EFFICIENT SCHEMES FOR ROBUST IMRT TREATMENT PLANNING

A ÓLAFSSON[1] AND S J WRIGHT[2]

Abstract. We use robust optimization techniques to formulate an IMRT treatment planning problem in which the dose matrices are uncertain, due to both dose calculation errors and inter-fraction positional uncertainty of tumor and organs. When the uncertainty is taken into account, the original linear programming formulation becomes a second-order cone program. We describe a novel and efficient approach for solving this problem, and present results to compare the performance of our scheme with a more conventional formulation in which planning tumor volume are used to account for positional uncertainties.

1. Introduction. Intensity modulated radiation therapy (IMRT) allows tailored doses of radiation to be delivered to a cancer patient via external beams, with the aim of destroying the tumor while sparing surrounding critical structures and normal tissue. After dividing the aperture through which the radiation is delivered into small rectangular regions, thus dividing each beam into so-called "beamlets," treatment planning procedures find the set of beamlet weights that most closely matches the dose requirements prescribed by an oncologist. Optimization techniques have proved to be useful for determining optimal beamlet weights. The increasing power and efficiency of these methods has allowed greater refinement and tuning of treatment plans.

A critical role in treatment planning is played by the *dose matrix*, in which each entry describes the amount of radiation deposited by a unit weight (intensity) from a particular beamlet into a particular voxel in the treatment region. Given the properties of the treatment region and the beamlet source and orientation, the dose matrix is calculated by techniques such as convolution-superposition and Monte Carlo. These are numerical/statistical techniques that give approximate results, and hence are one source of uncertainty in the problem. A second source of uncertainty is change in the positions of the internal tissues of the patient between treatment fractions, relative to the beams. These changes can be due to motion of the tumor and organs or to variations in the patient position on the treatment device between fractions. We focus on dose calculation and positional uncertainties in this paper. We do not deal with other sources of uncertainty, such as imaging errors, physician error, intra-fraction patient motion, registration errors, organ shrinkage, and possible microscopic extensions of the tumor, though our formulation techniques may be useful in handling some of these additional sources of uncertainty as well.

The two most accurate dose calculation methods available today are the convolution/superposition method (Lu, et al. 2005) and the Monte Carlo method (Ma, et al. 1999, Jeraj & Keall 1999, Sempau, et al. 2000). A general discussion of existing dose calculation methods can be found in the report of (Intensity Modulated Radiation Therapy Collaborative Working Group 2001). As described by (Kowalok 2004, Chapter 2), the dose in each voxel calculated from a Monte Carlo simulation is normally distributed. The variance can be approximated by a sample variance arising from the generation procedure.

Most of the existing literature on uncertainty in treatment planning focuses on organ movement and setup errors (Leong 1987, Beckham, et al. 2002, Stapleton, et al. 2005, Baum, et al. 2005, Chu, et al. 2005). In practice, organ movement is often dealt with by the conservative strategy of defining a margin around the ex-

pected tumor location and prescribing a high dose for this expanded region. (See our discussion of planning tumor volume (PTV) and clinical tumor volume (CTV) in subsection 2.1.) Setup errors can be dealt with in a similar manner. Alternative approaches are described in (Leong 1987, Beckham et al. 2002, Stapleton et al. 2005), where the uncertainty in setup is incorporated into the dose matrix calculations. Recent research (Chu et al. 2005, Unkelbach & Oelfke 2004) also has modeled organ movement as uncertainties in the dose matrix. These authors have considered multi-fraction treatment plans and assumed some knowledge of the distribution of organ positions. (van Herk 2004) discusses methods to estimate the probability distribution for organ movement. We discuss modeling organ movement further in subsection 2.3.

The focus of this paper is on modeling and efficient solution of an optimization problem in which the uncertainties from dose calculation and organ/tumor positions are accounted for explicitly and simultaneously in designing the treatment plan. By assuming the probability distributions for both types of uncertainty are known, we can set up and solve a robust optimization problem. Our goal is find solutions that are referred to in the optimization literature (Ben-Tal & Nemirovski 2000) as *almost reliable* solutions, in which the constraints are guaranteed to be satisfied to a certain level of confidence, regardless of how the uncertain data are resolved. This methodology represents a significant advance over techniques in which the uncertainty is ignored, as the beamlet weights calculated from such plans may result in unacceptable underdosing of tumors and overdosing of critical regions when the actual realization of the dose matrix differs appreciably from its expected value.

We also present an algorithmic advance. Like (Chu et al. 2005), we solve a second-order cone programming (SOCP) formulation, but rather than using standard SOCP software (which is based on interior-point algorithms) we use a sequential linear programming (SLP) approach that is well suited to these problems and apparently more efficient.

The remainder of the paper is structured as follows. In section 2, we outline the underlying LP formulation of the treatment planning problem used in this study and show how to obtain a robust formulation from it. Section 3 describes the sequential linear programming approach for solving the robust formulation. Computational results for a nasopharyngeal case are shown in section 4, while section 5 discusses conclusions and possible future research directions.

2. Linear Programming Formulations of IMRT.

2.1. Background and Notation. For planning purposes, the treatment region in the patient's body is divided into regions known as voxels, indexed as $i = 1, 2, \ldots, m$. Each of the apertures through which radiation can be delivered from various angles is divided into rectangular beamlets, which we index as $j = 1, 2, \ldots, n$. The dose of radiation delivered to voxel i by a unit weight (intensity) of beamlet j is denoted by A_{ij}; these quantities can be assembled into a *dose matrix* A of dimension $m \times n$. After assembling the voxel doses x_i into a vector x and the beamlet weights w_j into a vector w, we have the following linear relationship:

$$x = Aw. \tag{2.1}$$

In most approaches to IMRT treatment planning, the beamlet weight vector w is the main variable in the optimization formulation. Clinical data sets used in the computational experiments of (Hou, et al. 2003), (Llacer, et al. 2003), (Alber & Reemtsen 2004), and (Zhang, et al. 2004) typically involve 1000 to 5000 beamlets and up to 100000 voxels.

In treatment planning, the set of voxels $\{1, 2, \ldots, m\}$ is partitioned into one or more target volumes $\mathcal{T}$, one or more critical structures $\mathcal{C}$, and a region of normal tissue $\mathcal{N}$. The target volume is often subdivided into a clinical tumor volume (CTV), which is the region most likely to be tumor, and a planning tumor volume (PTV), which is a larger volume surrounding the CTV. The PTV includes the region into which the tumor may move during treatment, or into which extensions of the tumor may have spread at a sub-detectable level, so it is usually designated to receive some specified level of radiation, comparable to the CTV. The *critical voxels*, denoted by $\mathcal{C}$, are those that are part of a sensitive structure (such as the spinal cord or a vital organ) that we particularly wish to avoid irradiating. *Normal voxels*, denoted by $\mathcal{N}$, are those in the patient's body that fall into neither category. Ideally, normal voxels should receive little radiation, but it is less important to avoid dose to these voxels than to the critical voxels.

Generally, a dosimetrist tries to devise a plan in which the doses to target voxels are close to a specified value and/or within a given range, while constraints of different types are applied to discourage excessive dosage to critical and normal regions.

2.2. Nominal Linear Programming Formulation. In the following linear programming (LP) formulation, we introduce some additional notation. The vectors $x_{\mathcal{T}}^L$ and $x_{\mathcal{T}}^U$ denote lower and upper bounds on the doses to the target voxels, while $d_{\mathcal{T}}$ represents the prescribed dose to these voxels. When the treatment plan is divided into N fractions, the relevant quantities for each fraction are $x_{\mathcal{T}}^L/N$, $x_{\mathcal{T}}^U/N$, and $d_{\mathcal{T}}/N$, respectively. The per-fraction dose to target voxels in excess of $d_{\mathcal{T}}/N$ is denoted by the vector v, while t represents the dose under this level. The vector $d_{\mathcal{E}}$ represents a set of threshold values for doses to critical voxels, such that a penalty is imposed for the per-fraction dose $x_{\mathcal{E}}$ in excess of $d_{\mathcal{E}}/N$ to these voxels. Penalties vectors for doses to normal voxels, over-threshold doses to critical voxels, and overdose and underdose to target voxels, are denoted by $c_{\mathcal{N}}$, $c_{\mathcal{E}}$, c_v, and c_t, respectively. The dose matrix A is divided into three row submatrices $A_{\mathcal{T}} \in \mathbf{R}^{n_{\mathcal{T}} \times n}$, $A_{\mathcal{N}} \in \mathbf{R}^{n_{\mathcal{N}} \times n}$, and $A_{\mathcal{C}} \in \mathbf{R}^{n_{\mathcal{C}} \times n}$, corresponding to the treatment volumes $\mathcal{T}$, $\mathcal{N}$, and $\mathcal{C}$, respectively.

The linear programming problem for calculating the optimal weight vector for a single fraction can be written as follows:

$$\min_{w,s,t,x_{\mathcal{E}}} \; c_{\mathcal{N}}^T A_{\mathcal{N}} w + c_v^T v + c_t^T t + c_{\mathcal{E}}^T x_{\mathcal{E}} \tag{2.2a}$$

$$\text{subject to} \quad A_{\mathcal{T}} w \leq \frac{x_{\mathcal{T}}^U}{N} \tag{2.2b}$$

$$A_{\mathcal{T}} w \geq \frac{x_{\mathcal{T}}^L}{N} \tag{2.2c}$$

$$A_{\mathcal{C}} w - x_{\mathcal{E}} \leq \frac{d_{\mathcal{E}}}{N} \tag{2.2d}$$

$$v - t = A_{\mathcal{T}} w - \frac{d_{\mathcal{T}}}{N} \tag{2.2e}$$

$$w, x_{\mathcal{E}}, v, t \geq 0. \tag{2.2f}$$

The formulation is quite flexible, as the bounds and penalties can vary from voxel to voxel and, in particular, the penalty for overdosing a voxel can be different from the penalty for underdose. The threshold $d_{\mathcal{E}}$ and penalty $c_{\mathcal{E}}$ for excess doses to the critical voxels can be manipulated to ensure that dose-volume (DV) constraints are satisfied.

The first published usage of LPs in radiation treatment planning was the work by (Klepper 1966) and (Bahr, et al. 1968). Later works involving LP include (Hodes 1974), (Morrill, et al. 1990), (Langer, et al. 1990), and (Rosen, et al. 1991). The linear models vary in their definition of both objectives and constraints. For more recent work, see the surveys by (Shepard, et al. 1999) and (Reemtsen & Alber 2004) and the references therein.

We refer to the formulation above as a "nominal" LP formulation as it assumes that the dose matrix A is known with certainty, so that the constraints in the model can correspondingly be enforced with certainty. We now describe a formulation that accounts for the fact that A is actually uncertain, and uses our knowledge of this uncertainty to enforce the constraints in a probabilistic fashion.

2.3. Robust Linear Programming Formulation. We consider a model that incorporates two types of uncertainty in the dose matrix.

- Uncertainties due to Monte Carlo dose matrix computations. To model this uncertainty, we treat the dose to each voxel from each beamlet as a random variable, normally distributed with a known variance. Moreover, we assume that these random variables are independent.
- Uncertainties due to organ motion and patient positioning errors. These can be modeled by defining a number of scenarios for the positions of the organs relative to the beam and assigning a probability to each scenario. For instance, one scenario could be the "nominal" position identified by imaging, while other scenarios could be assigned by shifting the organs by various amounts in various directions. Each scenario is defined by a different dose matrix. Unless the number of scenarios is very large, there are fewer "degrees of freedom" in this model than in the dose-calculation model. It captures the correlation between doses to adjacent voxels when organ motion and patient positioning errors occur. In general, higher uncertainty will occur in voxels at the boundary between low-dose and high-dose regions, as such voxels will move into high-dose regions in some scenarios and into low-dose regions in other scenarios.

In accounting for the second type of error, we assume as in (Chu et al. 2005) that the dose is delivered in multiple fractions, with the scenario that occurs at each fraction being independent of the scenarios that occur on other fractions. Moreover, we assume that the number of fractions is large enough that the total dose is approximately normally distributed (using the central limit theorem). Thus, we need calculate only its mean and variance to characterize the dose distribution fully. This independence assumption is not fully satisfactory, as some systematic organ motion may occur as the patient geometry changes during the treatment period. Nevertheless, the model is reasonable given the information available at the initial time. Extensions to our formulation techniques could be devised to account for periodic adjustment of the plan during the treatment period, but these are beyond the scope of this paper.

Our model accounts for these two types of errors by assuming that we have dose matrices $A^1, A^2, \ldots, A^L$ associated with the L organ-position scenarios, with associated probabilities $p_1, p_2, \ldots, p_L$. However, in contrast to (Chu et al. 2005), we assume that each A^k is a (matrix) random variable rather than a fixed matrix. Specifically, we assume that each component A^k_{ij} is a normally distributed random variable with expected value $\tilde{A}^k_{ij}$ and variance $(\sigma^k_{ij})^2$, and that these random variables are independent of each other.

To obtain a formulation that takes account of these uncertainties, we need to calculate the expected value and variance of the dose to voxel i from a given weight vector w. For the expected value, we have

$$\tilde{d}_i = E(d_i) = E\left(\sum_{k=1}^{L} p_k A_{i\cdot}^k w\right) = \sum_{k=1}^{L} p_k E(A_{i\cdot}^k) w = \sum_{k=1}^{L} p_k \tilde{A}_{i\cdot}^k w = \hat{A}_{i\cdot} w, \tag{2.3}$$

where $\hat{A} = \sum_{k=1}^{L} p_k \tilde{A}^k$. Derivation of the variance of the dose d_i is shown in Appendix A. It has the form

$$\mathrm{Var}(d_i) = \|T_i w\|_2^2, \tag{2.4}$$

where the matrix T_i is defined as

$$T_i = \begin{bmatrix} \Sigma_i \\ R_i \tilde{A}_i \end{bmatrix},$$

where

$$\tilde{A}_i = \begin{bmatrix} \tilde{A}_{i\cdot}^1 \\ \tilde{A}_{i\cdot}^2 \\ \vdots \\ \tilde{A}_{i\cdot}^L \end{bmatrix}, \tag{2.5a}$$

$$\Sigma_i = \mathrm{diag}(\tilde{\sigma}_{i1}, \tilde{\sigma}_{i2}, \ldots, \tilde{\sigma}_{in}) \ \text{ with } \ \tilde{\sigma}_{ij}^2 = \sum_{k=1}^{L} p_k (\sigma_{ij}^k)^2, \tag{2.5b}$$

$$R_i = P^{1/2}[I - ep^T], \tag{2.5c}$$

with $p = (p_1, p_2, \ldots, p_L)^T$, $e = (1, 1, \ldots, 1)^T$, and $P = \mathrm{diag}(p_1, p_2, \ldots, p_L)$.

Applying the central limit theorem to the total dose to voxel i, aggregated over N fractions, we obtain an expected value of $N\hat{A}_{i\cdot} w$, and a variance of $N \|T_i w\|_2^2$. Hence, our robust reformulation of a constraint of the form $A_{i\cdot} w \le b_i/N$ in the nominal LP (2.2) is as follows:

$$N\hat{A}_{i\cdot} w + \Phi^{-1}(1-\delta)\sqrt{N}\, \|T_i w\|_2 \le N b_i,$$

where δ is a lower bound on probability that the constraint will be violated and Φ is the standard normal cumulative distribution function. Introducing the notation $z_{1-\delta} \equiv \Phi^{-1}(1-\delta)$ and rearranging, we rewrite this inequality as follows:

$$\hat{A}_{i\cdot} w + \frac{z_{1-\delta}}{\sqrt{N}} \|T_i w\|_2 \le b_i. \tag{2.6}$$

Using the notation above we can now write the robust formulation of the LP in (2.2) as follows:

$$\min_{w, x_\mathcal{E}, v, t} \ c_\mathcal{N}^T A_\mathcal{N} w + c_v^T v + c_t^T t + c_\mathcal{E}^T x_\mathcal{E} \tag{2.7a}$$

$$\text{subject to } \ c_U^i(w) \le 0, \ \ i \in \mathcal{T} \tag{2.7b}$$

$$c_L^i(w) \le 0, \ \ i \in \mathcal{T} \tag{2.7c}$$

$$c_\mathcal{C}^i(w) - (x_\mathcal{E})_i \le 0, \ \ i \in \mathcal{C} \tag{2.7d}$$

$$\hat{A}_\mathcal{T} w - v + t - \frac{d_\mathcal{T}}{N} = 0, \tag{2.7e}$$

$$w, x_\mathcal{E}, v, t \ge 0, \tag{2.7f}$$

where

$$c_L^i(w) = -\sum_{j=1}^{n} \hat{A}_{ij} w_j + \frac{z_{1-\delta}}{\sqrt{N}} \|T_i w\|_2 + \frac{(x_{\mathcal{T}}^L)_i}{N},$$

$$c_U^i(w) = \sum_{j=1}^{n} \hat{A}_{ij} w_j + \frac{z_{1-\delta}}{\sqrt{N}} \|T_i w\|_2 - \frac{(x_{\mathcal{T}}^U)_i}{N},$$

$$c_{\mathcal{C}}^i(w) = \sum_{j=1}^{n} \hat{A}_{ij} w_j + \frac{z_{1-\delta}}{\sqrt{N}} \|T_i w\|_2 - \frac{(d_{\mathcal{E}})_i}{N}$$

Because of the presence of the terms $\|T_i w\|_2$, the problem (2.7) is a second-order cone program (SOCP). This is a special class of convex optimization problem that has received a good deal of attention in recent years, as it is a useful way to express robust optimization problems and it can be solved efficiently by interior-point methods.

The constraints (2.7c) and (2.7b) ensure that for any given voxel $i \in \mathcal{T}$, the probability of violation of the constraint in question, given some set of realizations of the random dose matrices over the N-fraction treatment plan, is at most δ. Note that if δ is chosen to be very close to zero (so that $z_{1-\delta}$ is large), the constraints (2.7c) and (2.7b) might represent a significant tightening over the nominal case, possibly leading to an infeasible formulation. That is, there might be no weight vector w that satisfies all these constraints to the desired level of certainty.

For each critical voxel $i \in \mathcal{C}$, constraint (2.7d) ensures that with probability at least $1 - \delta$, the dose to voxel i is at most $(d_{\mathcal{E}})_i/N + (x_{\mathcal{E}})_i$. We continue use expected dose in the constraint (2.7e) which gauges the discrepancy between the delivered dose and the prescribed dose vector $d_{\mathcal{T}}$. It is not crucial to take account of the uncertainty in handling this aspect of the model, which is less critical than avoiding hot and cold spots in the target and overdose to the critical structures.

The constraints (2.7c) and (2.7b) in the robust formulation result in a tighter allowable dose interval when compared to the nominal formulation in (2.2), because of the addition of the standard deviation term to the bounds. If the interval between $x_{\mathcal{T}}^L$ and $x_{\mathcal{T}}^U$ is small and the allowable violation probability δ is small, there may be no w satisfying both constraints (2.7b) and (2.7c) simultaneously, in which case the robust linear program (2.7) will be infeasible.

3. Sequential Linear Programming Algorithm for Robust LP. Although good algorithms and software exist for solving second-order cone programs such as (2.7), they tend to require more computational resources than LPs of similar size and structure. (Chu et al. 2005) reported run times of 3.3 hours for a formulation similar to (2.7), while the corresponding nominal LP required about 17 minutes.

However, the particular problem (2.7) has a property that makes it unnecessary to use a general SOCP code. For data sets of interest, the terms $T_i w$ are nonzero for all $i \in \mathcal{T} \cup \mathcal{C}$. Since the constraint functions c_U^i, c_L^i, and $c_{\mathcal{C}}^i$ are nonsmooth only at points for which $T_i w = 0$, we can assume smoothness and apply a general algorithm for smooth nonlinear programming. Another property that ensures the success of this approach is that when the terms $\|T_i w\|_2$ are omitted, the nominal LP that remains has a finite solution. We now describe a *sequential linear programming (SLP)* approach for solving (2.7) which is extremely simple and is, we believe, applicable to other robust LP formulations with the two properties mentioned above.

We start by writing the nonsmooth penalty formulation of (2.7), in which all constraints except the nonnegativity $w \geq 0$ are replaced by nonsmooth penalty terms

in the objective:

$$\min_{w\geq 0} \tau(w) \equiv c_{\mathcal{N}}^T A_{\mathcal{N}} w + c_v^T \left(\hat{A}_{\mathcal{T}} w - \frac{d_{\mathcal{T}}}{N}\right)_+ + c_t^T \left(\hat{A}_{\mathcal{T}} w - \frac{d_{\mathcal{T}}}{N}\right)_- \\ + c_{\mathcal{E}}^T c_{\mathcal{C}}(w)_+ + \nu(||c_L(w)_+||_1 + ||c_U(w)_+||_1) \tag{3.1}$$

where $(x)_+ = \max(0, x)$ and $(x)_- = \max(0, -x)$. The value of the penalty parameter ν is chosen large enough to ensure that the constraints (2.7b) and (2.7c) are satisfied when the problem (2.7) is feasible. (If the underlying (2.7) is infeasible, then the minimization of the penalty function (3.1) for large ν will yield the w that comes closest to satisfying the constraints in an ℓ_1 sense.)

To ease description of the methodology used to solve this penalty formulation, let us note that this problem has the form

$$\min_{w\geq 0} p(h(w)),$$

where the vector function $h(w)$ contains linear and SOCP-type terms (so is smooth for w of interest) while $p(\cdot)$ is a polyhedral convex function of the general form

$$p(h) = \max_{j=1,2,\ldots,J} \phi_j^T h + \beta_j, \tag{3.2}$$

where ϕ_j, β_j, $j = 1, 2, \ldots, J$ are constant vectors and scalars, respectively. Note that $p(h)$ is piecewise linear as a function of h and that $p(h(w))$ is piecewise smooth as a function of w.

In the SLP approach, we obtain a step Δw from the current solution estimate w by solving the following subproblem:

$$\min_{\Delta w} p(h(w) + \nabla h(w)^T \Delta w) \text{ subject to } w + \Delta w \geq 0, \ \ \|\Delta w\|_\infty \leq \Delta. \tag{3.3}$$

The vector function h is replaced by a linear approximation around the current estimate w, which we can expect to be reasonably close to $p(h(w + \Delta w))$ provided that Δw is not too large. The last constraint is a trust region constraint, which bounds how far we can move away from the current iterate w at this iteration. Essentially, the bound Δ defines the region on which we "trust" the approximate objective $p(h(w) + \nabla h(w)^T \Delta w)$ to be close to the true objective $p(h(w + \Delta w))$. By making use of the definition (3.2) of $p(h)$, the subproblem (3.3) can be formulated as a linear program. When applied to (3.1), and after conversion to an LP, the subproblem (3.3) has the following form.

$$\min_{\Delta w, v, t, x_\mathcal{E}} \; m(\Delta w) \equiv {c_\mathcal{N}}^T A_\mathcal{N}(w + \Delta w) + c_v^T v + c_t^T t + c_\mathcal{E}^T x_\mathcal{E} + (\nu e)^T(\rho_L + \rho_U) \;\text{ s.t.}$$

$$-(\rho_L)_i + \sum_{j=1}^{n} \left(- \hat{A}_{ij} + \frac{z_{1-\delta}}{\sqrt{N}} \frac{(T_i w)^T (T_i)_{\cdot j}}{||T_i w||_2} \right) \Delta w_j + c_L^i(w) \le 0, \quad i \in \mathcal{T} \tag{3.4b}$$

$$-(\rho_U)_i + \sum_{j=1}^{n} \left(\hat{A}_{ij} + \frac{z_{1-\delta}}{\sqrt{N}} \frac{(T_i w)^T (T_i)_{\cdot j}}{||T_i w||_2} \right) \Delta w_j + c_U^i(w) \le 0, \quad i \in \mathcal{T} \tag{3.4c}$$

$$\hat{A}_{i\cdot} \Delta w + \frac{z_{1-\delta}}{\sqrt{N}} \frac{(T_i w)^T (T_i \Delta w)}{||T_i||_2} - (x_\mathcal{E})_i + c_{OAR}^i(w) \le 0, \quad i \in \mathcal{C} \tag{3.4d}$$

$$\hat{A}_T(w + \Delta w) - v + t - \frac{d_\mathcal{T}}{N} = 0, \tag{3.4e}$$

$$\Delta w + w \ge 0, \tag{3.4f}$$

$$||\Delta w||_\infty \le \Delta, \tag{3.4g}$$

$$\rho_L, \rho_U, v, t, x_\mathcal{E} \ge 0, \tag{3.4h}$$

where $e = (1, 1, \ldots, 1)^T$ as before.

The complete algorithm is shown as Algorithm 1. At each iteration, we compute a quantity s for purposes of a termination test. This quantity consists of a predicted relative decrease in the objective over the current step $((\tau(w^k) - m(w^k + \Delta w^k))/\tau(w^k))$ scaled by the current trust-region radius Δ_k, if the latter quantity is greater than 1. We note that the termination test is not critical. Later iterations of Algorithm 1 are relatively inexpensive, due to warm-start information being available from earlier iterations, so it makes little matter if the algorithm is allowed to run for some extra iterations.

Set $k \leftarrow 0$; choose initial trust region parameter Δ_0 and stopping parameter η;
Solve for Δw in (3.4) by omitting terms including w; Set $w^0 \leftarrow \Delta w$;
while $k = 0$ *or (*$k < 50$ and $s > \eta$*)* **do**
  Solve the subproblem (3.4) with $w = w^k$ to obtain Δw^k;
  Set $\tilde{w} \leftarrow w^k + \Delta w^k$;
  if $\tau(\tilde{w}) < \tau(w^k)$ **then**
    Set $w^{k+1} \leftarrow \tilde{w}$;
    $\Delta_{k+1} \leftarrow 1.5\Delta_k$;
  else
    $\Delta_{k+1} = .5\|\Delta w^k\|_\infty$;
  end
  Update $s = \frac{\tau(w^k) - m(\Delta w^k)}{\tau(w^k)\min(1, \Delta_k)}$;
  Set $k \leftarrow k + 1$;
end

Algorithm 1: Sequential Linear Programming

4. Computational Results. We report computational results obtained on a medium-scale clinical case, a nasopharyngeal tumor. This data set, also used by (Wu 2002), is a problem in which treatment is carried out in two axial planes through

the patient. The 24000 voxels are divided into five regions: CTV and PTV; parotids and spinal cord (the critical structures); and the normal region. There are a total of 1,989 beamlets, consisting of 39 beamlets from each of 51 angles.

For modeling purposes we used GAMS (Brooke, et al. 1998) with the linear programming solver CPLEX version 9.0. Following our earlier work on efficient LP formulations of treatment planning problems (Ólafsson & Wright 2006), we solve the primal formulation (2.2) of the nominal LP problem using dual simplex with steepest edge pricing. Experiments were performed on a computer with a 3.0 GHz Intel(R) Pentium(R) four-processor PC running CentOS 4. Cache size was 2 MB and total memory was 2 GB. We compare results obtained with the nominal planning procedure of section 2.2 and the robust planning methodology of section 2.3.

The costs, bounds, and thresholds in the nominal problem (2.2) were set as follows. For the costs, we have $c_{\mathcal{N}} = 3e$ and $c_{\mathcal{E}} = 10e$. The PTV and CTV components of c_v are set to 10 and 20, respectively, while the corresponding settings for c_t are 15 and 15. For the prescribed dose vector $d_{\mathcal{T}}$, the CTV components were set to 70 while the PTV components were set to 60. (The unit of dose is Gray (Gy).) Lower bound components $x_{\mathcal{T}}^L$ were set to 55 for PTV and 66 for CTV, while upper bound components are 74 for both PTV and CTV. The critical structure threshold $d_{\mathcal{E}}$ is set to 5 Gy for the spinal cord and 25 Gy for the parotids. The number of fractions used is $N = 40$.

We took the output of the Monte Carlo dose calculation engine to be the nominal doses $\tilde{A}_{ij}$ used in (2.2). In the robust planning setup, we took these values to be the expected doses to each voxel, while the dose calculation error variances were taken to be the square of

$$\sigma_{ij} = .02 \max_{k \in \mathcal{T}} \tilde{A}_{kj}, \quad \forall (i,j) \tag{4.1}$$

where $\mathcal{T}$ is the target region.

For robust planning, we used five scenarios for positional uncertainty, as shown in Table 4.1. These included the original specified positions of the organs and four other positional scenarios obtained by shifting the target and critical regions by 3.6 mm in the left, right, anterior, and posterior directions. (The shift happens to be twice the length of the side of a voxel, but non-integer multiples and non-axial directions of shift can be used.) The probability of the expected position is chosen to be approximately twice the probability of each of the shifted scenarios.

TABLE 4.1
The probability distribution for the given shift in considered directions.

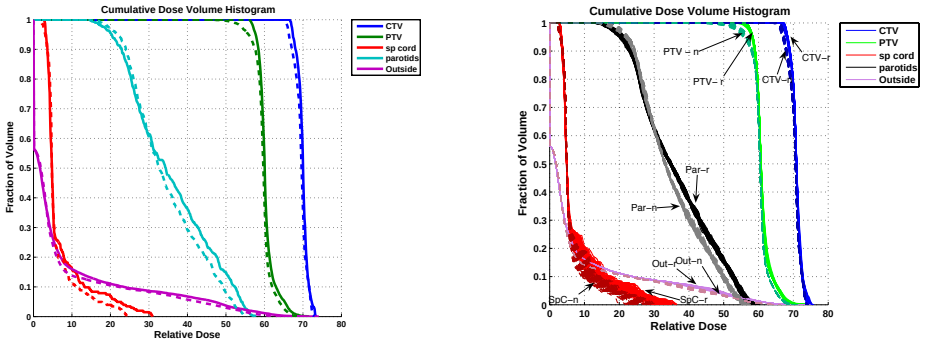
Scenario	Direction	Shift (mm)	Probability
1	None	0.0	0.32
2	Left	3.6	0.17
3	Right	3.6	0.17
4	Anterior	3.6	0.17
5	Posterior	3.6	0.17

In Algorithm 1 the trust region parameter Δ is initially set to 30 and the stopping parameter η to 10^{-3}. In the subproblem (3.4), we define $\delta = 0.05$, so that $z_{1-\delta} = 1.645$. The penalty parameter is $\nu = 1000$.

Having obtained an optimal weight vector w from either the nominal and robust planning procedures, we can evaluate its efficacy by simulating a dose distribution

obtained with w over a 40-fraction treatment period, in which the actual dose matrix on each fraction is chosen at random according to the positional distribution of table 4.1 and from a normal sampling with mean $\bar{A}_{ij}$ and variance σ_{ij}^2 at each entry (i, j). Intensity maps and DVH plots can be obtained for each such simulation, and the number of violations of upper and lower bounds can be tabulated. We can also approximate these simulations with *expected* results in which the proportion of each scenario is taken to be exactly the fraction appearing in Table 4.1 and the variance σ_{ij} is reset to zero.

4.1. Using PTV in both Normal and Robust Plans. In our data set, the specified PTV is quite extensive, suggesting that it was defined clinically to account not just for possible motion of the CTV but also widespread extensions of the tumor. Hence, we retained the PTV in the robust plan, shifting it along with the CTV and critical regions to obtain each scenario in Table 4.1. We would expect the robust plan to have the advantage that the PTV and CTV voxels would be much more likely to receive doses within the specified intervals. A possible price of this greater certainty of avoiding cold spots is that the normal and critical regions may receive more dose.



(a) (b)

FIG. 4.1. *(a) DVH for the expected position of organs. The solid line is the robust case and the dotted line is the nominal solution. (b)* 100 *cases of simulated dose for robust (solid) and nominal (dotted) plans.*

Figure 4.1(a) shows the expected DVH plot for the robust plan as solid lines and for the nominal plan as dotted lines. Figure 4.1(b) contains the corresponding plots for 100 simulations of the treatment plan, for both nominal and robust weight vectors.

The difference between the plans is minor in most respect, except that the nominal plan is more likely to violate the lower bounds on the PTV and CTV voxels, and the robust plan delivers somewhat more dose to the spinal cord. Table 4.2 shows the minimum, maximum, and average (taken over 100 simulations of the treatment plan) of both the minimum and maximum dose to the CTV and the PTV. It also shows the percentage of violated lower bound constraints in each case. The average minimum dose to the CTV is above the defined lower bound of 66 Gy for the robust solution while it is slightly below it for the nominal solution. For the PTV, the mean minimum dose for the robust solution is below 55 Gy, yet only 0.04% of constraints in the PTV are violated on average. (If we had chosen δ smaller than 0.05 the number of such violations would most likely have been decreased to zero.) The nominal solution on

the other hand violates on average 3.11% of the PTV lower-bound constraints and, in the worst case, 2.26% of the CTV lower-bound constraints.

TABLE 4.2
These results are obtained from the 100 *simulated cases. The line denoted by* Violation *represents the percentage of violated lower bound constraints.*

		Robust solution			Nominal solution		
		Min	Max	Mean	Min	Max	Mean
CTV	Minimum dose	66.17	67.39	66.82	64.27	65.97	65.43
	Maximum dose	73.64	75.80	74.72	73.65	74.35	73.89
	Violation	0.00%	0.00%	0.00%	0.11%	2.26%	0.58%
PTV	Minimum dose	45.20	56.15	54.24	36.22	48.76	44.62
	Maximum dose	70.80	75.59	72.31	69.48	71.21	70.14
	Violation	0.00%	0.42%	0.04%	1.98%	4.33%	3.11%

Computational results for Algorithm 1, which is used to solve the robust formulation, are shown in Table 4.3. The second column shows the trust region bound Δ. (That Δ increases in each iteration means that the step was accepted in all three iterations.) The third column shows the objective value $m(\Delta w)$ obtained from solving each LP subproblem (3.4). The fourth column shows the penalty function value (3.1), while the fifth columns shows the value at each iteration of the stopping criterion. The last two columns show the maximum violation of the current solution of the constraints (2.7c) and (2.7b). Even though there is a slight violation at the computed solution, it is quite reliable, as we discuss below.

The algorithm stopped after three iterations and required slightly under 10 minutes. (We do not include in this figure the approximately 4 minutes to perform preliminary setup and storage of the data, an operation that needs to be performed in both the robust and nominal cases.) The total number of LPs solved is 4, since one LP must be solved to obtain the initial point w^0. Each iteration takes less time to set up and solve than the one before it, as warm-start information is carried over from one LP to the next. By comparison, the nominal problem required slightly under 7 minutes to solve, not including the 4-minute setup time.

TABLE 4.3
Computational results for algorithm 1.

Iter	Δ	$m(\Delta w^k)$	$\tau(w^k)$	s	$\|\|c_L(w^k)_+\|\|_\infty$	$\|\|c_U(w^k)_+\|\|_\infty$
0	30.0	15332.04	21692.44	0.293	0.16	0.03
1	45.0	15349.42	15385.37	0.002	0.12	0.00
2	67.5	15350.12	15352.66	0.000	0.13	0.00
3			15350.92		0.13	0.00

4.2. Using Multiple Scenarios to Replace PTV. We performed a second set of tests in which we discarded the PTV supplied with the clinical data set, and defined a new one that consisted of a "collar" of four voxels in width (approximately 7.2 mm) around the specified CTV, to account for positional uncertainty of the CTV. The nominal plan was obtained by solving this case with the parameter settings above, except that the lower bound on the (newly defined) PTV was increased from 55 to 60 Gy and the prescribed value was increased from 60 to 65 Gy. (We wish to deliver a very similar dose to these voxels as to the CTV voxels, as there is a good chance that

the tumor may well occupy some of these voxels.) We increased the value of c_ε from $10e$ to $15e$ to further discourage dose to the critical structures. Finally, to obtain a more even intensity spread in the normal region, $c_\mathcal{N}$ is increased to $8e$ for the normal voxels in rows 35 to 68. This area surrounds the CTV, PTV and the parotids. In the robust planning procedure, we discarded the PTV altogether, instead using the potentially more precise positional uncertainty information contained in table 4.1. To decrease the probability that any CTV voxels drop below the lower bound, we used the more conservative value $\delta = 0.02$, thus $z_{1-\delta} = 2.054$.

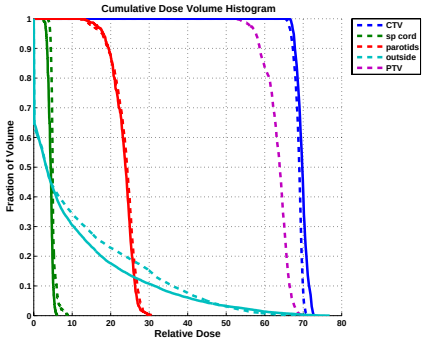
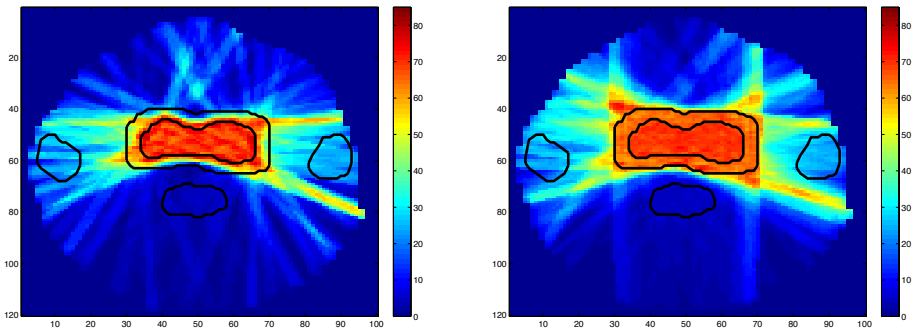


FIG. 4.2. *Dose volume histogram for the CTV case. The solid lines are for the robust formulation in which the PTV does not appear. The dotted lines are for the nominal solution.*

Figure 4.2 shows the expected-dose DVH plot for the robust and nominal plans. Since there is no PTV in the robust plan, only four solid lines are plotted. There is little difference between the two plans in coverage of the CTV, though for the nominal plan the minimum dose to the CTV is, at 65.50 Gy, a little below the specified minimum of 66 Gy. Since the robust plan no longer needs to cover the PTV with high dose, the doses to the critical structures and the normal tissue are diminished slightly in general. (For purposes of this figure, the normal voxels are taken to be just those for the nominal plan, though in fact the robust planning procedure contains additional normal voxels arising from the absence of the PTV.)



(a) (b)

FIG. 4.3. *Intensity plot for (a) the robust plan and (b) the nominal plan.*

Figure 4.3 shows intensity plots for the robust and nominal plans. The organs are plotted in their expected (unshifted) positions, with the CTV being the innermost outlined shape and the PTV surrounding it. The parotids appear to left and right, while the spinal cord appears below the CTV/PTV. As mentioned above, settings were tested by comparing results for expected nominal plans like the one in figure 4.3(b). The smearing outside out the boundary of the CTV in the robust plan results from the expected movement, but it targets these voxels more precisely than the robust plan, which delivers large doses throughout the PTV region. The unnecessary smearing, at the "corners" of the CTV, could probably be reduced by different settings of the weights.

Computational results for algorithm 1 are shown in table 4.4, which has the same format as table 4.3. The algorithm requires 13 iterations to meet the stopping criteria. As can be seen in the fourth column the improvement made is minimal even after the fourth iteration, so for this case a looser stopping criteria would have been more appropriate. However, the cost of solving the LP subproblem and the fourth and subsequent iterations is low, so not much is lost by continuing the iterations beyond the point of necessity. The robust problem took about 2 minutes to set up and then solved in 3.8 minutes. The nominal problem took about 2.5 minutes to set up and solved in 46 seconds. This case solves much faster than the case studied in subsection 4.1, since there are only 3751 rows in this problem compared to 16489 before.

TABLE 4.4
Computational results for the CTV case.

Iter	Δ	$m(\Delta w^k)$	$\tau(w^k)$	s	$\|\|c_L(w^k)_+\|\|_\infty$	$\|\|c_U(w^k)_+\|\|_\infty$
0	30.0	33487.14	44083.76	0.240	0.14	0.09
1	45.0	33623.72	34035.12	0.012	0.04	0.01
2	67.5	33618.42	33785.26	0.005	0.06	0.01
3	101.2	33626.18	33760.75	0.004	0.06	0.01
4	7.6	33633.13	33760.75	0.004	0.06	0.01
5	11.4	33627.36	33732.69	0.003	0.07	0.00
6	5.1	33632.80	33732.69	0.003	0.07	0.00
7	7.7	33637.37	33723.81	0.003	0.05	0.00
8	3.8	33640.89	33723.81	0.002	0.05	0.00
9	5.8	33640.43	33709.30	0.002	0.06	0.00
10	2.9	33644.03	33709.30	0.002	0.06	0.00
11	4.3	33644.60	33687.86	0.001	0.05	0.00
12	2.2	33646.87	33687.86	0.001	0.05	0.00
13	3.2	33647.17	33678.15	0.001	0.06	0.01
14			33678.15		0.06	0.01

5. Conclusions and Future Work. We have shown in this paper how two important sources of uncertainty in IMRT treatment planning problems—errors in dose matrix calculations and positional uncertainties—can be accounted for rigorously and tractably in a robust optimization formulation. We point out that the incorporation of dose calculation uncertainty into the model allows us to use quite inexact dose matrices (thus spending less time in the Monte Carlo dose calculation engine), yet achieve satisfaction of the important constraints in the model to a specified degree of certainty. We have also presented an algorithm for solving the robust formulation

whose solve time is a modest multiple of that requires by the corresponding linear programming formulation that ignores the uncertainty.

Future work will involved introducing robust formulation methodology to other, more elaborate formulations of the IMRT planning problem, and testing more extensively on a wider range of data sets. From an algorithmic point of view, we also wish to identify other robust linear programs from other applications for which Algorithm 1 is a more efficient solution approach than general SOCP codes, and to develop this algorithm further.

Appendix A. Dose Variance.

We describe here how the variance of the dose to a single voxel is calculated under our assumptions. For purposes of this discussion we denote the number of elements of the weight vector by n.

We have

$$
\begin{aligned}
\mathrm{Var}(d_i) &= \sum_{k=1}^{L} p_k E\left\{\left[A_{i\cdot}^k w - E(d_i)\right]^2\right\} \\
&= \sum_{k=1}^{L} p_k E\left\{\left[\sum_{j=1}^{n} A_{ij}^k w_j - \sum_{l=1}^{L} p_l \sum_{j=1}^{n} \tilde{A}_{ij}^l w_j\right]^2\right\} \\
&= \sum_{k=1}^{L} p_k E\left\{\left[\sum_{j=1}^{n} A_{ij}^k w_j - \sum_{j=1}^{n} (\sum_{l=1}^{L} p_l \tilde{A}_{ij}^l) w_j\right]^2\right\} \\
&= \sum_{k=1}^{L} p_k E\left\{\left[\sum_{j=1}^{n} \left(A_{ij}^k - \sum_{l=1}^{L} p_l \tilde{A}_{ij}^l\right) w_j\right]^2\right\} \\
&= \sum_{k=1}^{L} p_k E\left\{\sum_{j=1}^{n}\sum_{\hat{j}=1}^{n} \left(A_{ij}^k - \sum_{l=1}^{L} p_l \tilde{A}_{ij}^l\right)\left(A_{i\hat{j}}^k - \sum_{l=1}^{L} p_l \tilde{A}_{i\hat{j}}^l\right) w_j w_{\hat{j}}\right\} \\
&= \sum_{k=1}^{L} p_k E\left\{\sum_{j=1}^{n}\sum_{\hat{j}=1}^{n} \left(A_{ij}^k - \tilde{A}_{ij}^k + \tilde{A}_{ij}^k - \sum_{l=1}^{L} p_l \tilde{A}_{ij}^l\right)\right. \\
&\qquad\qquad \left.\left(A_{i\hat{j}}^k - \tilde{A}_{i\hat{j}}^k + \tilde{A}_{i\hat{j}}^k - \sum_{l=1}^{L} p_l \tilde{A}_{i\hat{j}}^l\right) w_j w_{\hat{j}}\right\} \\
&= \sum_{k=1}^{L} p_k E\left\{\sum_{j=1}^{n}\sum_{\hat{j}=1}^{n} \left(A_{ij}^k - \tilde{A}_{ij}^k\right)\left(A_{i\hat{j}}^k - \tilde{A}_{i\hat{j}}^k\right) w_j w_{\hat{j}}\right\} \\
&\qquad + \sum_{k=1}^{L} p_k \sum_{j=1}^{n}\sum_{\hat{j}=1}^{n} \left(\tilde{A}_{ij}^k - \sum_{l=1}^{L} p_l \tilde{A}_{ij}^l\right)\left(\tilde{A}_{i\hat{j}}^k - \sum_{l=1}^{L} p_l \tilde{A}_{i\hat{j}}^l\right) w_j w_{\hat{j}},
\end{aligned}
$$

where in the final equality we were able to eliminate the cross-terms because of $E(A_{ij}^k - \tilde{A}_{ij}^k) = 0$. Note that there is no expectation operator in the second term in the final line, as this is not a random variable.

For the first term in the final expression above, we can use independence of the

random variables A_{ij}^k, $j = 1, 2, \ldots, n$, and knowledge of the variances σ_{ij}^k to write

$$\sum_{k=1}^{n} p_k E\left\{\sum_{j=1}^{n}\sum_{\hat{j}=1}^{n}\left(A_{ij}^k - \tilde{A}_{ij}^k\right)\left(A_{i\hat{j}}^k - \tilde{A}_{i\hat{j}}^k\right) w_j w_{\hat{j}}\right\} = \sum_{k=1}^{L} p_k \sum_{j=1}^{n} (\sigma_{ij}^k)^2 w_j^2.$$

By defining $\tilde{\sigma}_{ij}$ as in (2.5b), we can write this term concisely as

$$\sum_{j=1}^{n} \tilde{\sigma}_{ij}^2 w_j^2. \tag{A.1}$$

For the second term in the variance expression, we define the matrix $\tilde{A}_i$ as in (2.5a), similarly to (Chu et al. 2005). We can then write

$$\begin{aligned}\sum_{k=1}^{L} p_k \sum_{j=1}^{n}\sum_{\hat{j}=1}^{n}\left(\tilde{A}_{ij}^k - \sum_{l=1}^{L} p_l \tilde{A}_{ij}^l\right)\left(\tilde{A}_{i\hat{j}}^k - \sum_{l=1}^{L} p_l \tilde{A}_{i\hat{j}}^l\right) w_j w_{\hat{j}} \\ = \left[\tilde{A}_i w - e p^T \tilde{A}_i w\right]^T P \left[\tilde{A}_i w - e p^T \tilde{A}_i w\right],\end{aligned}$$

where $p = (p_1, p_2, \ldots, p_L)^T$ and $P = \mathrm{diag}(p_1, p_2, \ldots, p_L)$. Hence, by defining $R_i = P^{1/2}[I - ep^T]$, we can write this term as

$$\|R_i \tilde{A}_i w\|_2^2. \tag{A.2}$$

By conjoining (A.1) with (A.2), we can express the variance of d_i as

$$\mathrm{Var}(d_i) = \sum_{j=1}^{n} \tilde{\sigma}_{ij}^2 w_j^2 + \|R_i \tilde{A}_i w\|_2^2 = \left\| \begin{bmatrix} \Sigma_i \\ R_i \tilde{A}_i \end{bmatrix} w \right\|_2^2, \tag{A.3}$$

where Σ_i is defined in (2.5b).

REFERENCES

M. Alber & R. Reemtsen (2004). 'Intensity modulated radiation therapy planning by use of a Lagrangian Barrier-Penalty algorithm'. To appear in Optimization Methods and Software.

G. Bahr, et al. (1968). 'The method of linear programming applied to radiation treatment planning'. *Radiology* **91**:686–693.

C. Baum, et al. (2005). 'Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities'. Technical report, University of Tübingen. To appear in *Radiotherapy and Oncology*.

W. A. Beckham, et al. (2002). 'A fluence-convolution method to calculate radiation therapy dose distributions that incorporate random set-up error'. *Phys. Med. Biol.* **47**:3465–3473.

A. Ben-Tal & A. Nemirovski (2000). 'Robust solutions of Linear Programming problems contaminated with uncertain data'. *Mathematical Programming, Series A* **88**:411–424.

A. Brooke, et al. (1998). *GAMS: A User's Guide*. GAMS Development Corporation, Washington DC. `http://www.gams.com`.

M. Chu, et al. (2005). 'Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty'. *Phys. Med. Biol.* **50**:5463–5477.

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

Q. Hou, et al. (2003). 'An optimization algorithm for intensity modulated radiotherapy - The simulated dynamics with dose-volume constraints'. *Medical Physics* **30**(1):61–68.

Intensity Modulated Radiation Therapy Collaborative Working Group (2001). 'Intensity-Modulated radiotherapy: Current status and issues of interest'. *Int. J. Radiation Oncology Biol. Phys.* **51**(4):880–914.

R. Jeraj & P. Keall (1999). 'Monte Carlo-based inverse treatment planning'. *Phys. Med. Biol.* **44**:1885–1896.

L. Y. Klepper (1966). 'Electronic computers and methods of linear programming in the choice of optimial conditions for radiation teletherapy'. *Medicinskay Radiologia* **11**:8–15.

M. E. Kowalok (2004). *Adjoint methods for external beam inverse treatment planning*. Ph.D. thesis, Medical Physics Department, University of Wisconsin, Madison, WI.

M. Langer, et al. (1990). 'Large scale optimization of beam weights under dose-volume restrictions'. *Int. J. Radiation Oncol. Biol. Phys.* **18**(4):887–893.

J. Leong (1987). 'Implementation of random positioning error in computerised radiation treatment planning systems as a result of fractionation'. *Phys. Med. Biol.* **32**(3):327–334.

J. Llacer, et al. (2003). 'Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints'. *Phys. Med. Biol.* **48**:183–210.

W. Lu, et al. (2005). 'Accurate convolution/superposition for multi-resolution dose calculation using cumulative tabulated kernels'. *Phys. Med. Biol.* **50**:655–680.

C.-M. Ma, et al. (1999). 'Clinical implementation of a Monte Carlo treatment planning system'. *Medical Physics* **26**(10):2133–2143.

S. Morrill, et al. (1990). 'The influence of dose constraint point placement on optimized radiation therapy treatment planning'. *Int. J. Radiation Oncol. Biol. Phys.* **19**(1):129–141.

A. Ólafsson & S. J. Wright (2006). 'Linear programing formulations and algorithms for radiotherapy treatment planning'. *Optimization Methods and Software* **21**(2):201–231.

R. Reemtsen & M. Alber (2004). 'Continuous optimization of beamlet intensities for photon and proton radiotherapy'. Tech. rep., Brandenburgische Technische Universität Cottbus.

I. I. Rosen, et al. (1991). 'Treatment plan optimization using linear programming'. *Medical Physics* **18**(2):141–152.

J. Sempau, et al. (2000). 'DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations'. *Phys. Med. Biol.* **45**:2263–2291.

D. M. Shepard, et al. (1999). 'Optimizing the delivery of radiation therapy to cancer patients'. *SIAM Review* **41**(4):721–744.

S. Stapleton, et al. (2005). 'Implementation of random set-up errors in Monte Carlo calculated dynamic IMRT treatment plans'. *Phys. Med. Biol.* **50**:429–439.

J. Unkelbach & U. Oelfke (2004). 'Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions'. *Phys. Med. Biol.* **49**:4005–4029.

M. van Herk (2004). 'Errors and margins in radiotherapy'. *Seminars in Radiation Oncology* **14**(1):52–64.

C. Wu (2002). *Treatment planning in adaptive radiotherapy*. Ph.D. thesis, Medical Physics Department, University of Wisconsin, Madison, WI.

X. Zhang, et al. (2004). 'Speed and convergence properties of gradient algorithms for optimization of IMRT'. *Medical Physics* **31**(5):1141–1152.