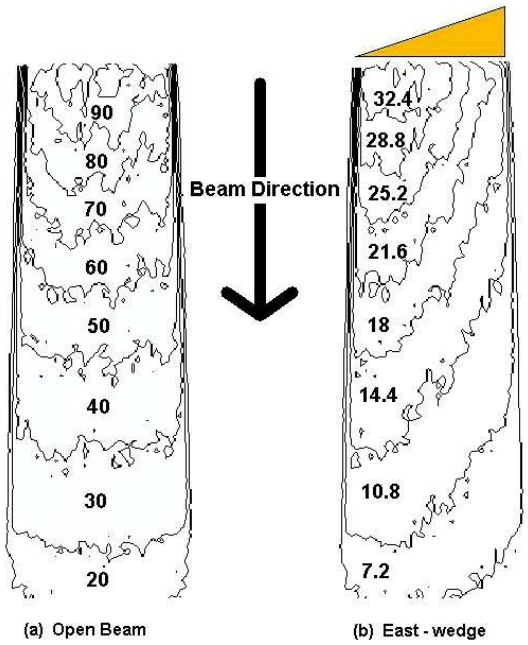


Figure 5: Dose contour maps: Effect of wedge on the dose distribution

Two different wedge systems are used in clinical practice. In the first system, four different wedges with angles 15°, 30°, 45°, and 60° are available, and the therapist is responsible for selecting one of these wedges and inserting it with the correct orientation. In the second system, a single 60° wedge (the *universal wedge*) is permanently located on a motorized mount located within the head of the treatment unit. This wedge can be rotated to the desired orientation or removed altogether, as required by the treatment plan. By using the universal wedge appropriately, all plans deliverable by the four-wedge system can be reproduced, as we show in Appendix A. Hence, we assume in this paper that the universal wedge is used.

3 Formulating the Optimization Problems

3.1 Beam Weight Optimization

We start with the simplest model, in which we assume that the angles from which beams are to be delivered are selected in advance, that wedges are not used, and that the apertures are chosen to be the beam's-eye view from each respective angle. It remains only to determine the intensities of the beams (that is, the beam weights) to be used from each angle.

We now introduce notation that is used below and in later sections. The set

of beam angles is denoted by $\mathcal{A}$. We use $\mathcal{T}$ to denote the set of all voxels that make up the target structure, $\mathcal{S}$ to denote the voxels in the sensitive structure, and $\mathcal{N}$ to be the voxels in the normal tissue. We use θ to denote the prescribed dose level for each target voxel, while the hot spot control parameter ϕ defines a dose level for each voxel in the critical structure that we would prefer not to exceed. The beam weight delivered from angle A is denoted by w_A, and the dose contribution to voxel (i, j, k) from a beam of weight 1 from angle A is denoted by $\mathcal{D}_{A,(i,j,k)}$. (It follows that a beam of weight w_A will produce a dose of $w_A\mathcal{D}_{A,(i,j,k)}$ in voxel (i, j, k).) We obtain the total dose $D_{(i,j,k)}$ to voxel (i, j, k) by summing the contributions from all angles $A \in \mathcal{A}$. We use $\mathcal{D}_{A,\Omega}$ (and D_Ω) to denote the submatrices consisting of the elements $\mathcal{D}_{A,(i,j,k)}$ (and $\mathcal{D}_{(i,j,k)}$) for all (i, j, k) in a given set of voxels Ω.

The beam weights w_A, $A \in \mathcal{A}$, which are of course nonnegative, are the unknowns in the optimization problem. The general form of this problem is as follows.

$$\begin{array}{ll} \min\limits_{w} & f(D_\Omega) \\ \text{s.t.} & \\ & D_\Omega = \sum\limits_{A\in\mathcal{A}} w_A \cdot \mathcal{D}_{A,\Omega}, \quad \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\ & w_A \geq 0, \qquad\qquad\qquad \forall A \in \mathcal{A}. \end{array} \tag{3}$$

The choice of objective function $f(D_\Omega)$ in (3) depends on the specific goal that the treatment planner wants to achieve. Two common goals are to control the integral dose to organs and to control *cold spots* (underdose to the target region) and *hot spots* (overdose). In general, the objective function measures the mismatch between the prescription and the delivered dose. For voxels in the target region $\mathcal{T}$, there are terms that penalize any difference between the delivered dose and the prescribed dose θ. For voxels in each sensitive structure $\mathcal{S}^i (i = 1, \ldots, |OAR|)$, there are terms that penalize the amount of dose in excess of ϕ_i, the desired upper bound on the dose to voxels in sensitive structure i. However, for simplicity of explanation, we only consider a single sensitive structure in the problem formulations in this paper. For voxels in the normal region $\mathcal{N}$, the desired dose is zero, so the objective usually includes terms that increase as the amount of dose delivered to these voxels increases. There may be more than one sensitive structure in a treatment planning problem.

Let parameters λ_t, λ_s, and λ_n be nonnegative weighting factors applied to the objective terms for the target, sensitive, and normal voxels, respectively. Two common ways to define the objective are to use the L_1-norm (which penalizes the absolute value of deviation from the prescribed dose on each voxel, weighted by the factors just defined) and the sum of squares of the deviations, again weighted according to the region in which each voxel lies. These techniques lead to the following two definitions:

$$\lambda_t \|D_\mathcal{T} - \theta e_\mathcal{T}\|_1 + \lambda_s \|(D_\mathcal{S} - \phi e_\mathcal{S})_+\|_1 + \lambda_n \|D_\mathcal{N}\|_1, \tag{4}$$

$$\lambda_t \|D_\mathcal{T} - \theta e_\mathcal{T}\|_2^2 + \lambda_s \|(D_\mathcal{S} - \phi e_\mathcal{S})_+\|_2^2 + \lambda_n \|D_\mathcal{N}\|_2^2. \tag{5}$$

The notation $(\cdot)_+ := \max(\cdot, 0)$ in the second term defines the overdose to voxels in the sensitive region, while $e_{\mathcal{T}}$ is the vector whose components are all 1 and whose dimension is the same as the cardinality of $\mathcal{T}$ (similarly for $e_{\mathcal{S}}$). The terms in (4) and (6) are approximations to the L_1 and squared-L_2 integrals of the deviations from prescription over each region of interest.

A planner can also use an average dose deviation for each structure by dividing the integral dose over the number of voxels in the structure:

$$\lambda_t \frac{\|D_{\mathcal{T}} - \theta e_{\mathcal{T}}\|_p}{\text{card}\,(\mathcal{T})} + \lambda_s \frac{\|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_p}{\text{card}\,(\mathcal{S})} + \lambda_n \frac{\|D_{\mathcal{N}}\|_p}{\text{card}\,(\mathcal{N})}, \quad p = 1, 2,$$

where $\text{card}\,(\mathcal{T})$, $\text{card}\,(\mathcal{S})$, and $\text{card}\,(\mathcal{N})$ denote the number of voxels in the target region, the sensitive structure, and the normal region, respectively. The use of these factors in the denominator facilitates easier choice of λ_t, λ_s, and λ_n, and removes some of the dependence of the plan on the relative sizes of each region.

An objective function based on L_∞-norm terms (6) allows effective penalization of "hot spots" in sensitive regions and of cold spots in the target. We define such a function as follows:

$$\lambda_t \|(D_{\mathcal{T}} - \theta e_{\mathcal{T}})\|_\infty + \lambda_s \|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_\infty + \lambda_n \|D_{\mathcal{N}}\|_\infty. \tag{6}$$

Combinations of the objective functions above can also be used to achieve specific treatment goals, as we describe later in this section.

3.1.1 Quadratic Programming Formulation

If we use a weighted sum-of-squares objective of the form (5), the 3D conformal radiation treatment planning problem is a quadratic program (QP). We slightly modify (5) by including the cardinality of the sets $\mathcal{T}$, $\mathcal{S}$, and $\mathcal{N}$ explicitly in the weighting terms. We arrive at the following QP formulation (a particular case of (3)):

$$\begin{array}{ll} \min\limits_{w} & \lambda_t \dfrac{\|D_{\mathcal{T}} - \theta e_{\mathcal{T}}\|_2^2}{\text{card}\,(\mathcal{T})} + \lambda_s \dfrac{\|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_2^2}{\text{card}\,(\mathcal{S})} + \lambda_n \dfrac{\|D_{\mathcal{N}}\|_2^2}{\text{card}\,(\mathcal{N})} \\ \text{s.t.} & \\ & \begin{array}{rcll} D_\Omega & = & \sum_{A \in \mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\ w_A & \geq & 0, & \forall A \in \mathcal{A}. \end{array} \end{array} \tag{7}$$

By introducing variables $V_{(i,j,k)}$, $(i,j,k) \in \mathcal{S}$ to denote the excess dose over the upper bound ϕ in the sensitive region $\mathcal{S}$, we can rewrite (7) as follows:

$$\begin{array}{ll} \min\limits_{w} & \lambda_t \dfrac{\|D_{\mathcal{T}} - \theta e_{\mathcal{T}}\|_2^2}{\text{card}\,(\mathcal{T})} + \lambda_s \dfrac{\|V_{\mathcal{S}}\|_2^2}{\text{card}\,(\mathcal{S})} + \lambda_n \dfrac{\|D_{\mathcal{N}}\|_2^2}{\text{card}\,(\mathcal{N})} \\ \text{s.t.} & \\ & \begin{array}{rcll} D_\Omega & = & \sum_{A \in \mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\ V_{\mathcal{S}} & \geq & D_{\mathcal{S}} - \phi e_{\mathcal{S}}, & \\ V_{\mathcal{S}} & \geq & 0, & \\ w_A & \geq & 0, & \forall A \in \mathcal{A}. \end{array} \end{array} \tag{8}$$

Note that the expression for $D_{\mathcal{N}}$ can be explicitly substituted into the objective function removing all such variables. Eliminating some of $D_{\mathcal{T}}$ and $D_{\mathcal{S}}$ may be done in certain cases, but the substitutions are more complex.

3.1.2 Least-Absolute-Value Formulation: Linear Programming

The absolute-value terms in (4) do not penalize large violations as much as the squared terms in (5). However, they allow the problem to be formulated as a linear program. By including the cardinalities of $\mathcal{T}$, $\mathcal{S}$, and $\mathcal{N}$ in the weighting factors of (4), we obtain another special case of (3):

$$\begin{array}{ll} \min\limits_{w} & \lambda_t \dfrac{\|D_{\mathcal{T}} - \theta e_{\mathcal{T}}\|_1}{\operatorname{card}(\mathcal{T})} + \lambda_s \dfrac{\|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_1}{\operatorname{card}(\mathcal{S})} + \lambda_n \dfrac{\|D_{\mathcal{N}}\|_1}{\operatorname{card}(\mathcal{N})} \\ \text{s.t.} & \\ & \begin{array}{lcll} D_\Omega & = & \sum_{A\in\mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ w_A & \geq & 0, & \forall A \in \mathcal{A}. \end{array} \end{array} \tag{9}$$

To recast this problem as a linear program, we introduce variables $V_{(i,j,k)}$ for $(i,j,k) \in \mathcal{T}\cup\mathcal{S}$ to represent violations from the desired doses or dose intervals on the PTV and the OAR. We can then write (9) equivalently as follows:

$$\begin{array}{ll} \min\limits_{w} & \lambda_t \dfrac{e_{\mathcal{T}}^T V_{\mathcal{T}}}{\operatorname{card}(\mathcal{T})} + \lambda_s \dfrac{e_{\mathcal{S}}^T V_{\mathcal{S}}}{\operatorname{card}(\mathcal{S})} + \lambda_n \dfrac{e_{\mathcal{N}}^T D_{\mathcal{N}}}{\operatorname{card}(\mathcal{N})} \\ \text{s.t.} & \\ & \begin{array}{lcll} D_\Omega & = & \sum_{A\in\mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ V_{\mathcal{T}} & \geq & D_{\mathcal{T}} - \theta e_{\mathcal{T}}, & \\ V_{\mathcal{T}} & \geq & \theta e_{\mathcal{T}} - D_{\mathcal{T}}, & \\ V_{\mathcal{S}} & \geq & D_{\mathcal{S}} - \phi e_{\mathcal{S}}, & \\ V_{\mathcal{S}} & \geq & 0, & \\ w_A & \geq & 0, & \forall A \in \mathcal{A}. \end{array} \end{array} \tag{10}$$

Note that since the elements $\mathcal{D}_{A,(i,j,k)}$ of the dose matrix and w_A of the weight vector are all nonnegative, the elements of the dose vector $D_{\mathcal{N}}$ are also nonnegative. Hence, in the last term of the objective, we are justified in making the substitution $\|D_{\mathcal{N}}\|_1 = e_{\mathcal{N}}^T D_{\mathcal{N}}$.

3.1.3 Min-Max Formulation: Linear Programming

Sometimes, it is important in radiation treatment to minimize the maximum dose violation on organs. Min-max formulations based on (6) can be used for this purpose:

$$\begin{array}{ll} \min\limits_{w} & \lambda_t \|D_{\mathcal{T}} - \theta\|_\infty + \lambda_s \|(D_{\mathcal{S}} - \phi)_+\|_\infty + \lambda_n \|D_{\mathcal{N}}\|_\infty \\ \text{s.t.} & \\ & \begin{array}{lcll} D_\Omega & = & \sum_{A\in\mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ w_A & \geq & 0, & \forall A \in \mathcal{A}. \end{array} \end{array} \tag{11}$$

An LP formulation for (11) can be generated by introducing extra scalar variables, V_t, V_s, and V_n into the problem as follows.

$$\begin{array}{ll} \min\limits_{w} & \lambda_t V_t \ + \ \lambda_s V_s \ + \ \lambda_n V_n \\ \text{s.t.} & \\ & \begin{array}{rcll} D_\Omega & = & \sum_{A\in\mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ V_t e_\mathcal{T} & \geq & D_\mathcal{T} - \theta e_\mathcal{T}, & \\ V_t e_\mathcal{T} & \geq & \theta e_\mathcal{T} - D_\mathcal{T}, & \\ V_s e_\mathcal{S} & \geq & D_\mathcal{S} - \phi e_\mathcal{S}, & \\ V_n e_\mathcal{N} & \geq & D_\mathcal{N}, & \\ 0 & \leq & V_t,\ V_s,\ V_n, & \\ 0 & \leq & w_A, & \forall A\in\mathcal{A}. \end{array} \end{array} \tag{12}$$

3.1.4 Composite Formulations

In the sections above, we introduced three possible problem formulations for the optimization problem (3) based on specific treatment goals. Often, the planner's goals are quite specific to the case at hand. For example, the planner may wish to keep the maximum dose violation on the target low, and also to control the integral dose violation on the OAR and the normal tissue. These goals can be met by defining the objective to be a weighted sum of the relevant terms. For the given example, we would obtain the following:

$$\begin{array}{ll} \min\limits_{w} & \lambda_t \|D_\mathcal{T} - \theta e_\mathcal{T}\|_\infty + \lambda_\mathcal{S} \dfrac{\|(D_\mathcal{S} - \phi e_\mathcal{S})_+\|_1}{\text{card}\,(\mathcal{S})} + \lambda_n \dfrac{\|D_\mathcal{N}\|_1}{\text{card}\,(\mathcal{N})} \\ \text{s.t.} & \\ & \begin{array}{rcll} D_\Omega & = & \sum_{A\in\mathcal{A}} w_A \mathcal{D}_{A,\Omega}, & \Omega = \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ w_A & \geq & 0, & \forall A\in\mathcal{A}. \end{array} \end{array} \tag{13}$$

In practice, voxels on the PTV that receive dose within specified limits may be acceptable as a treatment plan. Furthermore, voxels receive below the lower dose specification (cold spots) may get penalized more severely than *hot spots* on the PTV. Therefore, we consider the following two definitions of $f(D_\Omega)$ in (3) as follows:

$$\begin{array}{rcl} f(D_\Omega) & = & \lambda_t^+ \dfrac{\|(D_\mathcal{T} - \theta_u e_\mathcal{T})_+\|_1}{\text{card}\,(\mathcal{T})} + \lambda_t^- \dfrac{\|(\theta_L e_\mathcal{T} - D_\mathcal{T})_+\|_1}{\text{card}\,(\mathcal{T})} \\ & & + \lambda_s \dfrac{\|(D_\mathcal{S} - \phi e_\mathcal{S})_+\|_1}{\text{card}\,(\mathcal{S})} + \lambda_n \dfrac{\|D_\mathcal{N}\|_1}{\text{card}\,(\mathcal{N})}, \end{array} \tag{14}$$

$$\begin{array}{rcl} f(D_\Omega) & = & \lambda_t^+ \|(D_\mathcal{T} - \theta_u e_\mathcal{T})_+\|_\infty + \lambda_t^- \|(\theta_L e_\mathcal{T} - D_\mathcal{T})_+\|_\infty \\ & & + \lambda_s \dfrac{\|(D_\mathcal{S} - \phi e_\mathcal{S})_+\|_1}{\text{card}\,(\mathcal{S})} + \lambda_n \dfrac{\|D_\mathcal{N}\|_1}{\text{card}\,(\mathcal{N})}. \end{array} \tag{15}$$

In these objectives, θ_L is the target cold-spot control parameter. If the dosage of a voxel in $\mathcal{T}$ falls below θ_L, a penalty term for the violation is added to the objective. Likewise, a voxel on the PTV incurs a penalty if the dosage at the voxel exceeds θ_u.

It should be understood that in all models we describe in this paper such a separation of *hot* and *cold* spots is possible. However, we simplify the exposition throughout by using a combined objective function. Alternative objectives have been discussed elsewhere. For example, the papers [22, 26] use score functions to evaluate and compare different plans.

Building on the beam-weight optimization formulations described above, we now consider extended models in which beam angles and wedges are included in the optimization problem.

3.2 Beam orientation optimization

In the previous section, we showed how to choose the beam weights in an optimal fashion, given a set $\mathcal{A}$ of specified beam orientations. We now consider the problem of selecting a subset of at most K beam angles from a set $\mathcal{A}$ of candidates, simultaneously choosing optimal weights for the selected beams. A treatment plan involving few beams (say, 3 to 5) generally is preferable to one of similar quality that uses more beams because it requires less time and effort to deliver in the clinic.

We introduce binary variables ψ_A, $A \in \mathcal{A}$, that indicate whether or not angle A is selected to be one of the treatment beam orientations. The value $\psi_A = 0$ indicates that angle A is not used, so the weight for this beam must satisfy $w_A = 0$. When $\psi_A = 1$, on the other hand, the beam from angle A may have a positive weight. Both conditions are enforced by adding the constraint $w_A \leq M \cdot \psi_A$ to the model, where M is a upper bound on the beam weights (discussed below in Section 4.1). The resulting mixed programming formulation (10) is as follows:

$$
\begin{array}{lrcll}
\min\limits_{w,\psi} & f(D_\Omega) & & & \\
\text{s.t.} & & & & \\
& D_\Omega & = & \sum\limits_{A\in\mathcal{A}} \mathcal{D}_{A,\Omega} \cdot w_A, & \Omega = \{\mathcal{T} \cup \mathcal{S} \cup \mathcal{N}\} \\
& 0 & \leq & w_A \leq M\psi_A, & \forall A \in \mathcal{A}, \\
& \sum_{A\in\mathcal{A}} \psi_A & \leq & K, & \\
& \psi_A & \in & \{0,1\}, & \forall A \in \mathcal{A}.
\end{array}
\tag{16}
$$

Our approach relies on a mixed integer formulation for which standard optimization codes can be applied. Some theoretical considerations of optimizing beam orientations are also discussed in [1]. In general, using more beams typically produces better quality treatment plans. The down side, however, is that the time to treat the patients is longer when more beams are used. Furthermore,

it has been shown that, when many beams are used, (say ≥ 5), beam orientation becomes less important in the overall optimization [7, 8, 28]. In many cited cases, the objective is to find a minimum number of beams that satisfy the treatment goals.

The beam angles and the weights can be optimized either sequentially or simultaneously. Most of the earlier work in the literature uses sequential schemes [5, 15, 20, 24, 25], in which a certain number of beam angles are fixed first, and their weights are subsequently determined. Rowbottom *et al* [23] optimize both variables simultaneously. To reduce the initial search space, a heuristic approach to remove some beam orientations *a priori* is used, while the overall optimization problem is solved with the simplex method and simulated annealing.

A different approach has been proposed by Hass *et al* [16]. They address a geometrical formulation of the coplanar beam orientation problem combined with a hybrid multi-objective genetic algorithm. The approach is demonstrated by optimizing the beam orientation in two dimensions, with the objectives being formulated using planar geometry. Their algorithm attempts to replicate the approach of a treatment planner whilst reducing the amount of computation required. Hybrid genetic search operators have been developed to improve the performance of the genetic algorithm by exploiting problem-specific features. When the approach is applied without constraining the number of beams, the solution produces an indication of the minimum number of beams required. Webb [30] applies simulated annealing approach on a two dimensional treatment planning problem. Three-dimensional problems using simulated annealing approach are addressed in [23, 31, 32, 33].

3.3 Wedge orientation optimization

Wedges may be placed in front of a beam to deliver a nonuniform dose distribution across the aperture. Several researchers have studied treatment planning problem with wedge filters [9, 10, 19, 27, 36, 37]. Xing *et al* [36] demonstrate the use of optimizing the beam weights for an open field and two orthogonal wedged fields. Li *et al* [19] presents an optimization algorithm for the wedge orientation selection and the beam weights. Design of treatment plans involving wedges are discussed in [10, 19, 27, 36, 37]. The papers [27, 36, 37] discuss selection of wedge angles; in particular, Sherouse [27] describes a mathematical basis for selection of wedge angle and orientation. However, it has been noted that including wedge angle selection in the optimization makes for excessive computation time [37].

We consider four possible wedge orientations at each beam angle: "north", "south", "east", and "west". At each angle A, we calculate dose matrices for the beams-eye view aperture and for each of these four wedge settings, along with the dose matrix for the no-wedge setting (the open beam), as used in the formulations above. We use $\mathcal{F}$ to denote the set of wedge settings; $\mathcal{F}$ contains 5 elements in our case. Extending our previous notation, the dose contribution to voxel (i, j, k) from a beam delivered from angle A with wedge setting F is denoted by $\mathcal{D}_{A,F,(i,j,k)}$, and we use $\mathcal{D}_{A,F,\Omega}$ to denote the collection of doses for all

(i, j, k) in some set Ω. The weight assigned to a beam from angle A with wedge setting F is denoted by $w_{A,F}$, while the binary variable $\pi_{A,F}$ determines whether or not we use a beam from angle A with wedge setting F in the treatment plan.

The optimization problem is to select beams and optimizing weights when wedges are present. The new formulation is obtained not by simply replacing $\mathcal{A}$ by $\mathcal{A}\times\mathcal{F}$ in the discussion above, since there are some additional considerations. First, in selecting beams, we do not wish to place a limit on the total number of beams delivered, as in Section 3.2, but rather on the total number of distinct angles used. In other words, we are prepared to allow multiple beams to be delivered from a given angle for the same "cost" as a single beam from that angle; that is,

$$\psi_A \quad \geq \quad \pi_{A,F}, \quad \forall A \in \mathcal{A},\, \forall F \in \mathcal{F}.$$

This constraint models the clinical situation reasonably well, since changing the wedge orientation takes little time relative to the time required to move the gantry and possibly shift the patient.

The second consideration is that we do not wish to deliver two beams from the same angle for two diametrically opposite wedge settings. We enforce this restriction by adding the following constraints:

$$\begin{array}{rcl} 1 & \geq & \pi_{A,north} + \pi_{A,south}, \\ 1 & \geq & \pi_{A,west} + \pi_{A,east}. \end{array} \tag{17}$$

(17) limits the number of wedge orientations to be less than three in each angle for the treatment. The resulting mixed integer programming model is now as follows:

$$\begin{array}{lrcll} \min\limits_{w,\psi,\pi} & f(D_\Omega) & & & \\ \text{s.t.} & & & & \\ & D_\Omega & = & \displaystyle\sum_{A\in\mathcal{A},F\in\mathcal{F}} w_{A,F}\mathcal{D}_{A,F,\Omega}, & \Omega \in \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ & M\pi_{A,F} & \geq & w_{A,F}, & \\ & \psi_A & \geq & \pi_{A,F}, & \\ & K & \geq & \sum_{A\in\mathcal{A}} \psi_A, & \\ & 1 & \geq & \pi_{A,north} + \pi_{A,south}, & \\ & 1 & \geq & \pi_{A,west} + \pi_{A,east}, & \\ & w_{A,F} & \geq & 0, & \forall A \in \mathcal{A},\ \forall F \in \mathcal{F}, \\ & \psi_A\,, \pi_{A,F} & \in & \{0,1\}, & \forall A \in \mathcal{A},\ \forall F \in \mathcal{F}. \end{array} \tag{18}$$

In comparing (18) with (16), we see that the amount of data to be stored increases by a factor of $|\mathcal{F}|$. The number of binary variables also increases by a factor of $|\mathcal{F}| + 1$, although the nature of the new variables $\pi_{A,F}$ and the new constraints is such that the complexity of the problem is increased by less than this factor would suggest.

We show in Appendix B that a plan calling for two nonzero weights for two diametrically opposed beams can be replaced by an equivalent plan requiring a positive weight for an open beam along with a positive weight for one of the two original beams. Hence any solution with nonzero weights on diametrically opposed wedges can be postprocessed easily to reduce one of the weights to zero. This suggests an alternative to the formulation above, in which we dispense with the binary variables $\pi_{A,F}$ and the constraints in which they appear:

$$\begin{array}{ll} \min\limits_{w,\psi} & f(D_\Omega) \\ \text{s.t.} & \\ & \begin{array}{rcll} D_\Omega & = & \sum\limits_{A\in\mathcal{A},F\in\mathcal{F}} w_{A,F}\mathcal{D}_{A,F,\Omega}, & \Omega \in \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\ M\psi_A & \geq & w_{A,F}, & \\ K & \geq & \sum_{A\in\mathcal{A}}\psi_A, & \\ w_{A,F} & \geq & 0, & \forall A\in\mathcal{A},\ \forall F\in\mathcal{F}, \\ \psi_A & \in & \{0,1\}, & \forall A\in\mathcal{A}. \end{array} \end{array} \tag{19}$$

Post-processing can be used in the cases where

$$\{w_{A,south} > 0 \text{ and } w_{A,north} > 0 \ \} \text{ or } \{w_{A,west} > 0 \text{ and } w_{A,east} > 0\},$$

to avoid a treatment plan that calls for two nonzero weights for two diametrically opposite wedge settings as discussed in Appendix B. Note that if there are other constraints on the number of wedges being used, we typically cannot use (19) but must add the additional constraints on $\pi_{A,F}$ to (18).

4 Reducing the Solution Time

The formulation (19) includes beam angles, weights, and wedges as variables in the formulation. It involves a large amount of data—the beam shapes and dose matrices must be computed for each beam angle and wedge orientation—along with many discrete variables, and so is time-consuming to set up and solve. In this section, we describe a number of techniques for reducing the solution time. First, we show how to choose a reasonable value of M in the formulations (16), (18) and (19). (This choice is important in practice, as an excessively large value of M can lead to a significant increase in run time.) Second, we show how normal-tissue voxels in the treatment region some distance away from the target region can be merged, thereby reducing the number of variables without sacrificing solution quality. Third, we describe a scheme for solving a lower-resolution problem to identify the most promising beam angles, then consider only these angles in solving the full-resolution problem.

4.1 Computing tight upper bounds on the beam weights

The formulation (19) requires an upper bound M on the beam weights $w_{A,F}$ which is not known *a priori*. If M is too small, the optimization problem can

be infeasible or produce a suboptimal result. On the other hand, if the value is too large (usually the case), the algorithm can be considerably slower. A key preprocessing technique to overcome this problem is to use tight bounds on the decision variables [21].

Let ρ_A be the maximum dose deliverable to the target by a beam angle A with a unit beam intensity. Since the open beam delivers more radiation to a voxel (per unit beam weight) than any wedged beam, we have

$$\begin{aligned} \rho_A &:= \max_{F\in\mathcal{F},\,(i,j,k)\in\mathcal{T}} \mathcal{D}_{A,F,(i,j,k)} \\ &= \max_{(i,j,k)\in\mathcal{T}} \mathcal{D}_{A,(i,j,k)}, \quad A = 1, 2, \cdots, |\mathcal{A}|, \end{aligned} \tag{20}$$

where, as before, $\mathcal{D}_{A,(i,j,k)}$ denotes the dose delivered to voxel (i,j,k) from a unit weight of the open beam at angle A. In Section 2.2, we defined a constant $\tau_1 \in [0,1]$ as the largest radiation transmission factor by a wedge filter. Using this definition, we have for a given angle A that the maximum dose deliverable to a target voxel using wedge filters is

$$\rho_A \left(w_{A,0} + \tau_1 \sum_{F\in\mathcal{F}\setminus\{0\}} w_{A,F} \right), \tag{21}$$

where $0 \in \mathcal{F}$ denotes the open beam.

Suppose now that we modify the model in (19) to include explicit control of "hot spots" by introducing an upper bound u on the dose allowed in any target voxel. That is, we assume that the constraint

$$D_{\mathcal{T}} \le u e_{\mathcal{T}} \tag{22}$$

is added to (19). (Such a constraint may also be added to the other models of Section 3.) By combining (22) with (21), we deduce that

$$w_{A,0} + \tau_1 \sum_{F\in\mathcal{F}\setminus\{0\}} w_{A,F} \le \frac{u}{\rho_A}, \quad \forall A \in \mathcal{A}.$$

We can therefore use this constraint to bound $w_{A,F}$ for $F \in \mathcal{F}$, provided the angle A is selected. If the angle A is not selected, of course, we must enforce $w_{A,F} = 0$ for all $F \in \mathcal{F}$. We can accomplish these goals by replacing the somewhat arbitrary bound in (19):

$$M\psi_A \ge w_{A,F}$$

by

$$w_{A,0} + \tau_1 \sum_{F\in\mathcal{F}\setminus\{0\}} w_{A,F} \le \left(\frac{u}{\rho_A}\right) \psi_A, \quad \forall A \in \mathcal{A}, \tag{23}$$

where ψ_A is the binary variable that indicates whether or not the angle A is selected. Our modification of (19) that includes "hot spot" control and the

bound (23) is therefore as follows:

$$
\begin{array}{lrll}
\min\limits_{w,\psi} & f(D_\Omega) & & \\
\text{s.t.} & & & \\
& D_\Omega &= \sum\limits_{A\in\mathcal{A},F\in\mathcal{F}} w_{A,F}\mathcal{D}_{A,F,\Omega}, & \Omega \in \mathcal{T}\cup\mathcal{S}\cup\mathcal{N}, \\
& \frac{u}{\rho_A}\psi_A &\geq w_{A,0} + \tau_1 \sum\limits_{F\in\mathcal{F}\setminus 0} w_{A,F} & \\
& K &\geq \sum_{A\in\mathcal{A}}\psi_A, & \\
& w_{A,F} &\geq 0, & \forall A\in\mathcal{A},\, \forall F\in\mathcal{F}, \\
& \psi_A &\in \{0,1\}, & \forall A\in\mathcal{A}\,, \forall F\in\mathcal{F},
\end{array}
\tag{24}
$$

Note that if we also impose an upper bound on dose level to normal-tissue voxels, we can derive additional bounds on the beam weights using the same approach as is used for the target voxels above.

4.2 Reducing resolution in the normal tissue

The main focus in solving the optimization problem is to deliver enough dose to the target while avoiding organs at risk as much as possible. Therefore, the dosage to normal regions that are some distance away from the PTV does not need to be resolved to high precision. It suffices to compute the dose only on a representative subset of these normal-region voxels, and use this subset to enforce constraints and to formulate their contribution to the objective.

Given some parameter Δ, we define a neighborhood of the PTV as follows:

$$\mathcal{R}_\Delta(\mathcal{T}) := \{(i,j,k)\in\mathcal{N} \mid \text{dist}\,((i,j,k),\mathcal{T}) \leq \Delta,\}\,,$$

where dist $((i,j,k),\mathcal{T})$ denotes the Euclidean distance of the center of the voxel (i,j,k) to the target set. We also define a reduced version $\mathcal{N}_1$ of the normal region, consisting only of the voxels (i,j,k) for which i, j, and k are all even; that is:

$$\mathcal{N}_1 := \{(i,j,k)\in\mathcal{N} \mid \mod(i,2) = \mod(j,2) = \mod(k,2) = 0\}\,.$$

Finally, we include in the optimization problem only those voxels that are close to the target, or that lie in an OAR; or that lie in the reduced normal region. Formally, we consider voxels (i,j,k) with

$$(i,j,k)\in\mathcal{T}\cup\mathcal{S}\cup\mathcal{R}_\Delta(\mathcal{T})\cup\mathcal{N}_1.$$

Since each of the voxels $(i,j,k)\in\mathcal{N}_1$ effectively represents seven neighboring voxels, the weights applied to the terms for the voxels $(i,j,k)\in\mathcal{N}_1$ in the L_1 and sum-of-squares objective functions ((4) and (5), respectively) are increased. In effect, the objective quantity $\dfrac{\|D_\mathcal{N}\|_1}{\text{card}\,(\mathcal{N})}$ is smaller than $\dfrac{\|D_{\mathcal{N}_1\cup\mathcal{R}_\Delta(\mathcal{T})}\|_1}{\text{card}\,(\mathcal{N}_1\cup\mathcal{R}\Delta(\mathcal{T}))}$.

If this is an issue, it is possible to replace the latter by

$$\frac{\|\mathcal{D}_{\mathcal{R}_\Delta(\mathcal{T})}\|_1 + \|\mathcal{D}_{\mathcal{N}_1}\|_1 \left(\frac{\text{card}(\mathcal{N}\setminus\mathcal{R}_\Delta(\mathcal{T}))}{\text{card}(\mathcal{N}_1)}\right)}{\text{card}\,(\mathcal{N})}$$

in the objective function.

4.3 A three-phase approach

We now discuss an approach in which rather than attacking the full-scale optimization problem directly, we "ramp up" to the solution via a sequence of models. Each model in the sequence is easier to solve than the next model, and the solution of each provides a good starting point for the next model. The models differ from each other in the selection of voxels included in the formulation, and in the number of beam angles allowed. The idea is to first determine the treatment beam angles among all beam angles considered in the optimization. Then a linear program is solved to determine the intensities of these beam angles. One simple approach for removing unpromising beam angles is to remove from consideration those that pass directly through any sensitive structure [23]. A more elaborate approach [22] introduces a score function for each candidate angle, based on the ability of that angle to deliver a high dose to the target without exceeding the prescribed dose tolerance to OAR or to normal tissue located along its path. Only beam angles with the best scores are included in the model.

These heuristics can reduce solution time appreciably, but their effect on the quality of the final solution cannot be determined *a priori*. We propose instead the following incremental modeling scheme, which obtains a near-optimal solution within a small fraction of the time required to solve the original formulation directly. Our scheme proceeds by three phases.

Phase 1: Selection of Promising Beam Angles. Our aim in this phase is to construct a subset of beam angles A_1 that are likely to appear in the final solution of (19). A similar technique is applicable to (18) and (24). A mixed integer program (19) is solved r times using sampled data points. The sampled data points include voxels on the PTV, randomly sampled 10% of OAR ($\mathcal{S}'$), and $\mathcal{R}_\Delta(\mathcal{T})$, i.e.

$$\Omega_1 = \{\mathcal{T} \cup \mathcal{S}' \cup \mathcal{R}_\Delta(\mathcal{T})\}$$

We define $\mathcal{A}_1$ as a collection of beam angles found in r solutions of the mixed integer program.

Phase 2: Treatment Beam Angle Determination. In the next phase, we determine K or less treatment beam angles based on $\mathcal{A}_1$ (instead of $\mathcal{A}$). We solve the optimization model (19) using $\mathcal{A}_1$ and voxels on the PTV, OAR, and $\mathcal{N}_1$, i.e.

$$\Omega_2 = \{\mathcal{T} \cup \mathcal{S} \cup \mathcal{R}_\Delta(\mathcal{T}) \cup \mathcal{N}_1\}$$

Note that $|\mathcal{A}_1|$ is typically greater or equal than K.

Phase 3: Final Approximation. In the final phase, we fix the K beam angles and solve the resulting simplified optimization problem over the complete set of voxels. We have assumed that the K beam angles are a good approximation to those angles in the solution of the full-scale model. The final approximation typically takes much less time to solve than the full-scale model, both because of the smaller amount of data (due to fewer beam angles) and fewer binary variables.

We have found that this three-phase scheme reduces the total time required to compute the treatment plan considerably. Although it will not in general produce the same solution as the original full-scale model (19), we have found the quality of its approximate solution to be very close to optimal. Computational experience with this approach is given in Section 6.

5 Techniques for DVH control

Dose-volume histograms (DVH) are a compact way of representing dose distribution information for subsets of the treatment region. By placing simple constraints on the shape of the DVH for a particular region, radiation oncologists can exercise control over fundamental aspects of the treatment plan. For instance, the oncologist often is willing to sacrifice some specified portion of a sensitive structure (such as the lung) in order to provide an adequate probability of tumor control, when the sensitive structure lies near the tumor. This aim can be realized by requiring that at least a specified percentage of the sensitive structure must receive a dose less than a specified level. DVH constraints can also be used to control uniformity of the dose to the target, and to avoid cold spots (regions of underdose). For example, the planner may require all voxels in the target volume to receive doses of between 95% and 107% of the prescribed dose (θ).

DVH constraints that require some fraction of voxels in a region to receive less than a given dose, without specifying which individual voxels must satisfy this requirement, cannot be implemented in a straightforward way using traditional optimization formulations. However, by manipulating the objective function, we can set up and solve a sequence of problems that leads to a satisfactory approximate solution. We describe these techniques with reference to the formulation (16). The results are equally valid for (18), but the computational requirements are of course higher.

There are three typical requirements for the radiation treatment: *homogeneity*, *conformity*, and *avoidance* [11, 12]. In our formulations, homogeneity is controlled by the DVH control parameters θ_L and θ_u ($\theta_L \leq 1 \leq \theta_u$), which specify the lower and upper bounds on the dose to target voxels. (If the prescribed dose to the voxels in $\mathcal{T}$ is θ, then we wish to deliver at least $\theta_L \cdot \theta$ and at most $\theta_u \cdot \theta$ to each voxel.) The conformity constraints, which require that the dose to the normal tissue is as small as possible, can be controlled by the penalty parameter on the normal-tissue voxels in the objective function. As we increase the value of λ_n, it typically reduces the integral dose on the normal tissue. The

avoidance constraints, which require the dose to be below certain thresholds on at least some fraction of the sensitive structure, can be implemented by including terms in the objective that involve the OAR voxels and a hot-spot control parameter ϕ.

One might argue that the homogeneity and avoidance requirements can be controlled by adding hard constraints to the optimization model. However, the optimization problem might not be able to find a feasible solution with hard constraints. Even when it is possible to obtain a solution with a hard-constraint formulation, the solutions are typically too homogeneous, and physicians prefer the ability to relax or tighten the constraints using parametrized terms in the objective to achieve a specific treatment goal. We describe ways of controlling DVH on organs, and show results based on a clinical case in the following subsections.

5.1 Effects of different objective functions

We introduced different types of objective functions in Section 3.1; see in particular (4), (5), (6), (14), and (15). One can use infinity-norm penalty terms in the objective function to control hot and cold spots in the treatment region, while L_1-norm penalty terms are useful for controlling the integral dose over a region.

Here we illustrate the effectiveness of using *both* types of terms in the objective, by comparing results obtained from an objective with only L_1 terms, with results for an objective with both L_1 and infinity-norm terms. We use the typical values $\theta_L = 0.95$, $\theta_u = 1.07$, $\phi = 0.2$, and $K = 4$ in this experiment. As can be expected, Figure 6 shows that (15) has better control on the PTV; the infinity-norm terms yielded a stricter enforcement of the constraints on the PTV. The two objective functions can produce a similar solution if the values of λ_t's are chosen appropriately. However, the choice of such values is not intuitive. We believe that it is easier to choose the value of λ_t for the L_∞ penalty, and use these values in the sequel. We note that on the normal and OAR regions, the difference in quality of the solutions obtained from these two alternative objectives was insignificant.

5.2 DVH control on the PTV

Because of our experience reported in Section 5.1, we consider the optimization problem (16) with objective function $f(D_\Omega)$ defined by (15).

Modelers usually are advised to update the weights $(\lambda_t, \lambda_s, \lambda_n)$ to achieve DVH control on the PTV. However, based on extensive numerical experiments, we believe that this is a less effective way to provide DVH control. We suggest fixing $(\lambda_t, \lambda_s, \lambda_n)$ at appropriate value, say 1, and updating them only for fine tuning of a solution.

Our aim in controlling DVH on the PTV is to attain homogeneity of the dose on $\mathcal{T}$ without significant loss of quality in the dose profile for the normal region and OAR (that is, without significant change to the DVH plots for these

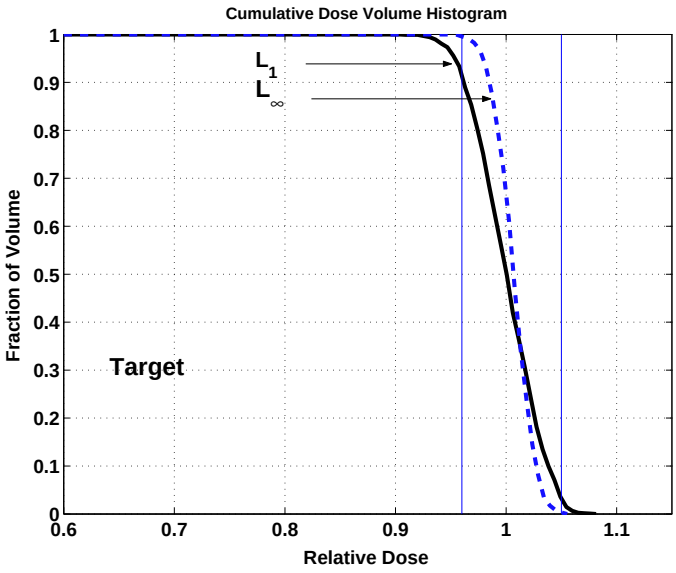
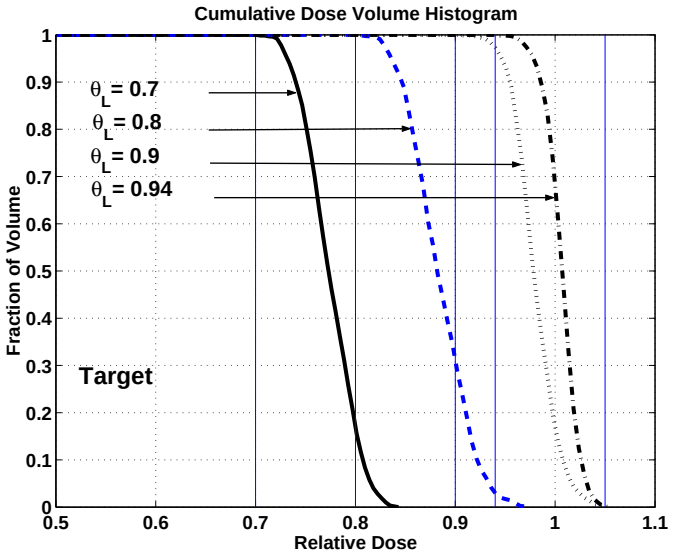


Figure 6: Dose Volume Histogram on the PTV

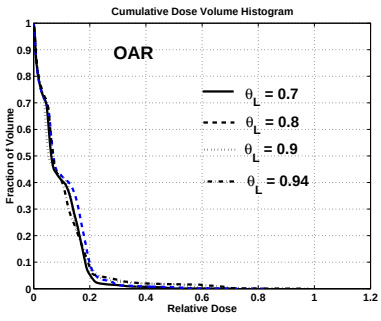
regions). As discussed above, the key parameters used to achieve this goal in (15) are θ_u and θ_L, which define the desired maximum and minimum fractions of the prescribed dose that the planner wishes to deliver to the target voxels. In this experiment, we fix $\theta_u = 1.07$, and try the values 0.7, 0.8, 0.9, 0.94 for the lower-bound fraction θ_L. Figure 7 shows four DVH plots based on the four different values of θ_L. For each value, we observe that in fact 100% of the target volume receives more that the desired lower bound θ_L. In other words, we manage to avoid completely cold spots in the PTV in this example. We may expect that larger values of θ_L (which lead us to confine the target dose to a tighter range) will result in a less attractive solution in the OAR and the normal tissue. However, as can be seen in Figure 7, the loss of treatment quality is not significant. We conclude that this technique for implementing homogeneity constraints is effective.

5.3 DVH control on the OAR

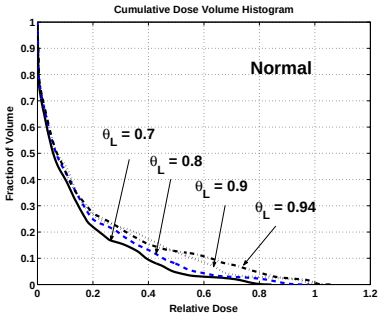
The objective (15) also contains terms that penalize the integral of the dose violation over the OAR and normal regions. Here, we show that the dose to the OAR can be controlled by means of the parameter ϕ, assuming that the weights λ_t, λ_s, and λ_n have been fixed appropriately. If our goal is for voxels in the OAR to receive a dose of at most β, where $\beta \in (0, 1)$, we set $\phi = \beta$ in (15). Figure 8(a) illustrates the effect of changing values of ϕ on DVH of the OAR. When ϕ is set to 0.5, most of the OAR receives dose less than 50% of the



(a) PTV



(b) OAR



(c) Normal

Figure 7: DVH plots: DVH control on the PTV

prescribed target dose. Similar results hold for the values 0.2 and 0.1, though constraint is not as "hard" in these cases. (For $\phi = 0.1$, about 20% of the OAR receives more than 10% of the prescribed dose, but only about 5% receives more than 20% of the prescribed dose.) As expected, the costs of achieving better control on the OAR is the loss of treatment quality on the PTV and the normal tissue. However, Figure 8 shows that there is little sacrifice in treatment quality.

5.4 Remarks

We conclude this section with several remarks.

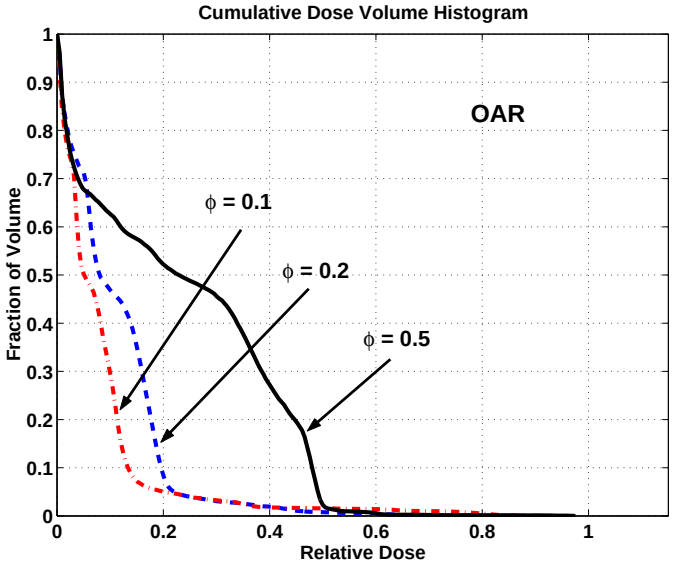
1. If our goal is to control hot spots in the OAR rather than the integral dose, we could replace the term $\|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_1$ in the objective (15) by its infinity-norm analogue $\|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_\infty$.

2. In applying the three-phase approach of Section 4.3 to the objective function (15), we can update ϕ on a per-organ basis and re-solve the optimization problem if the DVH requirement for the OAR is not satisfied at the end of Phase 3.

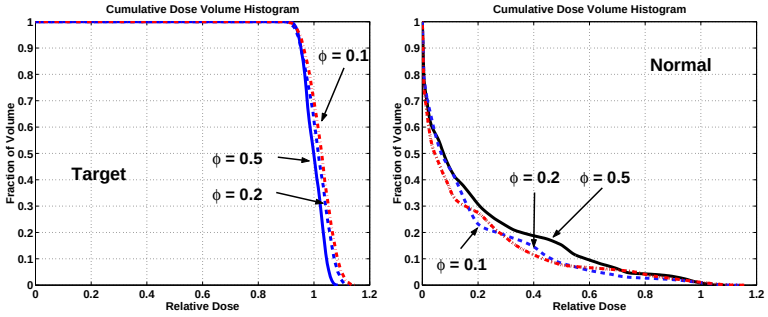
3. There can be some conflict between the goals of controlling DVH on target and non-target regions. Ideally, all target voxels should receive the exact prescription dose θ, while the non-target region should receive zero dose. In practice, this is not possible, as the target is always adjacent either to normal tissue or sensitive structures. Therefore, we need to reach a compromise based on the the relative priorities of meeting the prescription on the target and avoiding excessive dose to the OAR and normal tissues. If the PTV dose control is most important, as is usually the case, the control parameters θ_L, θ_u, ϕ should be chosen with $(\theta_u - \theta_L)$ small and ϕ as a fairly large (but smaller than 1) fraction of θ. However, if the OAR dose control is most important, a smaller value of ϕ can be used in conjunction with L_1-norm penalties for the OAR terms in the objective. In addition, a larger value of $(\theta_u - \theta_L)$ is appropriate in this case.

6 Application to Clinical Data

In this section, we use two sets of clinical data to explain how to use our model to achieve treatment planning goals.



(a) OAR



(b) PTV

(c) Normal

Figure 8: DVH plots: DVH control on the OAR

6.1 Solution time reduction

The specific optimization model considered in this section is as follows:

$$
\begin{aligned}
\min_{w,\psi} \quad & \lambda_t \left(\|(D_{\mathcal{T}} - \theta_u e_{\mathcal{T}})_+\|_\infty + \|(\theta_L e_{\mathcal{T}} - D_{\mathcal{T}})_+\|_\infty \right) \\
& + \lambda_s \frac{\|(D_{\mathcal{S}} - \phi e_{\mathcal{S}})_+\|_1}{\text{card}\,(\mathcal{S})} + \lambda_n \frac{\|D_{\mathcal{N}}\|_1}{\text{card}\,(\mathcal{N})} \\
\text{s.t.} \quad & \\
& D_\Omega = \sum_{A \in \mathcal{A}} \mathcal{D}_{A,\Omega} \cdot w_A, \qquad \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\
& D_{\mathcal{T}} \leq u, \\
& 0 \leq w_A \leq M \psi_A, \qquad \forall A \in \mathcal{A}, \\
& \sum_{A \in \mathcal{A}} \psi_A \leq K, \\
& \psi_A \in \{0,1\}, \qquad \forall A \in \mathcal{A}.
\end{aligned}
\tag{25}
$$

Note that we have introduced hard upper bound constraints on the target $D_{\mathcal{T}} \leq u$ (where u typically is somewhat larger than θ_u). We fix some of the control parameters in the optimization model (25) throughout the experiments: $\theta_L = 0.95$, $\theta_u = 1.07$, $\phi = 0.2$, $K = 4$, $\lambda_t = \lambda_s = \lambda_n = 1$, $u = 1.15$ and $|\mathcal{A}| = 36$. In fact, the set of angles $\mathcal{A}$ consists of angles equally spaced by 10° in a full 360° circumference.

We attempt to solve (25) using the full set of voxels. Note that the optimality criterion is set such that the solution process terminates with the relative error of the objective value being less than or equal to 1%. Figure 9 shows changes of upper and lower bounds of the objective values as the iteration number increases. We notice that a large number of iterations are used to slightly improve the feasible solution found at iteration 2.2×10^6. We also notice that the lower bound of the objective value increases slowly. We addressed techniques to overcome these problems in Section 4. Effects of using the techniques are discussed in the following paragraph.

Table 1 summarizes results of four different experiments using a data set from a patient with pancreatic cancer. Column I shows the results obtained by solving (25) directly, with M set to 2. In column II, we use the tight bound (23) on w_A, specialized to the case in which no wedges are used. That is, we replace the constraint $w_A \leq M\psi_A$ in (25) by $w_A \leq (u/\rho_A)\psi_A$. (This tighter bound is also used in columns III and IV.) Column III shows the solution time for the reduced-voxel version of the problem discussed in Section 4.2. Finally, column IV shows results obtained with the three-phase approach of Section 4.3 using $r = 10$ samples of the OAR. Note that the objective values were calculated on the full set of voxels for the comparison.

Table 1 shows that the final objective values obtained from the all three schemes were the same, to at least three significant digits. The next rows in Table 1 show the CPU times required (in hours) for each of the four experiments, and the savings in comparison with the time in column I. By comparing columns I and II, we see that a modest reduction was obtained by using the tighter bound.

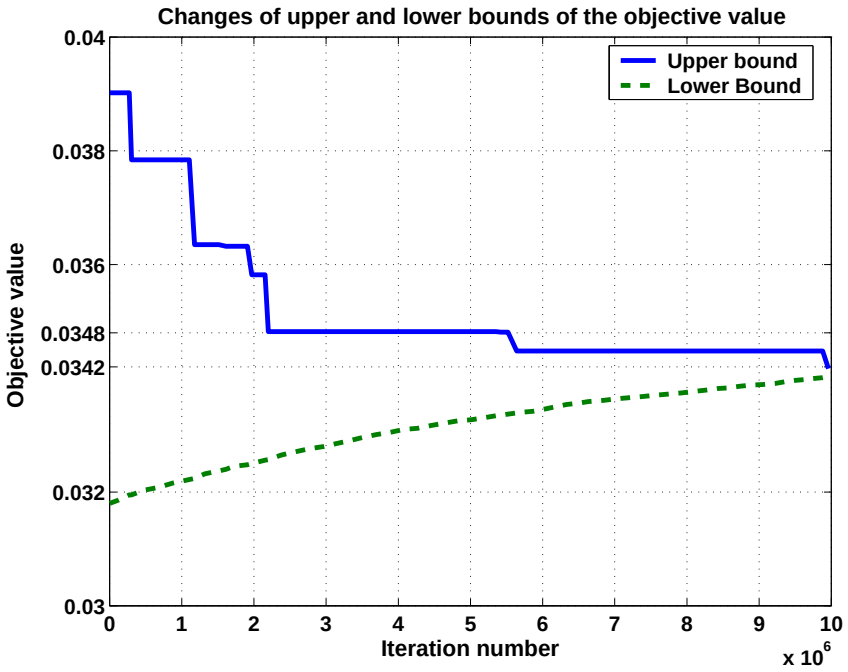


Figure 9: Changes of upper and lower bounds of the objective value

Table 1: Comparisons among different solution schemes.

	I	II	III	IV
Approach	Single Solve	Single Solve	Reduced Model	Three-Phase
Bound (M)	2	u/ρ_A	u/ρ_A	u/ρ_A
Final Objective	0.0342	0.0342	0.0342	0.0342
Time (hours)	112.3	93.5	29.9	0.5
Time saved(%)	-	16.8	73.3	99.5

Column III shows that more significant savings were obtained, with essentially no degradation in the quality of the solution plan, by using a reduced model. The full problem contains 1244 voxels in the PTV, 69270 voxels in the OAR, and 747667 voxels in the normal region, while the reduced model has 1244 voxels in the PTV, 14973 voxels in the OAR, and 96154 voxels in the normal tissue. The reduction in computing time was over 73%. Column IV shows that the use of the three-phase scheme resulted in a savings of 99.5% over the direct solution scheme with no effect on the quality of the solution.

Note that, if the solution time is very important, we could relax the cold-spot and hot-spot control parameter values on the PTV. Relaxing these parameter values typically speeds up the the solution time.

We believe our iterative technique is equally effective in the general case in which wedges are included in the formulation. Hence, our subsequent computations used the iterative scheme with wedges.

6.2 The effect of using wedges on DVH

In general, the use of wedges gives more flexibility in achieving adequate coverage of the tumor volume while sparing normal tissues. To show the effect of wedges, we test our optimization models on a different set of data, from a prostate cancer patient. Figure 10 shows DVH graphs obtained for a treatment plan using wedges (24) and one using no wedges (25). Three beam angles, $K = 3$, are used in both cases. As can be seen in Figure 10(a), a significant improvement on DVH on the OAR is achieved by adding wedges. In Figure 10(b), we see that there is also a slight improvement in the DVH for the PTV. The line is closer to the prescribed dose level of one when wedges are used. The DVH on the normal tissue, however, does not show much difference between the wedges and no-wedges cases.

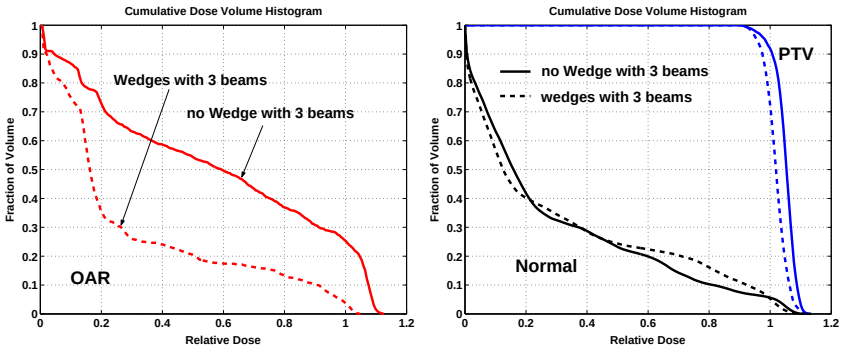
6.3 A Clinical case study - Pancreas

We now apply the full optimization approach (including DVH controls and wedges) to a pancreatic tumor. This case is made particularly difficult by the close proximity to the PTV of several sensitive structures, including the spinal cord, liver, left kidney, and right kidney. The set $\mathcal{A}$ contains 36 equispaced candidate beam angles. Wedges are also used for the beam angles. The goals of the treatment plan are as follows:

1. Choose four beam angles for the treatment.
2. As the first priority, the target volume should receive dose between 95% and 107% of the presribed dose.
3. 90% of each organ-at-risk should receive less than 20% of the target prescribed dose level.
4. The integral dose delivered to the normal tissue should be minimized.

To achieve these goals, we set DVH control parameters as follows:

$$\theta = 1.0,\ \theta_L = 0.95,\ \theta_u = 1.07,\ r = 10,\ \ K = 4,\ \text{and}$$



(a) Organ at Risk (b) Target & Normal

Figure 10: Dose Volume Histogram: effect of wedges with 3 beam angles

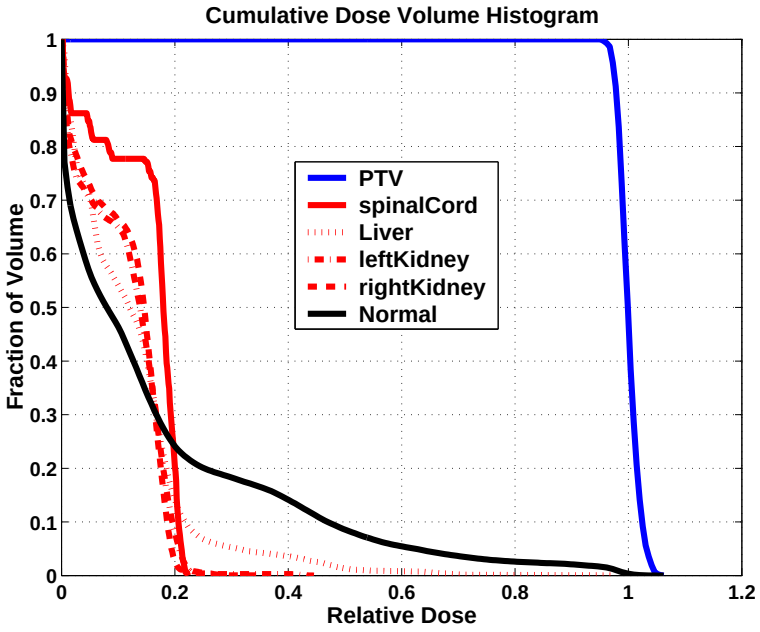
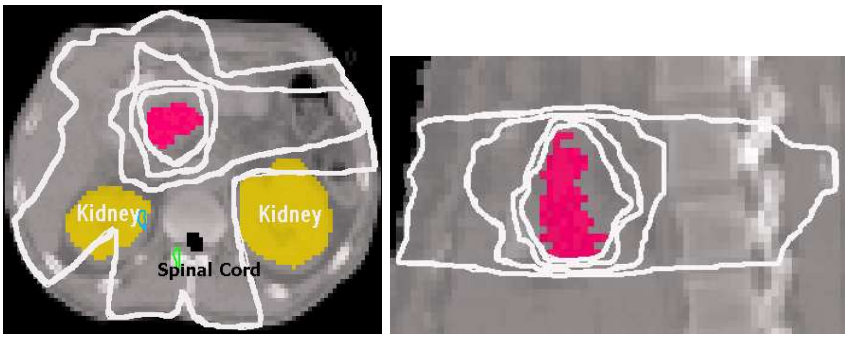


Figure 11: Dose Volume Histogram: optimal solution



(a) Axial (b) Sagittal

Figure 12: Isodose Plots: four lines represent 20%, 50%, 80%, and 95% isodose lines; the 20% line is outermost.

$$\phi_i = 0.2,\ i \in \{\text{spinal cord, liver, left kidney, right kidney}\}.$$

Figure 11 shows DVH plots of this experiment. Note first that the homogeneity constraints are satisfied for the PTV: every voxel in the target volume receives 95% and 107% of the prescribed dose. It is also clear that approximately 90% of each sensitive structure receives at most 20% of the target prescribed dose, as specified; the DVH plot for each sensitive structure passes very close to the point $(0.2, 0.1)$ that corresponds to the aforementioned treatment goal.

Figure 12 shows isodose lines on the CT slices. The target (tumor) is outlined within four isodose lines. The outermost line is 20% isodose line, which encloses a region in which the voxels receive a dose of 20% of the target prescribed dose level. Moving inwards towards the target, we see 50%, 80%, and 95% isodose lines. Figure 12(a) shows an axial slice. The kidneys are outlined as two circles right below the target. As can be seen, the target lies well inside the 95% isodose line, while the dose to the organs at risk remains reasonable. Figure 12(b) shows a sagittal view of the target with those four isodose lines also.

All computations in this paper were performed on Pentium 4 1.8 GHz machine running on Linux. All optimization problems were modeled in the GAMS modeling language [4]. We use CPLEX 7.1 as LP and MIP solver, and MINOS 5.5 for QP solver.

7 Summary

We have developed an optimization framework for 3D conformal radiotherapy. The key features of our methodology are as follows:

1. Simultaneous optimization of three key parameters (beam angles, wedge

orientation, and beam weight);
2. Fast delivery of the treatment plan; and
3. Capability of controlling DVH on organs implicitly depending on the specific treatment goal of the planner.

The optimization problems were formulated as mixed integer linear programming and quadratic programming problems. We presented different objective function formulations for different treatment goals. Since the data set required by the obvious optimization formulations was very large, techniques were introduced to reduce the data requirements and the complexity of the problem. Specifically, we introduced tighter *a priori* bounds on the beam weights, reduction of the number of voxels to be considered in the optimization, and a three-phase scheme in which a sequence of progressively more realistic optimization models is solved to obtain an approximate solution. Using all these techniques, we demonstrated a 99.5% improvement in computational time over direct solution of the full-resolution problem on a clinical data set.

A Using a Universal Wedge

In this section we show that a treatment plan that requires the use of a wedge at a certain orientation can be expressed equivalently by a plan that uses a different wedge at the same orientation, provided that certain conditions are satisfied by the two wedges. This result implies that a single "universal wedge" can be used to design and deliver treatment plans; an array of wedges with different properties (i.e., different values of τ_0 and τ_1) is unnecessary.

Suppose that at some angle A and some wedge at a given orientation with parameters τ_0' and τ_1' (with $0 \leq \tau_0' < \tau_1' \leq 1$) we have a treatment plan that calls for delivering a weight $w'_{A,\text{open}}$ through the open beam, and $w'_{A,\text{west}}$ through the wedge. (We have supposed without loss of generality that the wedge is oriented to the west, so the attenuation parameter τ_{ij} for beamlet (i,j) is given by the formula (1).) We now ask whether it is possible to deliver an equivalent dose through every beamlet using a different wedge with the same (west) orientation, and different parameters τ_0 and τ_1, with $0 \leq \tau_0 < \tau_1 \leq 1$.

Using (1), we find that the total dose delivered through beamlet (i,j) is

$$\begin{aligned} & w'_{A,\text{open}} + w'_{A,\text{west}} \left[\tau_0' + \frac{j-0.5}{N}(\tau_1' - \tau_0') \right] \\ = \quad & w'_{A,\text{open}} + w'_{A,\text{west}} \left[\tau_0' - 0.5/N(\tau_1' - \tau_0') \right] + j w'_{A,\text{west}}(\tau_1' - \tau_0')/N. \end{aligned}$$

If we were to use the alternative wedge with parameters τ_0 and τ_1, and weights $w_{A,\text{open}}$ and $w_{A,\text{west}}$, we would find that the total dose delivered through beamlet (i,j) is

$$w_{A,\text{open}} + w_{A,\text{west}} \left[\tau_0 - 0.5/N(\tau_1 - \tau_0) \right] + j w_{A,\text{west}}(\tau_1 - \tau_0)/N.$$

By equating the constant terms and the coefficient of j in the last two formulae, we find that the plans are equivalent if

$$w_{A,\text{west}}(\tau_1 - \tau_0) = w'_{A,\text{west}}(\tau'_1 - \tau'_0)$$

and

$$w_{A,\text{open}} + w_{A,\text{west}}\left[\tau_0 - 0.5/N(\tau_1 - \tau_0)\right] = w'_{A,\text{open}} + w'_{A,\text{west}}\left[\tau'_0 - 0.5/N(\tau'_1 - \tau'_0)\right].$$

By rearranging and substituting, we find that the weights for the new beam must be

$$w_{A,\text{west}} = \frac{\tau'_1 - \tau'_0}{\tau_1 - \tau_0} w'_{A,\text{west}}$$

and

$$w_{A,\text{open}} = w'_{A,\text{open}} + w'_{A,\text{west}}\left[\tau'_0 - \frac{\tau'_1 - \tau'_0}{\tau_1 - \tau_0}\tau_0\right]. \tag{26}$$

Note that $w_{A,\text{west}}$ is always nonnegative whenever $w'_{A,\text{west}}$ is nonnegative, but that $w_{A,\text{open}}$ is not necessarily nonnegative, even when the weights for the original wedge are both nonnegative. However, a sufficient condition for $w_{A,\text{open}}$ to be nonnegative for *any* nonnegative values of $w'_{A,\text{open}}$ and $w'_{A,\text{west}}$ is that

$$\frac{\tau'_0}{\tau_0} \geq \frac{\tau'_1 - \tau'_0}{\tau_1 - \tau_0},$$

since this condition ensures that the bracketed term on the right-hand side of (26) is nonnegative. This condition implies that given a solution using a particular wedge, we can always identify an equivalent plan using an alternative wedge with the same (or smaller) value of τ_0 and a larger value of $\tau_1 - \tau_0$.

B Transforming the Wedge Orientation Solution

As mentioned earlier, we would prefer not to deliver two beams from the same angle for two diametrically opposite wedge settings. There is nothing in the formulations of Section 3.3 to prevent such solutions arising. However, a simple transformation of the solution results in an equivalent treatment plan in which no such beam pairs are present.

To illustrate the technique, consider the "west" and "east" wedge orientations. For beamlet (i, j), $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$, the attenuation factor when the west wedge is present is given by (1). For the east wedge, it is as follows.

$$\tau_{ij}^{\text{west}} = \tau_0 + \frac{N - j + 0.5}{N}(\tau_1 - \tau_0), \quad i = 1, 2, \ldots, M, \;\; j = 1, 2, \ldots, N. \tag{27}$$

Suppose now that we have a treatment plan in which, at some angle A, the weight corresponding to the open beam (no wedge) is $w_{A,\text{open}} \geq 0$, while the weights corresponding to the west and east beams are $w_{A,\text{west}} > 0$ and

$w_{A,\text{east}} > 0$, respectively. Suppose for the moment that $w_{A,\text{west}} \geq w_{A,\text{east}}$. The contribution of these three weights to the total intensity delivered by beamlet (i, j) is then

$$w_{A,\text{east}} \left[\tau_0 + \frac{N - j + 0.5}{N}(\tau_1 - \tau_0)\right] + w_{A,\text{west}} \left[\tau_0 + \frac{j - 0.5}{N}(\tau_1 - \tau_0)\right] + w_{A,\text{open}}$$

which is equal to

$$(w_{A,\text{west}} - w_{A,\text{east}}) \left[\tau_0 + \frac{j - 0.5}{N}(\tau_1 - \tau_0)\right] + (w_{A,\text{open}} + w_{A,\text{east}}(\tau_1 - \tau_0)).$$

Hence, the same beamlet intensity could be delivered at every (i, j) pair by using weight $w_{A,\text{open}} + w_{A,\text{east}}(\tau_1 - \tau_0)$ for the open beam, $(w_{A,\text{west}} - w_{A,\text{east}})$ for the west wedge, and 0 for the east wedge. Similarly, for the case of $w_{A,\text{west}} \leq w_{A,\text{east}}$, we achieve identical beamlet intensities by using weight $w_{A,\text{open}} + w_{A,\text{west}}(\tau_1 - \tau_0)$ for the open beam, 0 for the west wedge, and $(w_{A,\text{east}} - w_{A,\text{west}})$ for the east wedge.

We conclude that when there are positive beam weights for two diametrically opposed wedge orientations, we can obtain an equivalent treatment plan by zeroing the smaller of the two weights, and adjusting the weights on the open beam and on the remaining wedge orientation.

References

[1] T. Bortfeld and W. Schlegel. Optimization of beam orientations in radiation-therapy - some theoretical considerations. *Physics in Medicine and Biology*, 38(2):291–304, 1993.

[2] Th. Bortfeld, J. Burkelbach, R. Boesecke, and W. Schlegel. Methods of image reconstruction from projections applied to conformation radiotherapy. *Physics in Medicine and Biology*, 25(10):1423–1434, 1990.

[3] Thomas Bortfeld, Arthur L. Boyer, Wolfgang Schlegel, Darren L. Kahler, and Timothy J. Waldron. Realization and verification of three-dimensional conformal radiotherapy with modulated fields. *International journal of radiation oncology: Biology, Physics*, 30(4):899–908, 1994.

[4] A. Brooke, D. Kenderick, A. Meeraus, and R. Raman. *GAMS: User's Guide*. GAMS Development Corporation: http://www.gams.com/docs/, 1998.

[5] G.T.Y. Chen, D. R. Spelbring, C.A. Pelizzari, J.M. Balter, L. C. Myrianthopoulous, S. Vijayakumar, and H. Halpern. The use of beam eye view volumetrics in the selection of noncoplanar radiation portals. *International Journal of Radiation Oncology: Biology, Physics*, 23:153–163, 1992.

[6] Yan Chen, Darek Michalski, Christopher Houser, and James M Galvin. A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy. *Physics in Medicine and Biology.*, 47:1647–1658, 2002.

[7] B. C. J. Cho, W. H. Roa, D. Robinson, and B. Murray. The development of target-eye-view maps for selection of coplanar or noncoplanr beams in conformal radiotherapy treatment planning. *Medical Physics*, 26(11):2367–2372, 1999.

[8] Steven M. Crooks, Andrei Pugachev, Christopher King, and Lei Xing. Examination of the effect of increasing the number of radiation beams on a radiation treatment plan. *Physics in Medicine and Biology*, 47:3485–3501, 2002.

[9] Jianrong Dai and Yuping Zhu. Selecting beam weight and wedge filter on the basis of dose gradient analysis. *Medical Physics*, 27:1746–1752, 2001.

[10] Jianrong Dai, Yuping Zhu, and Qing Ji. Optimizing beam weights and wedge filters with the concept of the super-omni wedge. *Medical Physics*, 27(12):2757–2762, 2000.

[11] Michael C. Ferris, Jinho Lim, and David M. Shepard. An optimization approach for radiosurgery treatment planning. *SIAM Journal On Optimization, Forthcoming*, 2002.

[12] Michael C. Ferris, Jinho Lim, and David M. Shepard. Radiosurgery optimization via nonlinear programming. *Annals of Operations Research, Forthcoming*, 2002.

[13] Michael C. Ferris, Robert R. Meyer, and Warren D'Souza. Radiation treatment planning: Mixed integer programming formulations and approaches. *Optimization Technical Report 02-08, University of Wisconsin*, 2002.

[14] Michael Goitein, Mark Abrams, Serek Rowell, Helen Pollari, and Judy Wiles. Multi-dimensional treatment planning: Ii. beam's eye-view, back projection, and projection through ct sections. *International Journal of Radiation Oncology: Biology, Physics*, 9:789–797, 1983.

[15] P Gokhale, E.M.A. Hussein, and N Kulkarni. Determination of beam orientation in radiotherapy planning. *Medical Physics*, 21(3):393–400, 1994.

[16] OCL Hass, K. J. Burnham, and J. A. Mills. Optimization of beam orientation in radiotherapy using planar geometry. *Physics in Medicine and Biology*, 43(8):2179–2193, 1998.

[17] Intensity Modulated Radiation Therapy Collaborative Working Group. Intensity-modulated radiotherapy: Current status and issues of interest. *International Journal of Radiation Oncology: Biology, Physics*, 51(4):880–914, 2001.

[18] Thomas J. Jordan and Peter C. Williams. The design and performance characteristics of a multileaf collimator. *Physics in Medicine and Biology.*, 39:231–251, 1994.

[19] Jonathan G. Li, Artheru L. Boyer, and Lei Xing. Clinical implementation of wedge filter optimization in three-dimensional radiotherapy treatment planning. *Radiotherapy and Oncology*, 53:257–264, 1999.

[20] L.C. Myrianthopoulos, G.T.Y. Chen, S. Vijayakumar, H. Halpern, D. R. Spelbring, and C.A. Pelizzari. Beams eye view volumetrics - an aid in rapid treatment plan development and evaluation. *International Journal of Radiation Oncology: Biology, Physics*, 23(367-375), 1992.

[21] George L. Nemhauser and Laurence A. Wolsey. *Integer and Combinatorial Optimization*. John Wiley & Sons, 1988.

[22] Andrei Pugachev and Lei Xing. Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *International Journal of Radiation Oncology: Biology, Physics*, 51(5):1361–1370, 2001.

[23] C.G. Rowbottom, V.S. Khoo, and S. Webb. Simultaneous optimization of beam orientations and beam weights in conformal radiotherapy. *Medical Physics*, 28(8):1696–1702, 2001.

[24] C.G. Rowbottom, S. Webb, and M. Oldham. Improvements in prostate radiotherapy from the customization of beam directions. *Medical Physics*, 25(1171-1179), 1998.

[25] C.G. Rowbottom, S. Webb, and M. Oldham. Beam-orientation customization using an artificial neural network. *Physics in Medicine and Biology*, 44(2251-2262), 1999.

[26] S. Shalev, D. Viggars, M. Carey, and P. Hahn. The objective evaluation of alternative treatment plans 2 score functions. *International Journal of Radiation Oncology: Biology, Physics*, 20(5):1067–1073, 1991.

[27] George W. Sherouse. A mathematical basis for selection of wedge angle and orientation. *Medical Physics*, 20(4):1211–1218, 1993.

[28] S. Soderstrom, A. Gustafsson, and A. Brahme. Few-field radiation-therapy optimization in the phase-space of complication-free tumor central. *International Journal of imaging systems and technology*, 6(1):91–103, 1995.

[29] J. Tervo and P. Kolmonen. A model for the control of a multileaf collimator in radiation therapy treatment planning. *Inverse Problems*, 16:1875–1895, 2000.

[30] S. Webb. Optimization of conformal radiotherapy dose distributions by simulated annealing. *Physics in Medicine and Biology*, 34:1349–1369, 1989.

[31] S. Webb. Optimization by simulated annealing of three-dimensional, conformal treatment planning for radiation fields defined by a multileaf collimator. *Physics in Medicine and Biology*, 36(9):1201–1226, 1991.

[32] S. Webb. Optimization by simulated annealing of three-dimensional, conformal treatment planning for radiation fields defined by a multileaf collimator: Ii. inclusion of the two-dimensional modulation of the x-ray intensity. *Physics in Medicine and Biology*, 37(8):1992, 1992.

[33] S. Webb. *The Physics of Conformal Radiotherapy.* Institute of Physics Publishing, 1997.

[34] S. Webb. Configuration options for intensity-modulated radiation therapy using multiple static fields shaped by a multileaf collimator. *Physics in Medicine and Biology.*, 43:241–260, 1998.

[35] Xingen Wu and Yuping Zhu. A global optimization method for three-dimensional conformal radiotherapy treatment planning. *Physics in Medicine and Biology*, 46:109–119, 2001.

[36] Lei Xing, R.J. Hamilton, C. Pelizzari, and G.T.Y. Chen. A three-dimensional algorithm for optimizing beam weights and wedge filters. *Medical Physics*, 25(10):1858–1865, 1998.

[37] Lei Xing, C. Pelizzari, F.T. Kuchnir, and G.T.Y. Chen. Optimization of relative weights and wedge angles in treatment planning. *Medical Physics*, 24(2):215–221, 1997.