

An Optimization Framework for Conformal Radiation Treatment Planning

Gino J. Lim

Department of Industrial Engineering, University of Houston, Houston, Texas 77204, ginolim@uh.edu

Michael C. Ferris, Stephen J. Wright

Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706
{ferris@cs.wisc.edu, swright@cs.wisc.edu}

David M. Shepard

Swedish Cancer Institute, Seattle, Washington 98104, david.shepard@swedish.org

Matthew A. Earl

Department of Radiation Oncology, University of Maryland School of Medicine,
Baltimore, Maryland 21201, mearl@umm.edu

An optimization framework for three-dimensional conformal radiation therapy is presented. In conformal 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. Wedge filters may be used to produce a gradient in beam intensity across the aperture. 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, the weight (exposure time) to be used for each beam, and the type and orientation of wedges to be used. Practical solution techniques for this problem are described; they include strengthening of the formulation and solution of smaller approximate problems obtained by a reduced parametrization of the treatment region. In addition, techniques for controlling the dose-volume histogram implicitly for various parts of the treatment region using hot- and cold-spot control parameters are presented. Computational results are given that show the effectiveness of the proposed approach on practical data sets.

Key words: optimization; treatment planning; conformal; sampling

History: Accepted by John W. Chinneck, Area Editor for Modeling: Methods and Analysis; received January 2003; revised March 2004, April 2005, December 2005; accepted January 2006. Published online in *Articles in Advance* July 20, 2007.

1. Introduction

Optimization of radiation therapy for cancer treatment has become an active research topic in recent years (Bortfeld et al. 1990, Bortfeld and Schlegel 1993, Bortfeld et al. 1994, Chen et al. 2002, Intensity Modulated Radiation Therapy Collaborative Working Group 2001, Jordan and Williams 1994, Tervo and Kolmonen 2000, Webb 1998, Wu and Zhu 2001, Xiao et al. 2003). Several survey articles cover the essential elements of the problem; see Holder (2004), Lodwick et al. (1998), Rosen et al. (1990), and Shepard et al. (1999). Many types of cancer are treated by aiming beams of ionizing radiation at the patient from a number of different angles. The fundamental goal is to apply a significant total dose of radiation to the cancerous region (the tumor) while sparing from excessive radiation the surrounding normal tissues (especially sensitive structures near the tumor).

External-beam radiation treatments are typically delivered using a linear accelerator with a multileaf

collimator (see Figure 2) 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 beam is shaped at each angle to match the shape of the tumor, as viewed from that angle. We refer to this method of beam-shape selection as the *beam's-eye view (BEV)* technique. Wedges can be placed in front of the beams to induce a gradient across the radiation field to help treat 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 used in the treatment of most patients, until recently automated treatment planning systems have typically used naïve optimization procedures. We focus on the conformal approach because it requires little alteration to current clinical practices, and therefore has a high probability of rapid adoption. A more sophisticated treatment planning approach known as intensity modulated radiation therapy (IMRT) 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. The approaches outlined in this paper are also pertinent to IMRT, but will benefit from the reduced number of aperture shapes that are generated, for example, in the direct aperture optimization (DAO) approach (Earl et al. 2003, Shepard et al. 2002).

Our process consists of the following steps: 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 on organs. Data for the model consist of the dose-distribution matrices for beams from each angle, along with 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. Calculation of dose matrices is described in Section 2.1. 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 a “sensitive structure” or a “critical structure”), and specifies the desired or required dose for each region. In Section 2.2 we describe the attenuation of radiation dose caused by wedges.

In Section 3, we present several formulations of the treatment planning problem using linear programming (LP) and mixed-integer linear programming (MIP) approaches. In these optimization models, a lower bound is typically specified for the dose to each voxel in the target (PTV), while an upper bound is used for voxels in the OAR and normal tissue. Since sensitive structures often are located close to the PTV, it is often difficult or impossible to determine a treatment plan that satisfies the required bounds at every voxel. Accordingly, these “hard” bounds are often replaced by penalty terms in the objective, or by constraints that limit the fraction of the sensitive structure

that receives an unacceptably high dose. These important formulation issues are discussed in Ehr Gott and Burjony (2001), Holder (2001, 2003), and Sonderman and Abrahamson (1985).

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. In Section 3.2, we discuss the “angle-selection” problem, in which the most effective angles (and their 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 is used in the treatment. Treatments with fewer beams can be delivered more rapidly, and hence are generally preferred. We consider treatment plans involving wedges in Section 3.3, using an extension of the MIP formulation for the angle-selection problem.

The quality of a treatment plan is typically specified and evaluated using a dose-volume histogram (DVH), a plot that shows what fraction of volume of a structure receives dosage in a given range. Optimization techniques utilizing dose-volume restrictions (often known as *DVH constraints*) are described in Langer et al. (1990), Leong and Langer (1987), and Morrill et al. (1990). The use of binary variables to enforce such constraints has been investigated, e.g., in Ferris et al. (2005) and Lee et al. (2000, 2003). In Section 4, we demonstrate various techniques for DVH control including techniques that depend on penalization rather than on additional integer variables in the formulation. We also demonstrate the usefulness of wedges in DVH control. In Section 5, we describe several techniques for improving the formulation and reducing the solution time without degrading the solution quality for this model. In Section 6 we present computational results for a complete data set, showing the effectiveness of our three-phase approach and the quality of the generated solutions.

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 (Goitein et al. 1983, Webb 1997). For purposes of dose-matrix calculation, the full aperture is divided into rectangular subfields arranged in a regular $M \times N$ pattern (see Figure 1). Each subfield is called a *pencil beam* or *beamlet*, and is identified by an index pair (i, j) , where $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$. In our examples, the full aperture is square with sides of length 10 cm, while $M = 10$ and $N = 20$. Therefore, each of the 200 pencil-beam apertures has dimension 1 cm \times 0.5 cm.

The dose-distribution matrix for each pencil beam from each angle is calculated via a Monte Carlo technique that simulates the track of a large number of

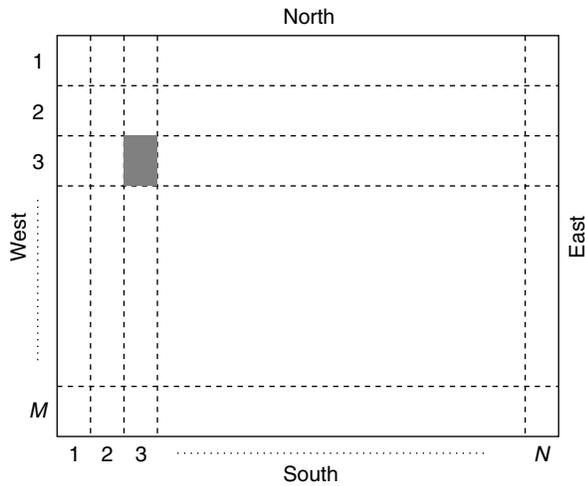


Figure 1 Division of Aperture into Pencil Beams (Shaded Area Represents One Beamlet)

individual radiation particles. A unit-intensity, non-wedged beam is assumed for the purpose of these calculations.

In conformal radiotherapy, the shape of the beam from each angle corresponds to the BEV of the tumor. We define the BEV to contain all beamlets for which at least one voxel in the PTV receives a dose of at least $T\%$ of the maximum dose delivered by that pencil beam to *any* voxel, where T is a threshold parameter. The choice of T is critical. If too small, the BEV overestimates the PTV, producing an aperture that irradiates not only the target but also nearby normal tissue and organs at risk. If too large, the BEV underestimates the PTV, and the optimizer might not be able to find a solution that delivers adequate radiation to all of the PTV. The best value of T to use depends somewhat on the shape of the tumor. We choose T to be the minimum value such that the resulting BEVs provide a complete two-dimensional coverage of the PTV from all beam angles considered in the problem. Values in the range 10%–15% appear to be appropriate. A BEV is shown in Figure 2.

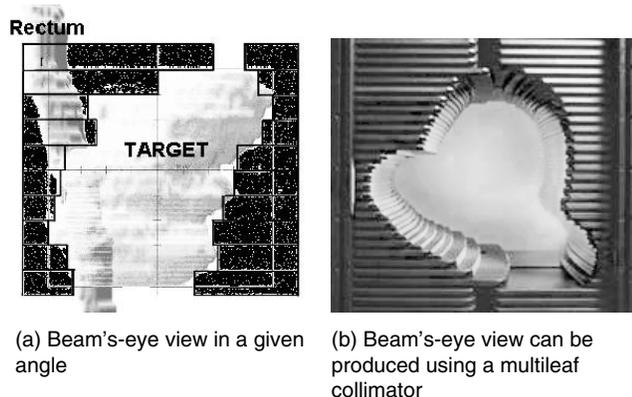


Figure 2 Beam's-Eye View

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

2.2. Wedges

As shown in Figure 3, a wedge (also called a “wedge filter”) is a tapered metallic block with a thick side (the *heel*) and a thin edge (the *toe*). When the wedge is placed in front of the aperture, less radiation is transmitted through the heel of the wedge than through the toe. Figure 3 also shows an external 45° wedge, so named because it produces isodose lines that are oriented at approximately 45°. We assume that the wedge can be oriented in four ways: with its heel aligned with each of the four sides of the full rectangular aperture. We refer to these orientations as “north,” “south,” “east,” and “west,” as indicated in Figure 1.

Our models include a wedge-transmission factor τ that defines the reduction in dose caused by the wedge. Wedges are characterized by 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 pencil beams along the heel of the wedge, while τ_1 is 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:

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

When the heel of the wedge lies along the north edge, the transmission factor is

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

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

Two different wedge systems are used in clinical practice. In the first system, four different wedges

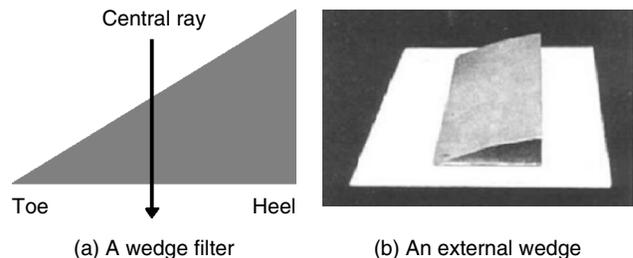


Figure 3 Wedges

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; see the appendix. We assume that the universal wedge is used. Note that the formulation assumes that the wedge rotates and not the collimator.

3. Formulating the Optimization Problems

3.1. Optimizing Beam Weights

We start with the simplest model, in which the angles from which beams are to be delivered are pre-selected, wedges are not used, and the apertures are chosen as the beam’s-eye view from each angle. All that remains is to determine the beam weights at each angle.

We now introduce some notation. The set of beam angles is denoted by \mathcal{A} . We let \mathcal{T} denote the set of all voxels that comprise the PTV, \mathcal{S} denote the voxels in the OAR (typically this is a collection of organs, $\mathcal{S} = \mathcal{S}^1 \cup \dots \cup \mathcal{S}^m$), and \mathcal{N} be the voxels in the normal tissue. We use Δ to denote the prescribed dose level at each PTV voxel, and 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 w_A , and the dose contribution to voxel (i, j, k) from a beam of unit weight from angle A is denoted by $\mathcal{D}_{A,(i,j,k)}$. (It follows that a beam of weight w_A produces 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 $A \in \mathcal{A}$ and use $\mathcal{D}_{A,\Omega}$ (and D_Ω) to denote submatrices consisting of the elements $\mathcal{D}_{A,(i,j,k)}$ (and $D_{(i,j,k)}$) for all (i, j, k) in a given voxel set Ω .

The beam weights w_A , for $A \in \mathcal{A}$, are nonnegative and are the unknowns in the optimization problem. The general form of this problem is:

$$\begin{aligned} \min_w \quad & f(D_\Omega) \\ \text{s.t.} \quad & D_\Omega = \sum_{A \in \mathcal{A}} w_A \mathcal{D}_{A,\Omega}, \quad \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\ & w_A \geq 0, \quad \forall A \in \mathcal{A}. \end{aligned} \quad (3)$$

The choice of objective function $f(D_\Omega)$ in (3) depends on the specific goal of the treatment planner. In general, the objective function measures the

mismatch between the prescription and the delivered dose. For voxels in the PTV region \mathcal{T} , there may be terms that penalize any difference between the delivered dose and the prescribed dose. For the voxels in each OAR \mathcal{S}^p ($p = 1, \dots, m$), there may be terms that penalize the amount of dose in excess of $\phi_p \Delta$, the desired upper bound on the dose to voxels in \mathcal{S}^p . (For simplicity of exposition, we consider only a single OAR.) The objective often includes terms that penalize any dose to voxels in the normal region \mathcal{N} .

The L_1 -norm (sum of absolute values) and squared L_2 norm (sum of squares; see Cormack and Quinto 1990), are both used to penalize differences between delivered and desired doses in the objective $f(D_\Omega)$. Two possible definitions of f based on these norms are

$$\begin{aligned} f(D_\Omega) = \lambda_t \frac{\|D_{\mathcal{T}} - \Delta e_{\mathcal{T}}\|_1}{|\mathcal{T}|} + \lambda_s \frac{\|(D_{\mathcal{S}} - \phi \Delta e_{\mathcal{S}})_+\|_1}{|\mathcal{S}|} \\ + \lambda_n \frac{\|D_{\mathcal{N}}\|_1}{|\mathcal{N}|}, \end{aligned} \quad (4)$$

$$\begin{aligned} f(D_\Omega) = \lambda_t \frac{\|D_{\mathcal{T}} - \Delta e_{\mathcal{T}}\|_2^2}{|\mathcal{T}|} + \lambda_s \frac{\|(D_{\mathcal{S}} - \phi \Delta e_{\mathcal{S}})_+\|_2^2}{|\mathcal{S}|} \\ + \lambda_n \frac{\|D_{\mathcal{N}}\|_2^2}{|\mathcal{N}|}. \end{aligned} \quad (5)$$

The notation $(\cdot)_+ := \max(\cdot, 0)$ in the second term defines the overdose to voxels in the OAR, 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 parameters λ_t , λ_s , and λ_n are non-negative weighting factors applied to the objective terms for the PTV, OAR, and normal voxels, respectively, while $|\mathcal{T}|$, $|\mathcal{S}|$, and $|\mathcal{N}|$ denote the number of voxels in these respective regions.

An objective function based on L_∞ -norm terms (6) allows effective penalization of hot spots in the OAR and of cold spots in the PTV. We define such a function by

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

Combinations of these objective functions can be used to achieve specific treatment goals, as described later.

Problems of the form (3) in which f is defined by (4) or (6) can be formulated as linear programs using standard techniques. For example, the term $\lambda_s \|(D_{\mathcal{S}} - \phi \Delta e_{\mathcal{S}})_+\|_1 / |\mathcal{S}|$ in (4) can be modeled by introducing a vector $V_{\mathcal{S}}$ into the formulation, along with the constraints $V_{\mathcal{S}} \geq D_{\mathcal{S}} - \phi \Delta e_{\mathcal{S}}$ and $V_{\mathcal{S}} \geq 0$, and including the term $(\lambda_s / |\mathcal{S}|) e_{\mathcal{S}}^T V_{\mathcal{S}}$ in the objective. Problems in which f is defined by (5) can be formulated as convex quadratic programs.

The treatment planner’s goals are often case-specific. For example, the planner may wish to keep the maximum dose violation on the PTV 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 might obtain the following definition of $f(D_\Omega)$ in (3):

$$\lambda_t \|D_{\mathcal{T}} - \Delta e_{\mathcal{T}}\|_\infty + \lambda_s \frac{\|(D_{\mathcal{T}} - \phi \Delta e_{\mathcal{T}})_+\|_1}{|\mathcal{S}|} + \lambda_n \frac{\|D_{\mathcal{N}}\|_1}{|\mathcal{N}|}. \quad (7)$$

In practice, voxels in the PTV that receive a dose within specified limits may be acceptable as a treatment plan. Furthermore, voxels that receive below the lower dose specification (cold spots) may be penalized more severely than hot spots in the PTV. Therefore, we consider the following definition of f :

$$f(D_\Omega) = \lambda_t^+ \|(D_{\mathcal{T}} - \theta_U \Delta e_{\mathcal{T}})_+\|_\infty + \lambda_t^- \|(\theta_L \Delta e_{\mathcal{T}} - D_{\mathcal{T}})_+\|_\infty + \lambda_s \frac{\|(D_{\mathcal{T}} - \phi \Delta e_{\mathcal{T}})_+\|_1}{|\mathcal{S}|} + \lambda_n \frac{\|D_{\mathcal{N}}\|_1}{|\mathcal{N}|}. \quad (8)$$

In this objective, θ_L is the PTV cold-spot control parameter. If the dosage delivered to a voxel in \mathcal{T} falls below $\theta_L \Delta$, a penalty term for the violation is added to the objective. Likewise, a voxel in the PTV incurs a penalty if the dose exceeds $\theta_U \Delta$.

All the models described in this paper can accommodate this separation of *hot* and *cold* spots. However, we simplify the exposition throughout by using a combined objective function. Alternative objectives have been discussed elsewhere. For example, Pugachev and Xing (2001) and Shalev et al. (1991) use score functions to evaluate and compare different plans, while Hamacher and Küfer (2002) use a multi-objective approach.

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. Optimization Beam Angles

We now consider the problem of selecting a subset of at most K beam angles from a candidate set \mathcal{A} , while simultaneously choosing optimal weights for the selected beams. In this model, binary variables ψ_A , $A \in \mathcal{A}$ indicate whether or not angle A is selected to be one of the treatment-beam orientations. The constraint $w_A \leq M\psi_A$ (for some large M) ensures that w_A is nonzero only if $\psi_A = 1$. The resulting mixed programming formulation is as follows:

$$\begin{aligned} \min_{w, \psi} \quad & f(D_\Omega) \\ \text{s.t.} \quad & D_\Omega = \sum_{A \in \mathcal{A}} w_A \mathcal{D}_{A, \Omega}, \quad \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N} \\ & 0 \leq w_A \leq M\psi_A, \quad \forall A \in \mathcal{A}, \\ & \sum_{A \in \mathcal{A}} \psi_A \leq K, \quad \psi_A \in \{0, 1\}, \quad \forall A \in \mathcal{A}. \end{aligned} \quad (9)$$

Some theoretical considerations of optimizing beam orientations are also discussed in Bortfeld and Schlegel (1993). A treatment plan involving few beams (say, three to five) generally is preferable to one of similar quality that uses more beams because it requires less time and effort to deliver. Furthermore, it has been shown that, when many beams are used, (say ≥ 5), beam orientation becomes less important in the overall optimization (Cho et al. 1999, Crooks et al. 2002, Ehr Gott and Johnston 2003, Soderstrom et al. 1995). In many cited cases, the objective is to find a minimum number of beams that satisfy the treatment goals.

Beam angles and weights can be selected either sequentially or simultaneously. Most of the earlier work in the literature uses sequential schemes (Chen et al. 1992; Gokhale et al. 1994; Myriantopoulos et al. 1992; Rowbottom et al. 1998, 1999), in which a certain number of beam angles are decided first, and their weights are subsequently determined. Rowbottom et al. (2001) optimize both variables simultaneously. To reduce the initial search space, a heuristic approach removes some beam orientations a priori, while the overall optimization problem is solved with the simplex method and simulated annealing. Prior information is included in the simultaneous optimization scheme outlined in Pugachev and Xing (2002).

A different approach has been proposed by Hass et al. (1998). They address a geometric formulation of the coplanar beam-orientation problem by means of a hybrid multi-objective genetic algorithm, which attempts to replicate the approach of a (human) treatment planner while reducing the amount of computation required. When the approach is applied without constraining the number of beams, the solution produces an indication of the minimum number of required beams. Webb (1989) applies simulated annealing to a two-dimensional treatment-planning problem. Three-dimensional problems using a simulated-annealing approach are described in Rowbottom et al. (2001) and Webb (1991, 1992, 1997), while column-generation approaches are discussed in Preciado-Walters et al. (2004).

3.3. Optimizing Wedge Orientations

Several researchers have studied the treatment planning problem with wedges. Xing et al. (1998) optimize the beam weights for an open field and two orthogonal wedged fields. Li et al. (1999) describe an algorithm for selecting both wedge orientation and beam weights, while Sherouse (1993) describes a mathematical basis for selection of wedge angle and orientation. It is noted in Xing et al. (1997) that including wedge-angle selection in the optimization makes for excessive computation time. Design of treatment plans

involving wedges are also discussed in Dai et al. (2000).

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 beam’s-eye view aperture and for each of these four wedge settings, along with the dose matrix for the open beam, as used in the formulations above. We let \mathcal{F} denote the set of wedge settings; \mathcal{F} contains five 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 $w_{A,F}$.

To include wedges in the optimization problem, we do not simply replace \mathcal{A} by $\mathcal{A} \times \mathcal{F}$ in (9); 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 the clinical situation, changing the wedge orientation takes relatively little time.) It follows that a single binary variable suffices for each angle A , so we can state the MIP model that includes beam-orientation selection as follows:

$$\begin{aligned} \min_{w, \psi} \quad & f(D_\Omega) \\ \text{s.t.} \quad & D_\Omega = \sum_{A \in \mathcal{A}, F \in \mathcal{F}} w_{A,F} \mathcal{D}_{A,F,\Omega}, \quad \Omega \in \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\ & 0 \leq w_{A,F} \leq M\psi_A, \quad \forall A \in \mathcal{A}, \forall F \in \mathcal{F}, \\ & \sum_{A \in \mathcal{A}} \psi_A \leq K, \quad \psi_A \in \{0, 1\}, \quad \forall A \in \mathcal{A}. \end{aligned} \quad (10)$$

A second consideration is that we do not wish to deliver two beams from the same angle for two diametrically opposite wedge settings. We can accommodate this restriction by introducing separate binary variables $\pi_{A,F}$ for each angle A and orientation F . A less-expensive approach is to postprocess the solution whenever

$$\begin{aligned} & \{w_{A,\text{south}} > 0 \text{ and } w_{A,\text{north}} > 0\} \quad \text{or} \\ & \{w_{A,\text{west}} > 0 \text{ and } w_{A,\text{east}} > 0\}, \end{aligned}$$

for any A , to zero out one of the weights for each pair. To illustrate the postprocessing technique, consider the “west” and “east” wedge orientations. For beamlet (i, j) , $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$, the attenuation factor when the west wedge is present is given by (1). For the east wedge, we have

$$\begin{aligned} \tau_{ij}^{\text{east}} &= \tau_0 + \frac{N-j+0.5}{N}(\tau_1 - \tau_0), \\ & i = 1, 2, \dots, M, j = 1, 2, \dots, N. \end{aligned} \quad (11)$$

Suppose now that we have a treatment plan in which, for some 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

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

which is equal to

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

Hence, an identical dose could be delivered to each voxel (i, j) 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. A similar result holds for the case of $w_{A,\text{west}} \leq w_{A,\text{east}}$.

Note that if there are other constraints on the number of wedges being used, we need to replace (10) by a formulation with additional binary variables $\pi_{A,F}$.

3.4. Computing Tight Upper Bounds on the Beam Weights

If the upper bound M on the beam weights $w_{A,F}$ is too large (as is usually the case), the feasible set is larger and the algorithm often takes longer to solve the problem. A key preprocessing technique to overcome this problem is to calculate a stringent bound on the continuous decision variables (Nemhauser and Wolsey 1988) that allows M to be chosen sufficiently large to produce an optimal solution, but not larger than necessary. We now describe a technique of this type for problem (10).

Let μ_A be the maximum dose deliverable to the PTV 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

$$\begin{aligned} \mu_A &:= \max_{F \in \mathcal{F}, (i,j,k) \in \mathcal{T}} \mathcal{D}_{A,F,(i,j,k)} = \max_{(i,j,k) \in \mathcal{T}} \mathcal{D}_{A,(i,j,k)}, \\ & A = 1, 2, \dots, |\mathcal{A}|, \end{aligned} \quad (12)$$

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 . Using the definition of τ_1 from Section 2.2, we have for a given angle A that the maximum dose

deliverable to a PTV voxel using wedge filters is

$$\mu_A \left(w_{A,0} + \tau_1 \sum_{F \in \mathcal{F} \setminus \{0\}} w_{A,F} \right), \quad (13)$$

where $0 \in \mathcal{F}$ denotes the open beam. Suppose now that we modify the model in (10) to include explicit control of hot spots by introducing an upper bound u on the dose allowed in any PTV voxel. We add the constraint

$$D_{\mathcal{F}} \leq ue_{\mathcal{F}} \quad (14)$$

to (10). By combining (14) with (13), we deduce that

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

Accordingly, we can replace the constraint $M\psi_A \geq w_{A,F}$ in (10) by

$$w_{A,0} + \tau_1 \sum_{F \in \mathcal{F} \setminus \{0\}} w_{A,F} \leq \left(\frac{u}{\mu_A} \right) \psi_A, \quad \forall A \in \mathcal{A}, \quad (15)$$

where ψ_A is the binary variable that indicates whether or not the angle A is selected. Our problem becomes

$$\begin{aligned} \min_{w, \psi} \quad & f(D_{\Omega}) \\ \text{s.t.} \quad & D_{\Omega} = \sum_{A \in \mathcal{A}, F \in \mathcal{F}} w_{A,F} \mathcal{D}_{A,F,\Omega}, \quad \Omega \in \mathcal{F} \cup \mathcal{F} \cup \mathcal{N}, \\ & (u/\mu_A)\psi_A \geq w_{A,0} + \tau_1 \sum_{F \in \mathcal{F} \setminus \{0\}} w_{A,F} \\ & K \geq \sum_{A \in \mathcal{A}} \psi_A, \\ & w_{A,F} \geq 0, \quad \forall A \in \mathcal{A}, \forall F \in \mathcal{F}, \\ & \psi_A \in \{0, 1\}, \quad \forall A \in \mathcal{A}, \forall F \in \mathcal{F}. \end{aligned} \quad (16)$$

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.

4. Techniques for DVH Control

Dose-volume histograms (DVH) are a compact way to represent 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 attempt to control the fundamental aspects of the treatment plan. For instance, the oncologist is often willing to sacrifice some specified portion of an OAR (such as the lung) in order to provide an adequate probability of tumor control (especially if the OAR lies near the tumor). This aim is realized by requiring that at least a specified percentage of the OAR must receive a dose less than a specified level. DVH constraints are used to control uniformity of the

dose to the PTV, and to avoid cold spots. For example, the planner may require all voxels in the PTV to receive doses of between 95% and 107% of the prescribed dose Δ .

Modelers usually are advised to update the weights $(\lambda_t^+, \lambda_t^-, \lambda_s, \lambda_n)$ to achieve DVH control. However, as pointed out in Dennis and Das (1997), understanding the relationship between the λ values and their intended consequences is far from straightforward. Rather than focusing our tuning efforts on these weights, we suggest manipulating other parameters in the model; specifically, the PTV control parameters θ_U and θ_L and the hot-spot control parameter ϕ in (8). We describe these techniques with reference to the problem in (9).

There are three typical requirements for the radiation treatment: *homogeneity*, *conformity*, and *avoidance* (Ferris et al. 2003a, b). In our formulations, homogeneity is controlled by θ_L and θ_U , which define the lower and upper bounds on the dose to PTV voxels (we have $\theta_L \leq 1 \leq \theta_U$). The conformity constraints, which require the dose to the normal tissue to be as small as possible, can be implemented by increasing the weight λ_n on the normal-tissue term in the objective. Avoidance constraints, which take the form of DVH constraints on the OAR, are implemented via the hot-spot control parameter ϕ .

We show results for DVH control of a clinical case (a pancreatic tumor) in most of the following sections. This problem, which is illustrated in Figure 10, includes four critical structures (two kidneys, the spinal cord, and the liver) in the vicinity of the tumor. For illustrative purposes, we condense all the OARs in some figures into a single OAR and graph its dose distribution with a single line. Of course, the techniques we present in this section can be used to control doses to each organ individually, as we demonstrate in Section 6.

4.1. Choice of Norms in the Objective Functions

We introduced different types of objective functions in Section 3.1; see in particular (4), (5), (6), and (8). One uses infinity norms 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. Specifically, we compare the function in (8) (with $\lambda_t^+ = \lambda_t^- = \lambda_s = \lambda_n = 1$) against a function in which the infinity norms in the first two terms are replaced by L_1 norms, scaled by the cardinality of the target set. We use $\theta_L = 0.95$, $\theta_U = 1.07$, $\phi = 0.2$, and $K = 4$ in this experiment. As might be expected,

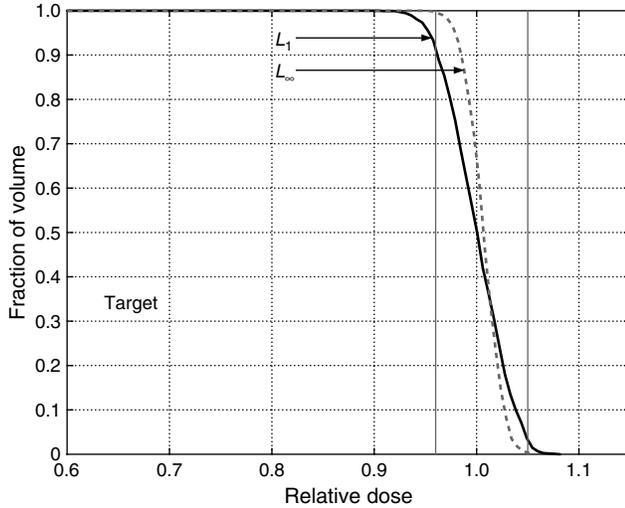


Figure 4 Cumulative Dose-Volume Histogram on the PTV

(8) has better control on the PTV as shown in Figure 4; the infinity norm yielded a stricter enforcement of the constraints on the PTV. The two objective functions can produce a similar solution if the values of the λ_i s are chosen appropriately. However, our experience indicates that it is easier to choose an appropriate value of λ_i for the L_∞ penalty than it is to tune this parameter for the L_1 norm. (In the normal and OAR regions, the difference in quality of the solutions obtained from these two alternative objectives was insignificant.)

4.2. DVH Control on the PTV

Here we consider the optimization problem (9) with objective function $f(D_\Omega)$ defined by (8). We aim to attain homogeneity of the dose on \mathcal{T} without sacrificing too much quality in the dose profile for the normal region and OAR. As discussed above, the key parameters in (8) with respect to this goal are θ_U and θ_L . In this experiment, we fix $\theta_U = 1.07$, and try the values 0.7, 0.8, 0.9, and 0.94 for the lower-bound fraction θ_L . Figure 5 shows the DVH plots for the four different values of θ_L , for the PTV, OAR, and normal regions. For each value, we find that 100% of the PTV receives more than the desired lower bound θ_L ; we manage to avoid PTV cold spots completely in this example. We might expect that larger values of θ_L (which confine the target dose to a tighter range) would result in a less attractive solution in the OAR and the normal tissue, but Figure 5 shows that the loss of treatment quality is not significant. We conclude that the use of θ_U and θ_L to implement homogeneity constraints is effective.

4.3. DVH Control on the OAR

We show here that that the dose to the OAR can be controlled by means of the parameter ϕ in (8),

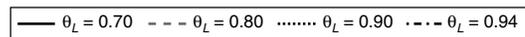
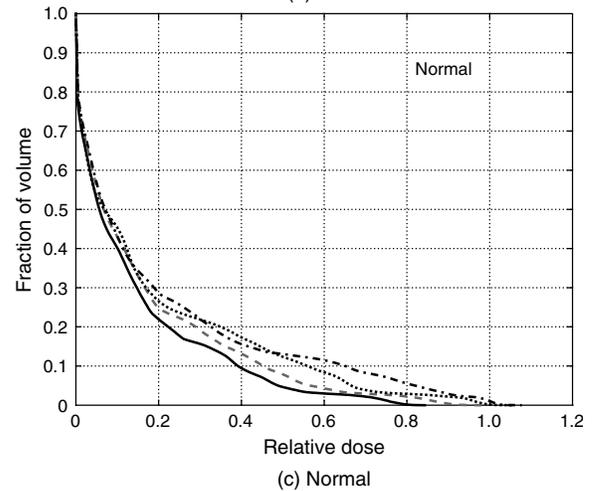
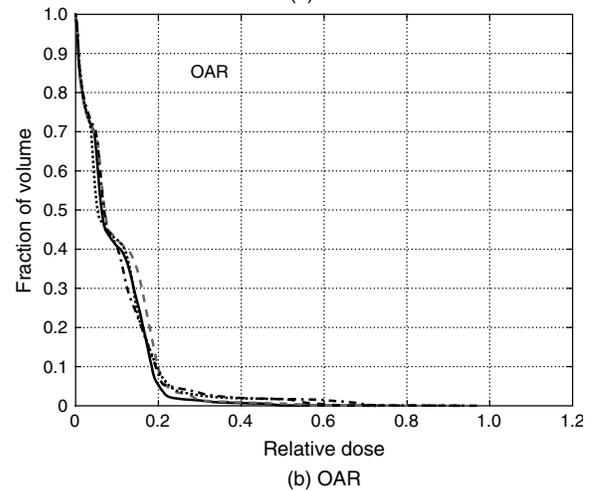
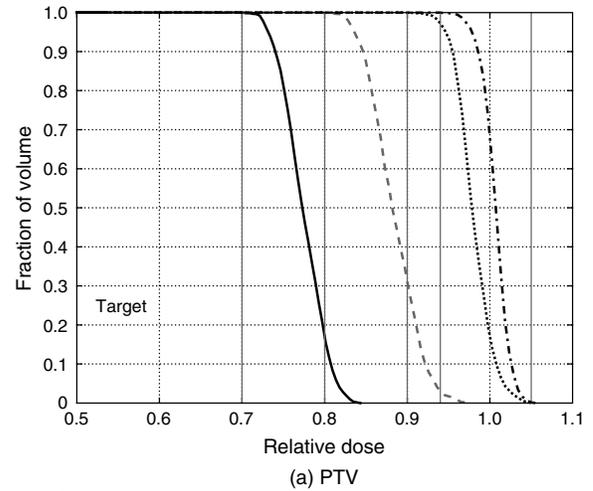


Figure 5 Cumulative DVH Control for Different Choices of Parameter θ_L

assuming that the weights λ_t , $\lambda_{s,t}$ and λ_n have been fixed appropriately. As shown in Figure 6, we set ϕ to various values in the range $[0, 1]$. For $\phi = 0.5$, almost all of the OAR receives dose less than 50% of the prescribed target dose. Similar results hold for the

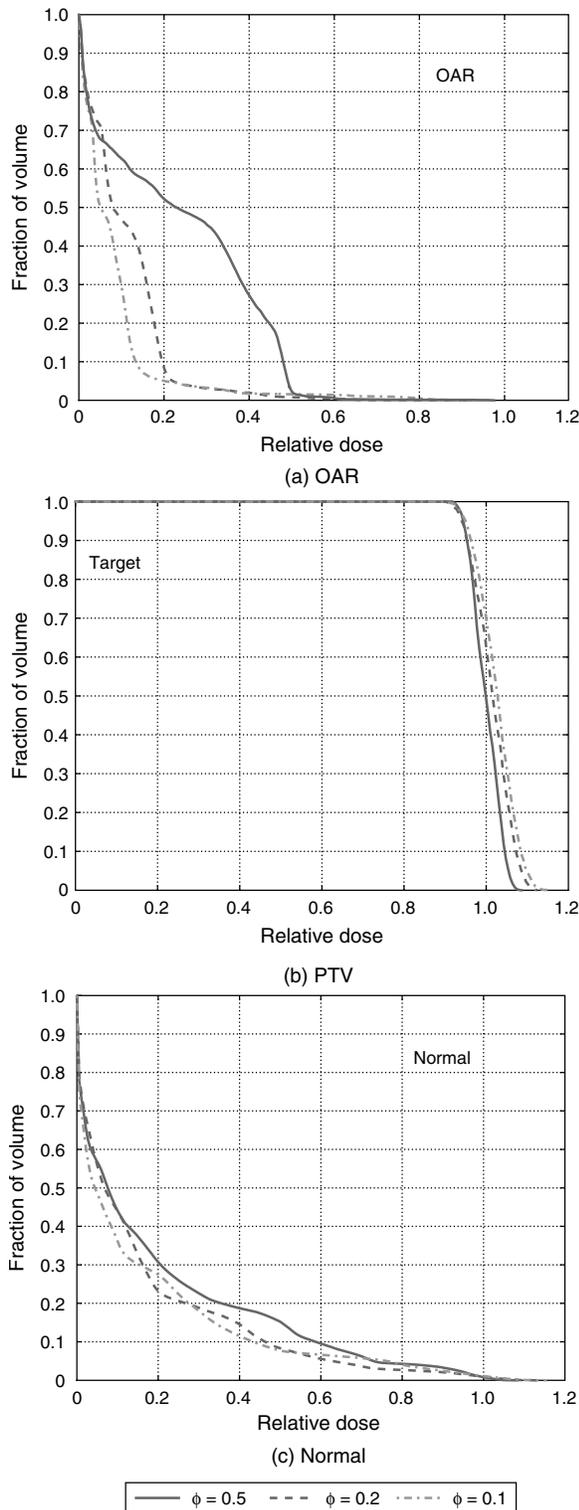


Figure 6 Cumulative DVH Control for Different Values of Parameter ϕ

values $\phi = 0.2$ and $\phi = 0.1$. (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.) Better control of the dose to OAR causes loss of treatment quality on the PTV and the

normal tissue, but Figure 6 shows that the degradation is not significant.

We note that if our goal is to control hot spots in the OAR rather than the integral dose, we could replace the term $\|(D_{\mathcal{O}} - \phi \Delta e_{\mathcal{O}})_+\|_1$ in the objective (8) by its infinity-norm counterpart $\|(D_{\mathcal{O}} - \phi \Delta e_{\mathcal{O}})_+\|_{\infty}$.

The parameter ϕ can be updated on a per-organ basis if the DVH requirement for a given OAR is not satisfied. Furthermore, there can be some conflict between the goals of controlling DVH on target and nontarget regions, as the proximity of PTV to normal regions and OAR makes it inevitable that some nontarget voxels will receive high doses. If the PTV dose control is most important (as is usually the case), the control parameters θ_L , θ_U , and ϕ should be chosen with $\theta_U - \theta_L$ small and ϕ as a fairly large (but smaller than 1) fraction of the prescribed target dose Δ . However, if the OAR dose control is most important, a smaller value of ϕ should 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.

4.4. DVH Control via Wedges

In general, the use of wedges gives more flexibility in achieving adequate coverage of the tumor 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 7 shows DVH graphs obtained for a treatment plan using wedges (16) and one using no wedges (17). Conventionally, four or six beams are usually used to treat cases of this type. However, we use three beam angles ($K = 3$) to emphasize the effect of wedges. Figure 7(a) shows that a significant improvement on the OAR is achieved by adding wedges. In Figure 7(b), we see that there is also a slight improvement in the DVH for the PTV, and little difference between the wedge and no-wedge cases for the normal tissue.

5. Reducing the Solution Time

The problem (10) involves numerous variables (some of them discrete) and a large amount of data, mostly in the form of dose matrices. Therefore, the optimization problem is time-consuming to construct and solve. In this section, we describe a number of techniques to reduce the solution time. First, we show how normal-tissue voxels that are distant from the PTV can be merged, thereby reducing the number of variables without an appreciable change in solution quality. Second, 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.

We note that Legras et al. (1982) describe a multi-phase approach in which linear programs are solved

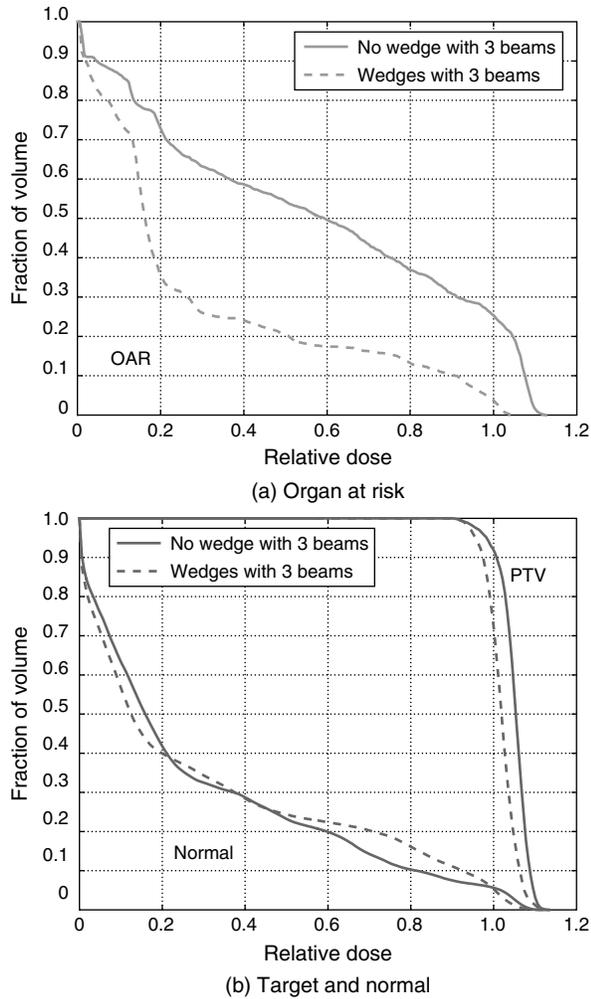


Figure 7 Cumulative DVH: Effect of Wedges in a Prostate Cancer Case with Three Beam Angles

to determine the most promising beam angles, with a refined solution being obtained from a final nonlinear program. Our approach in Section 5.2 is different in that each of our phases involves the solution of MIPs in which the angles are selected explicitly. The phases differ from each other in the reduced sets of voxels that are used as the basis of the problem formulation.

5.1. Reducing Resolution in the Normal Tissue

Since the main focus of the planning problem is to deliver enough dose to the PTV while avoiding organs at risk, the dosage to normal regions that are some distance away from the PTV need not 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 ρ , we define a neighborhood of the PTV as follows:

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

where $\text{dist}((i, j, k), \mathcal{T})$ denotes the Euclidean distance of the center of the voxel (i, j, k) to the PTV. 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:

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

Finally, we include in the optimization problem only those voxels that are close to the PTV, or that lie in an OAR, or that lie in the reduced normal region; that is,

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

(see related work in Bahr et al. 1968 and Morrill et al. 1990). Since each of the voxels $(i, j, k) \in \mathcal{N}_1$ effectively represents itself and seven neighboring voxels, the weights applied to the voxels $(i, j, k) \in \mathcal{N}_1$ in the objective functions (4) and (5) should be increased correspondingly. An appropriate replacement for the term $\|D_{\mathcal{N}}\|_1/|\mathcal{N}|$ in (4) could then be

$$\frac{\|D_{\mathcal{R}_\rho(\mathcal{T})}\|_1 + \|D_{\mathcal{N}_1}\|_1(|\mathcal{N} \setminus \mathcal{R}_\rho(\mathcal{T})|/|\mathcal{N}_1|)}{|\mathcal{N}|}.$$

5.2. A Three-Phase Approach

We now discuss our multiphase approach that “ramps up” to the solution of the full problem via a sequence of models. Essentially, the models are solved in increasing order of difficulty, with the solution of one model providing a good starting point for the next. The models differ from each other in the selection of voxels included in the formulation, and in the number of beam angles allowed.

If the most promising beam angles can be identified in advance, the full problem can be solved with a small number of discrete variables. A simple approach for removing unpromising beam angles is to remove from consideration those that pass directly through any OAR (Rowbottom et al. 2001). A more elaborate approach (Pugachev and Xing 2001) introduces a score function for each candidate angle, based on the ability of that angle to deliver a high dose to the PTV without exceeding prescribed dose tolerances to OAR or to normal tissue located along its path. Only beam angles with good 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 as follows.

5.2.1. Phase 1: Selection of Promising Beam Angles. Our aim in this phase is to construct a subset of beam angles \mathcal{A}_1 that are likely to appear in the final solution of (10). (A similar technique can be applied to (16).) We solve a collection of r MIPs, where each MIP is constructed from a reduced set of voxels consisting of the voxels in the PTV, a randomly sampled 10% of the OAR voxels (\mathcal{S}'), and the voxels in $\mathcal{R}_\rho(\mathcal{T})$; that is, $\Omega_1 = \{\mathcal{T} \cup \mathcal{S}' \cup \mathcal{R}_\rho(\mathcal{T})\}$. We define \mathcal{A}_1 as the set of all angles A with $w_A > 0$ in at least one of the r sampled problems.

5.2.2. Phase 2: Treatment-Beam-Angle Determination. In the next phase, we select K or fewer treatment beam angles from \mathcal{A}_1 . We solve a version of (10) using \mathcal{A}_1 in place of \mathcal{A} and a reduced set of voxels defined as $\Omega_2 = \{\mathcal{T} \cup \mathcal{S} \cup \mathcal{R}_\rho(\mathcal{T}) \cup \mathcal{N}_1\}$. Note that $|\mathcal{A}_1|$ is typically greater than or equal to K , so the binary variables play a nontrivial role in this phase.

5.2.3. Phase 3: Final Approximation. In the final phase, we fix the K beam angles (by fixing $\psi_A = 1$ for the angles A selected in Phase 2 and $\psi_A = 0$ otherwise) and solve the resulting simplified optimization problem over the complete set of voxels. This final approximation typically takes much less time to solve than does the full-scale model, both because of the smaller amount of data (due to fewer beam angles) and the absence of binary variables.

Although there is no guarantee that this technique will produce the same solution as the original full-scale model (10), we have found that the quality of its approximate solution is close to optimal. Computational experience with this approach is given in Section 6.

6. Computational Performance

In this section, we discuss the performance of our proposed approaches (including the three-phase approach) on the pancreatic data set introduced in Section 4. We also describe the treatment plan obtained for this data set for the full model in which DVH controls and wedges are all applied.

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

$$\begin{aligned}
 & \min_{w, \psi} f(D_\Omega) \\
 & \text{s.t. } D_\Omega = \sum_{A \in \mathcal{A}} w_A \mathcal{D}_{A, \Omega}, \quad \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}, \\
 & D_{\mathcal{T}} \leq u e_{\mathcal{T}}, \\
 & 0 \leq w_A \leq M \psi_A, \quad \forall A \in \mathcal{A}, \\
 & \sum_{A \in \mathcal{A}} \psi_A \leq K, \\
 & \psi_A \in \{0, 1\}, \quad \forall A \in \mathcal{A},
 \end{aligned} \tag{17}$$

where $f(D_\Omega)$ is defined by (8). Note that we have introduced the hard upper bound $D_{\mathcal{T}} \leq u$ (see (14)) on the target, allowing us to experiment using the tightened upper bounds of Section 3.4. We fix the parameters in (17) as follows: $\theta_L = 0.95$, $\theta_U = 1.07$, $\phi = 0.2$, $K = 4$, $\lambda_i^+ = \lambda_i^- = \lambda_s = \lambda_n = 1$, $u = 1.15$, and $|\mathcal{A}| = 36$. The set of angles \mathcal{A} consists of angles equally spaced by 10° in a full 360° circumference.

First, we solve (17) using the full set of voxels. The MIP solver is set to terminate when the gap between the upper and lower bounds of the objective value falls below 1% (in relative terms). This calculation and the others in this section were performed on a Pentium 4, 1.8 GHz PC running Linux. The problems were modeled in the GAMS modeling language (Brooke et al. 1988) and CPLEX 7.1 was used as the LP and MIP solver.

Figure 8 shows changes of upper and lower bounds on the optimal objective value as the iteration number increases, where iteration count is the total number of branch-and-bound nodes explored. Only slight improvements to the upper bound (which represents the best integer solutions found to date) occur after the first 220,000 iterations, and the lower bound of the objective value increases slowly beyond this point. We set the “big M ” value to 2 for this experiment; the total computation time of over 112 hours is shown in column I of Table 1. This table also shows the effects of the computational speedups described in Section 5. In columns II, III, and IV we use the tight bound (15) 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 (17) by $w_A \leq (u/\mu_A) \psi_A$.) In addition, column III shows the effects of using the reduced-voxel version of the problem discussed in Section 5.1. Finally, column IV shows results obtained with the three-phase approach of Section 5.2 using $r = 10$ samples of the OAR.

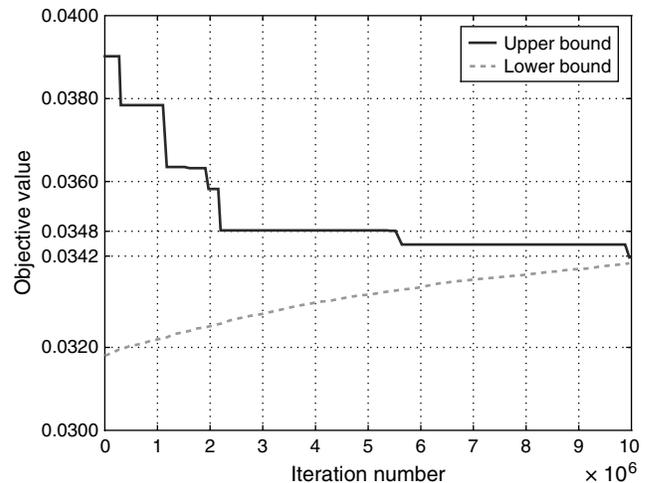


Figure 8 Progress of Upper and Lower Bounds during MIP Algorithm

Table 1 Comparisons Among Different Solution Schemes

	I	II	III	IV
Approach	Single solve	Single solve	Reduced model	Three-phase
Bound (M)	2	u/μ_A	u/μ_A	u/μ_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

For purposes of comparing the quality of the computational results obtained with these four approaches, we calculated the final objective values on the full set of voxels. To three significant figures, these values were the same. 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 is obtained by using the tighter bound. Column III shows a computational savings of almost three quarters, without degradation of solution quality, when a reduced model is used. The full problem contains 1,244 voxels in the PTV, 69,270 voxels in the OAR, and 747,667 voxels in the normal region, while the reduced model has 1,244 voxels in the PTV, 14,973 voxels in the OAR, and 96,154 voxels in the normal tissue. The most dramatic savings, however, were for the three-phase scheme, which yielded a savings of 99.5% over the direct solution scheme with no appreciable effect on the quality of the solution. The difficulty of the full problem arises in large part from the hot-spot and cold-spot control terms. Using looser values for these parameter values speeds up the the solution time considerably.

We believe our three-phase technique is equally effective for the case in which wedges are included in the formulation. Hence, we performed a final test with the pancreatic data set on a full model that includes wedges. This case is made particularly difficult by the close proximity of the PTV to several OARs, including the spinal cord, liver, left kidney, and right kidney. Besides the 36 beam angles, we allow wedges with four possible orientations (at points of the compass) at each angle.

The specific goals of the treatment plan were defined as follows:

1. Four beam angles.
2. As the highest priority, the target volume should receive a dose of between 95% and 107% of the prescribed dose.
3. 90% of each OAR should receive less than 20% of the target prescribed dose level.
4. The integral dose delivered to the normal tissue should be kept as small as possible.

To achieve these goals, we set DVH control parameters as follows: $\theta_L = 0.95$, $\theta_U = 1.07$, $r = 10$, $K = 4$, and $\phi = 0.2$, for each organ in {spinal cord, liver,

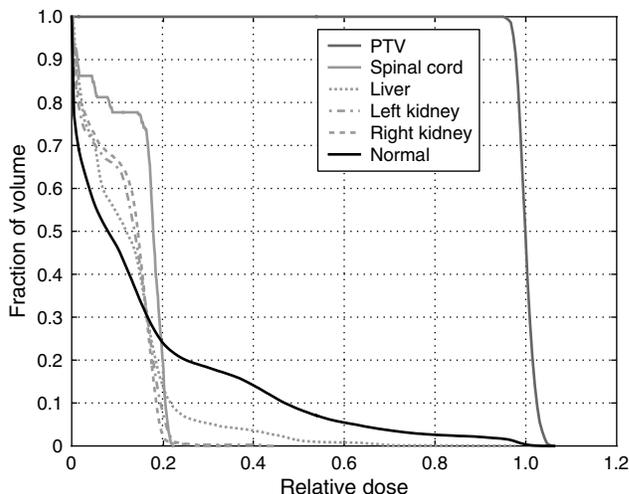


Figure 9 Cumulative Dose-Volume Histogram at Optimum

left kidney, right kidney}. (We included a separate term of the form $\lambda_s \|(D_{\mathcal{S}} - \phi \Delta e_{\mathcal{S}})_+\|_1 / |\mathcal{S}|$ in the objective function for each of the four individual critical structures.)

Figure 9 shows DVH plots of this experiment. The homogeneity constraints are satisfied for the PTV; every voxel in the PTV receives between 95% and 107% of the prescribed dose. It is also clear that approximately 90% of each OAR receives at most 20% of the target prescribed dose, as specified; the DVH plot for each OAR passes very close to the point (0.2, 0.1) that corresponds to the aforementioned treatment goal.

Figure 10 shows isodose lines on the slices through the treatment region obtained by computerized tomography. The PTV is outlined within four isodose lines. The outermost line is the 20% isodose line, which encloses a region in which the voxels receive a dose of at least 20% of the PTV prescribed dose. Moving inwards towards the PTV, we see 50%, 80%, and 95% isodose lines. Figure 10(a) shows an axial slice. The kidneys are outlined as two circles right below the PTV. As can be seen, the PTV lies well inside the 95% isodose line, while the dose to the organs at risk remains reasonable. Figure 10(b) shows a sagittal view of the PTV with those four isodose lines also.

The three-phase approach outlined here has been used in a number of other studies. Examples of the benefits of this procedure on breast, pancreatic, head, and neck cases, for example, can be found in Ferris et al. (2004).

7. Summary

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

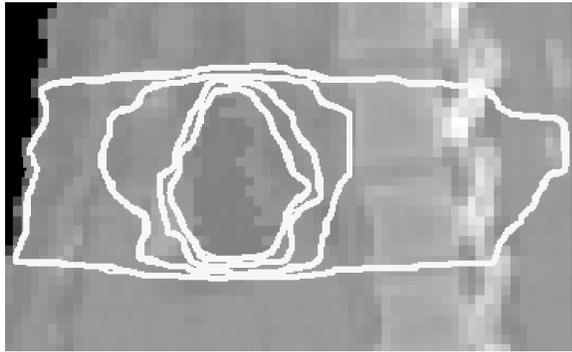
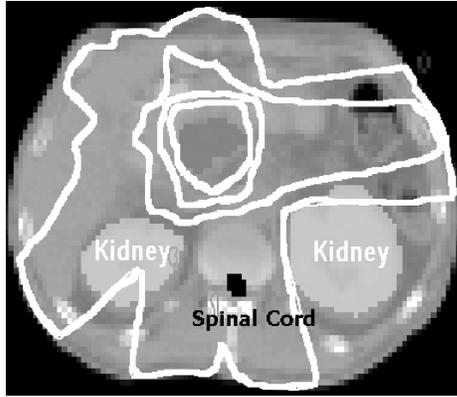


Figure 10 Isodose Plots: Lines Represent 20%, 50%, 80%, and 95% Isodoses (20% Line Outermost)

1. Simultaneous optimization of three key parameters (beam angles, wedge orientations, and beam weights) via a sequential-sampling and angle-set-reduction technique.
2. The use of a penalization scheme (rather than integer variables) to control the DVH.
3. A three-phase algorithm to find the solution rapidly using standard software tools for LP and MIP, and a technique for bounding maximum beam weights a priori, to increase the efficiency of the MIP solver. Our approach yielded a 99.5% improvement in runtime over direct solution of a full-resolution problem on a clinical data set.

Acknowledgments

The authors acknowledge the support of the National Science Foundation under Grants CCF-0113051 and DMS-0427689. The authors also acknowledge support from the Air Force Office of Scientific Research under Grant FA9550-04-1-0192.

Appendix. Universal Wedge

We show here that a treatment plan that requires the use of a wedge is in some cases equivalent to one that uses a wedge with different properties in combination

with an open (unwedged) beam of the same shape. This result implies that a single “universal” wedge suffices in designing a wide range of treatment plans; not much is to be gained by using a range of wedges with different properties.

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 $w'_{A,open}$ through the open beam, and $w'_{A,west}$ through the wedge. (The attenuation parameter τ_{ij} for beamlet (i, j) is given by (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

$$\begin{aligned} w'_{A,open} + w'_{A,west} \left[\tau'_0 + \frac{j-0.5}{N} (\tau'_1 - \tau'_0) \right] \\ = w'_{A,open} + w'_{A,west} [\tau'_0 - 0.5/N(\tau'_1 - \tau'_0)] \\ + jw'_{A,west} (\tau'_1 - \tau'_0)/N. \end{aligned}$$

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

$$\begin{aligned} w_{A,open} + w_{A,west} [\tau_0 - 0.5/N(\tau_1 - \tau_0)] \\ + jw_{A,west} (\tau_1 - \tau_0)/N. \end{aligned}$$

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

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

and

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

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

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

and

$$w_{A,open} = w'_{A,open} + w'_{A,west} \left[\tau'_0 - \frac{\tau'_1 - \tau'_0}{\tau_1 - \tau_0} \tau_0 \right]. \quad (18)$$

Note that $w_{A,west}$ is always nonnegative whenever $w'_{A,west}$ is nonnegative, but that $w_{A,open}$ is not necessarily nonnegative, even when the weights for the original wedge are both nonnegative. However, a sufficient condition for $w_{A,open}$ to be nonnegative for *any* nonnegative values of $w'_{A,open}$ and $w'_{A,west}$ 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 (18) 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 $\tau_1 - \tau_0$.

References

- Bahr, G. K., J. G. Kereiakes, H. Horwitz, R. Finney, J. Galvin, K. Goode. 1968. The method of linear programming applied to radiation treatment planning. *Radiology* **91** 686–693.
- Bortfeld, Th., W. Schlegel. 1993. Optimization of beam orientations in radiation therapy: Some theoretical considerations. *Phys. Medicine Biol.* **38** 291–304.
- Bortfeld, Th. R., J. Burkelbach, R. Boesecke, W. Schlegel. 1990. Methods of image reconstruction from projections applied to conformation radiotherapy. *Phys. Medicine Biol.* **25** 1423–1434.
- Bortfeld, Th. R., A. L. Boyer, W. Schlegel, D. L. Kahler, T. J. Waldron. 1994. Realization and verification of three-dimensional conformal radiotherapy with modulated fields. *Internat. J. Radiation Oncology, Biol. Phys.* **30** 899–908.
- Brooke, A., D. Kendrick, A. Meeraus. 1988. *GAMS: A User's Guide*. The Scientific Press, South San Francisco, CA.
- Chen, G. T. Y., D. R. Spelbring, C. A. Pelizzari, J. M. Balter, L. C. Myriantopoulos, S. Vijayakumar, H. Halpern. 1992. The use of beam eye view volumetrics in the selection of noncoplanar radiation portals. *Internat. J. Radiation Oncology: Biol., Phys.* **23** 153–163.
- Chen, Y., D. Michalski, C. Houser, J. M. Galvin. 2002. A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy. *Phys. Medicine Biol.* **47** 1647–1658.
- Cho, B. C. J., W. H. Roa, D. Robinson, B. Murray. 1999. The development of target-eye-view maps for selection of coplanar or noncoplanar beams in conformal radiotherapy treatment planning. *Medical Phys.* **26** 2367–2372.
- Cormack, A., E. Quinto. 1990. The mathematics and physics of radiation dose planning using X-rays. *Contemporary Math.* **113** 41–55.
- Crooks, S. M., A. Pugachev, C. King, L. Xing. 2002. Examination of the effect of increasing the number of radiation beams on a radiation treatment plan. *Phys. Medicine Biol.* **47** 3485–3501.
- Dai, J., Y. Zhu, Q. Ji. 2000. Optimizing beam weights and wedge filters with the concept of the super-omni wedge. *Medical Phys.* **27** 2757–2762.
- Dennis, J., I. Das. 1997. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural Optim.* **14** 63–69.
- Earl, M. A., D. M. Shepard, S. Naqvi, X. A. Li, C. Yu. 2003. Inverse planning for intensity modulated arc therapy using direct aperture optimization. *Phys. Medicine Biol.* **48** 1075–1089.
- Ehrgott, M., M. Burjony. 2001. Radiation therapy planning by multicriteria optimization. *Proc. 36th Annual Conf. Oper. Res. Soc. New Zealand*, 244–253.
- Ehrgott, M., R. Johnston. 2003. Optimisation of beam directions in intensity modulated radiation therapy planning. *OR Spectrum* **25** 251–264.
- Ferris, M. C., J.-H. Lim, D. M. Shepard. 2003a. Optimization approaches for treatment planning on a Gamma Knife. *SIAM J. Optim.* **13** 921–937.
- Ferris, M. C., J.-H. Lim, D. M. Shepard. 2003b. Radiosurgery treatment planning via nonlinear programming. *Ann. Oper. Res.* **119** 247–260.
- Ferris, M. C., R. R. Meyer, W. D'Souza. 2005. Radiation treatment planning: Mixed integer programming formulations and approaches. H. P. Williams, G. Appa, L. Pitsoulis, eds. *Handbook on Modelling for Discrete Optimization*. Kluwer Academic Publishers, Boston, MA.
- Ferris, M. C., R. Einarsson, Z. Jiang, D. Shepard. 2004. Sampling issues for optimization in radiotherapy. Technical Report, Comput. Sci. Dept., University of Wisconsin, Madison, WI.
- Goitein, M., M. Abrams, S. Rowell, H. Pollari, J. Wiles. 1983. Multi-dimensional treatment planning: II. Beam's eye-view, back projection, and projection through ct sections. *Internat. J. Radiation Oncology, Biol., Phys.* **9** 789–797.
- Gokhale, P., E. M. A. Hussein, N. Kulkarni. 1994. Determination of beam orientation in radiotherapy planning. *Medical Phys.* **21** 393–400.
- Hamacher, H. W., K.-H. Küfer. 2002. Inverse radiation therapy planning—A multiple objective optimization approach. *Discrete Appl. Math.* **118** 145–161.
- Hass, O. C. L., K. J. Burnham, J. A. Mills. 1998. Optimization of beam orientation in radiotherapy using planar geometry. *Phys. Medicine Biol.* **43** 2179–2193.
- Holder, A. 2001. Partitioning multiple objective optimal solutions with applications in radiotherapy design. Math. Technical Report 54, Trinity University, San Antonio, TX.
- Holder, A. 2003. Designing radiotherapy plans with elastic constraints and interior point methods. *Health Care and Management Sci.* **6** 5–16.
- Holder, A. 2004. Radiotherapy treatment design and linear programming. M. L. Brandeau, F. Sainfort, W. P. Pierskalla, eds. *Oper. Res. and Health Care: A Handbook of Methods and Applications*. Kluwer Academic Publishers, Boston, MA, 741–774.
- Intensity Modulated Radiation Therapy Collaborative Working Group. 2001. Intensity-modulated radiotherapy: Current status and issues of interest. *Internat. J. Radiation Oncology: Biol., Phys.* **51** 880–914.
- Jordan, T. J., P. C. Williams. 1994. The design and performance characteristics of a multileaf collimator. *Phys. Medicine Biol.* **39** 231–251.
- Langer, M., R. Brown, M. Urie, J. Leong, M. Stracher, J. Shapiro. 1990. Large scale optimization of beam weights under dose-volume restrictions. *Internat. J. Radiation Oncology, Biol. Phys.* **18** 887–893.
- Lee, E. K., T. Fox, I. Crocker. 2000. Optimization of radiosurgery treatment planning via mixed integer programming. *Medical Phys.* **27** 995–1004.
- Lee, E. K., T. Fox, I. Crocker. 2003. Integer programming applied to intensity-modulated radiation therapy treatment planning. *Ann. Oper. Res.* **119** 165–181.
- Legras, J., B. Legras, J. Lambert. 1982. Software for linear and nonlinear optimization in external radiotherapy. *Comput. Programs Biomedicine* **15** 233–242.
- Leong, J., M. Langer. 1987. Optimization of beam weights under dose-volume restrictions. *Internat. J. Radiation Oncology, Biol. Phys.* **13** 1255–1260.
- Li, J. G., A. L. Boyer, L. Xing. 1999. Clinical implementation of wedge filter optimization in three-dimensional radiotherapy treatment planning. *Radiation Therapy and Oncology* **53** 257–264.
- Lodwick, W., S. McCourt, F. Newman, S. Humphries. 1999. Optimization methods for radiation therapy plans. C. Borgers, F. Natterer, eds. *Computational Radiology and Imaging: Therapy and Diagnosis*. IMA Series in Appl. Math., Springer-Verlag, 229–250.
- Morrill, S., I. Rosen, R. Lane, J. Belli. 1990. The influence of dose constraint point placement on optimized radiation therapy treatment planning. *Internat. J. Radiation Oncology, Biol. Phys.* **19** 129–141.

- Myriantopoulos, L. C., G. T. Y. Chen, S. Vijayakumar, H. Halpern, D. R. Spelbring, C. A. Pelizzari. 1992. Beams eye view volumetrics—An aid in rapid treatment plan development and evaluation. *Internat. J. Radiation Oncology: Biol., Phys.* **23** 367–375.
- Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization*. Wiley, New York.
- Preciado-Walters, F., R. Rardin, M. Langer, V. Thai. 2004. A coupled column generation, mixed-integer approach to optimal planning of intensity modulated radiation therapy for cancer. *Math. Programming* **101** 319–338.
- Pugachev, A., L. Xing. 2001. Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *Internat. J. Radiation Oncology, Biol., Phys.* **51** 1361–1370.
- Pugachev, A., L. Xing. 2002. Incorporating prior knowledge into beam orientation optimization in IMRT. *Internat. J. Radiation Oncology, Biol. Phys.* **54** 1565–1574.
- Rosen, I., R. Lane, S. Morrill, J. Belli. 1990. Treatment plan optimization using linear programming. *Medical Phys.* **18** 141–152.
- Rowbottom, C. G., V. S. Khoo, S. Webb. 2001. Simultaneous optimization of beam orientations and beam weights in conformal radiotherapy. *Medical Phys.* **28** 1696–1702.
- Rowbottom, C. G., S. Webb, M. Oldham. 1998. Improvements in prostate radiotherapy from the customization of beam directions. *Medical Phys.* **25** 1171–1179.
- Rowbottom, C. G., S. Webb, M. Oldham. 1999. Beam-orientation customization using an artificial neural network. *Phys. Medicine Biol.* **44** 2251–2262.
- Shalev, S., D. Viggars, M. Carey, P. Hahn. 1991. The objective evaluation of alternative treatment plans 2 score functions. *Internat. J. Radiation Oncology: Biol., Phys.* **20** 1067–1073.
- Shepard, D. M., M. A. Earl, X. A. Li, C. Yu. 2002. Direct aperture optimization: A turnkey solution for step-and-shoot IMRT. *Medical Phys.* **29** 1007–1018.
- Shepard, D. M., M. C. Ferris, G. Olivera, T. R. Mackie. 1999. Optimizing the delivery of radiation to cancer patients. *SIAM Rev.* **41** 721–744.
- Sherouse, G. W. 1993. A mathematical basis for selection of wedge angle and orientation. *Medical Phys.* **20** 1211–1218.
- Soderstrom, S., A. Gustafsson, A. Brahme. 1995. Few-field radiation-therapy optimization in the phase-space of complication-free tumor central. *Internat. J. Imaging Systems Tech.* **6** 91–103.
- Sonderman, D., P. Abrahamson. 1985. Radiotherapy treatment design using mathematical programming models. *Oper. Res.* **33** 705–725.
- Tervo, J., P. Kolmonen. 2000. A model for the control of a multileaf collimator in radiation therapy treatment planning. *Inverse Problems* **16** 1875–1895.
- Webb, S. 1989. Optimisation of conformal radiotherapy dose distributions by simulated annealing. *Phys. Medicine Biol.* **34** 1349–1370.
- Webb, S. 1991. Optimization by simulated annealing of three-dimensional, conformal treatment planning for radiation fields defined by a multileaf collimator. *Phys. Medicine Biol.* **36** 1201–1226.
- Webb, S. 1992. 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. *Phys. Medicine Biol.* **37** 1689–1704.
- Webb, S. 1997. *The Physics of Conformal Radiotherapy: Advances in Technology*. Institute of Physics Publishing Ltd., Bristol, England.
- Webb, S. 1998. Configuration options for intensity-modulated radiation therapy using multiple static fields shaped by a multileaf collimator. *Phys. Medicine Biol.* **43** 241–260.
- Wu, X., Y. Zhu. 2001. A global optimization method for three-dimensional conformal radiotherapy treatment planning. *Phys. Medicine Biol.* **46** 109–119.
- Xiao, Y., Y. Censor, D. Michalski, J. M. Galvin. 2003. The least-intensity feasible solution for aperture-based inverse planning in radiation therapy. *Ann. Oper. Res.* **119** 183–203.
- Xing, L., R. J. Hamilton, C. A. Pelizzari, G. T. Y. Chen. 1998. A three-dimensional algorithm for optimizing beam weights and wedge filters. *Medical Phys.* **25** 1858–1865.
- Xing, L., C. A. Pelizzari, F. T. Kuchnir, G. T. Y. Chen. 1997. Optimization of relative weights and wedge angles in treatment planning. *Medical Phys.* **24** 215–221.