Optimization in Intensity Modulated Radiation Therapy

Eva K. Lee
Center for Operations Research in Medicine,
School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332-0205, USA
Winship Cancer Institute and Dept. of Radiation Oncology,
Emory University School of Medicine
(evakylee@isye.gatech.edu).

Joseph O. Deasy
Division of Bioinformatics and Outcomes Research,
Department of Radiation Oncology,
Washington University School of Medicine,
St. Louis, MO 63110, USA
(deasy@radonc.wustl.edu).

Abstract: An overview and some computational challenges in intensity modulated radiation therapy are presented. Experience with a mixed-integer programming treatment planning model is described. The MIP model allows simultaneous optimization over the space of beamlet intensity weights and beam and couch angles. The model uses two classes of decision variables to capture the beam configuration and intensities simultaneously. Binary (0/1) variables are used to capture "on" or "off" or "yes" or "no" decisions for each field, and nonnegative continuous variables are used to represent intensities of beamlets. Binary and continuous variables are also used for each voxel to capture dose level and dose deviation from target bounds. The treatment planning model was designed to explicitly incorporate the following planning constraints: (a) upper/lower/mean dose-based constraints, (b) dose-volume and equivalent-uniform-dose constraints for critical structures, (c) homogeneity constraints (underdose/overdose) for the planning target volume (PTV), (d) coverage constraints for PTV, and (e) maximum number of beams allowed. Results of applying the MIP Model to a patient case are presented. Brief discussions of recent linear programming and nonlinear programming treatment planning models are also described, as is an MIP approach for direct aperture optimization.

1. Introduction

Every year over 1.4 million new cancer cases are diagnosed [1] in the United States, and over half of the patients receive radiation treatment at some point during the course of their disease. The key to the effectiveness of radiation therapy for the treatment of cancer lies both in the fact that the repair mechanisms for cancerous cells are less efficient than that of normal cells, and the ability to deliver higher doses to the target volume using "cross-fired" radiation beam. Thus, a dose of radiation sufficient to kill cancerous cells may not be lethal for nearby healthy tissue. Nevertheless, avoiding or minimizing radiation exposure to healthy tissue is extremely important.

Using multiple beams of radiation from multiple directions to cross-fire at the tumor volume provides a method to keep radiation exposure to normal tissue at relatively low levels, while dose to tumor cells is elevated. The crux of the treatment planning process involves designing beam profiles (*i.e.*, a collection of beams) that delivers a sterilizing dose of radiation to the tumor volume, while dose levels to critical normal tissues are kept below established tolerance levels. Often, one attempts to design a plan for which the prescription dose isodose surface conforms to the geometric shape of the specified tumor volume [28, 67]. (The term *prescription dose* typically refers to the minimum dose desired to be delivered to the tumor volume; it is generally physician specified.)

Linear accelerators (LINAC) are common beam delivery units used for external beam radiotherapy. The table on which the patient lies and the beam delivery mechanism for the LINAC rotate about separate orthogonal axes, providing the ability to adjust the angle and entry point of radiation fields used during treatments. Each field is further defined by a bank of multileaf collimators (MLC), small metallic leaves located inside the LINAC treatment unit. These leaves can be opened or closed, and used to shape the radiation beam as it exits the machine. Figure 1 shows a linear accelerator.

Intensity-modulated radiation therapy (IMRT) is an important recent advance in radiation therapy [68]. In conventional radiotherapy treatment, the planning process consists of determining a set of external beams that meet, as best as possible, the clinical dose distribution criteria. In many cases, significant compromises to critical structure function have to be made to enable a tumoricidal dose to

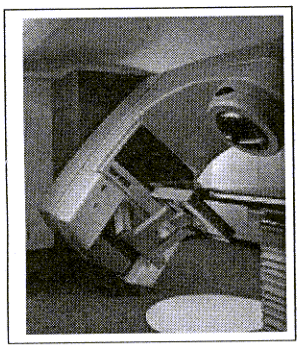

Figure 1: A linear accelerator used for external beam radiotherapy treatment

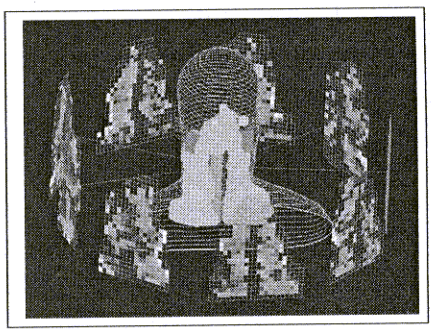

Figure 2: The treatment of a head-and-neck case via IMRT. Shown is a 3D view of the patient, the planning tumor volume (PTV), yellow; the spinal cord, pink; and the parotid glands, red. The 9 beams, shown with gray levels, reflect the modulated radiation intensity. (Use with permission from [2])

be delivered to the targets. In IMRT, the radiation fluence is varied across the beam, which allows a higher degree of conformation to the tumor than previously possible and allows concave isodose profiles to be generated, which may block, for example, dose to critical structure anterior or posterior to the target from that view. Specifically, not only is the shape of the beam controlled, but combinations of open and closed multileaf collimators modulate the intensity as well. For this reason, IMRT provides improved delivery control over conventional treatment. Indeed, it provides an unprecedented capability to dynamically vary the dose to accommodate the shape of the tumor from different angles, and to spare normal tissues and organs-at-risk (OAR) that may be potentially harmed during treatment.

Due to the complexity of the beam intensity profile associated with IMRT, there has been a tremendous research effort among medical physicists and radiation oncologists related to IMRT treatment planning and delivery, and there remain many opportunities for computational advances, particularly in treatment design. A computer-driven optimization algorithm must be used to determine the beam fluences (intensity maps) that provide the best compromise between target underdosing, target overdosing and critical structure overdosing. The textbook by Webb [68] has a good list of references for IMRT optimization.

In Sections 2.1 and 2.2, we describe the treatment planning problem for IMRT, and discuss relevant input data and the dose matrix. In Section 2.3, we discuss our experience of a mixed integer programming treatment planning model. The mixed integer programming model allows one to simultaneously incorporate dose coverage, underdose, overdose, homogeneity and conformity criteria on the tumor volume; dose volume restrictions on the critical structures (how much volume can receive more than a specified dose); and physical constraints on the total number of beams. Section 2.4 describes briefly the associated clinical results, and Section 2.5 provides a very brief discussion of current mathematical programming approaches. Summary and discussion is presented in Section 3.

2. Intensity-modulated radiation therapy treatment planning

Treatment planning in intensity modulated radiation therapy consists of a sequence of steps:

- Acquiring a 3D image of the affected region
- Delineating target volumes and healthy anatomical structures
- Selecting the appropriate radiation source and energy
- Selecting a set of beam angles for use in treatment
- Computing dose from each beam
- Performing intensity map optimization for the selected beams
- Developing optimal collimator sequences for actual delivery

These steps can be performed sequentially, or some can be combined together, resulting in complex numerical problems. In the sections presented herein, much of the description follows our recent work on mixed integer programming in this area [35, 36]. Very brief discussions on linear programming and nonlinear programming approaches are included.

2.1 Input data and dose calculation

Image Acquisition and Segmentation. The planning process begins when the patient is diagnosed with a tumor mass and radiation is selected as part of the treatment regime. A 3D image, or volumetric studyset, of the affected region, which contains the tumor mass and the surrounding areas, is acquired via computed tomography (CT) scans. These CT data are used for treatment planning, and electron density information derived from them are used in the photon dose calculations. Additionally, magnetic resonance imaging (MRI) scans may be acquired, fused with the CT volumetric studyset, and used to more accurately identify the location and extent of some tumors — especially those in the brain. Based on these scans, the physician outlines the tumor, and also outlines anatomic structures that need to be held to a low dose during treatment.

It is common practice to identify three "volumes" associated with the tumor. The gross tumor volume (GTV) represents the volume that encompasses the imageable or palpable macroscopic disease; that is, the disease that can be detected and localized by the oncologist. The clinical target volume (CTV) expands the GTV to include regions of suspected microscopic disease. The delineation of the CTV depends heavily on *a priori* knowledge of the behavior of a given tumor type. For a given GTV, tumor histologic features, and patient type, a set of probabilities exist (imperfectly known) that the tumor will, or will not, extend microscopically into a given regional organ or lymph node. However, accurate specific data are usually not available to the radiation oncologist, only general principles are known. A more quantitative and consistent definition of the CTV is an important need. The planning target volume (PTV) includes additional margins for anatomical and patient setup uncertainties related to organ and patient movement over time. All volumetric data is discretized into voxels (point representations of volume properties) at a granularity that is conducive both to generating a realistic model and to ensuring that the resulting treatment planning instances are tractable (*i.e.*, capable of being solved in a reasonable amount of computational time for practical clinical usage).

Dose Calculation Radiation dose, measured in Gray (Gy), is energy (Joules) deposited locally per unit mass (kg). Fluence for external beam photon radiation is defined mathematically by the number of photon crossings per surface area. Dose tends to be proportional to fluence, but is also influenced by photons and electrons scattered in the patient's tissues as well as the incident energy and media involved.

The calculation of the dose distribution associated with IMRT delivery is a critical aspect of the IMRT optimization and delivery processes. The calculated dose distribution from each candidate set of plan parameters is evaluated at each iteration or at the end of the optimization process, and the objective function values (costs or scores) for the iterative optimization are typically obtained by analysis of the dose distribution. For most systems, after the fluence-optimized plan is obtained, another dose calculation/optimization procedure, called leaf sequencing, is performed which first breaks the beams up into machine-deliverable multileaf sequencing steps, and then includes a final dose calculation step based on the details of the multileaf field shapes.

One of the most commonly used IMRT dose calculation algorithms involves a simple pencil beam method and is usually part of a broader class of correction-based dose-calculation algorithms [40, 4]. While these models offer significant speed advantages for use in the optimization code, they have varying limitations in accuracy.

In contrast, convolution/superposition, energy deposition kernel-based approaches can take into account beam energy, geometry, beam modifiers, patient contour, and electron density distribution [41, 10, 3, 6, 43]. Both the convolution method and the Monte Carlo method compute the dose per unit energy fluence (or fluence) incident on the patient.

Although it is clear that improved dose-calculation accuracy afforded by the convolution-type calcula-

tions may be important for IMRT, the long calculation times make this difficult.

Recently, significant progress has been made in the development of Monte Carlo calculation algorithms for photon beams, which simulate particle tracks individually, that are fast enough to compete with other current methods [24, 39, 70, 57]. In several situations, the Monte Carlo method is likely to be even more accurate than the convolution method [71]. For example, multiple scatter (second and higher order scatter) may be perturbed near the surface of a patient and the Monte Carlo method may be able to account for this as long as the number of simulated particles is sufficient. Direct Monte Carlo simulation may be the only option for achieving accurate dose computations in these complex situations. However, the application of Monte Carlo methods to optimization for IMRT is an area that requires much more work before relevant results will be available.

Access and usage of realistic radiotherapy data can be facilitated by using an open-source toolbox, developed by Deasy *et al.* [21], which enables users to import clinical plan data into Matlab for viewing and manipulation, and furthermore includes tools to generate the dose influence matrices.

2.2 Treatment planning strategies

In a strategy known as forward treatment planning, the beam geometry (beam orientation, shape, modifier, beam weights, etc.) is first defined, followed by calculation of the 3D dose distribution. After qualitative review of the dose distribution by the treatment planner and/or radiation oncologist, plan improvement is often attempted by modifying the initial geometry (*e.g.*, changing the beam weights and/or modifiers, adding another beam), to improve the target dose coverage and/or decrease the dose in the organs at risk. This forward planning process is repeated until a satisfactory plan is generated. As one can imagine, this is a time consuming approach to treatment planning.

In newer inverse treatment planning, the focus is on the desired outcome (*e.g.*, a specified dose distribution or tumor control probability (TCP) and normal tissue complication probability (NTCP)) rather than on how the outcome is achieved. The user of the system specifies the goals; the computer (optimization system) then adjusts the beam parameters (mainly the intensities) iteratively in an attempt to achieve the desired outcome. After review of the computer optimized dose distribution, some modification of the desired outcome and adjustment of the relative importance of each end point might be needed if the physician is not satisfied with the dose to the target volume or organs-at-risk (OARs).

Clearly, optimization is a classical inverse planning approach: constraints and an objective function are utilized to guide the optimization solver to select a plan with pre-specified clinical properties. Beginning with the work of Bahr *et al.* [5] in the late 60's, a number of research articles, authored primarily by medical researchers, discussed the use of mathematical programming and other optimization techniques in conventional external beam radiation treatment planning [18, 30, 31, 32, 34, 56, 61, 65, 73].

Much of IMRT treatment planning research has focused on the determination of the fluence map [67, 2, 7, 8, 11, 13, 44, 45, 72, 26, 12, 15, 19, 25, 27, 23]; that is, the radiation intensity or beam weights associated to each of the small beamlets of a selected radiation field/beam. However, the determination of beam angles, shapes, modifiers, couch positions and radiation energy to be used are best modeled using discrete variables.

At present, most IMRT optimization systems use dose-based and/or dose-volume-based criteria. One method commonly used to create dose-based and dose-volume objective functions involves minimizing the variance of the dose relative to the prescribed dose for the target volumes or dose limits for the organs at risk. This type of objective function has been used for traditional radiation therapy treatment optimization for the past several decades [62]. Variance is defined as the sum of the squares of the differences between the calculated dose and the prescribed dose or dose limit. Thus, a typical dose-based or dose-volume-based objective function is the sum of the variance terms representing each anatomic structure multiplied by appropriate penalty factors (*i.e.*, importance factors). Just as in conventional radiation therapy [14], the resulting unconstrained quadratic programming problem is often solved via the gradient method [61, 72], although the inclusion of dose-volume constraints makes the problem non-convex [20].

Within the optimization community, linear programming and nonlinear programming have been used to determine the optimal intensity map [55, 58], while mixed integer programming has been introduced to simultaneously determine the optimal beam angles and beam intensities [35, 36], and in finding optimal apertures for radiation delivery [52]. Below, we describe the MIP models formulated for simultaneous beam angle and intensity map optimization, closely following the presentation in Lee *et al.* [35, 36]. Results from a patient case will be briefly summarized. We then briefly describe linear and nonlinear programming approaches by others. Besides mathematical programming approaches, heuristic approaches — such as simulated annealing and genetic algorithms — have been commonly used for radiation therapy treatment optimization.

2.3 Mixed integer programming treatment planning models

The treatment planning models in [35, 36] use two classes of decision variables to capture the beam configuration and intensities simultaneously: Binary (0/1) variables are used to capture "on" or "off" or "yes" or "no" decisions for each field, and nonnegative continuous variables are used to represent intensities of beamlets. Binary and continuous variables are also used for each voxel to capture dose level and dose deviation from target bounds. Below, we provide the mathematical description of the treatment planning models.

Let $\mathcal{B}$ denote the set of candidate beams (each with an associated beam angle), and let $\mathcal{N}_i$ denote the set of beamlets (discretized sub-beams — usually rectangular in cross-section — which comprise the beam) associated with beam $i \in \mathcal{B}$. Beamlets associated with a beam can only be used when the beam is chosen to be "on". If a beam is on, the beamlets with positive dose intensity will contribute a certain amount of radiation dosage to each voxel in the target volume and other anatomical structures. Once the set of potential beamlet intensities is specified, the total radiation dose received at each voxel can be modelled. Let $w_{ij} \geq 0$ denote the intensity of beamlet j from beam i (in calibrated monitor units). Then the total radiation dose at a voxel P is given by

$$D_P(w) = \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{N}_i} D_{P,ij}\, w_{ij}, \tag{1}$$

where $D_{P,ij}$ denotes the dose per monitor unit intensity contribution to voxel P from beamlet j in beam i. Various dose constraints are involved in the design of treatment plans. Clinically prescribed lower and upper bounds, say L_P and U_P, for dose at voxel P are incorporated with (1) to form the basic dosimetric constraints:

$$\sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{N}_i} D_{P,ij}\, w_{ij} \geq L_P, \quad \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{N}_i} D_{P,ij}\, w_{ij} \leq U_P. \tag{2}$$

Our model also allows selection of optimal beam angles out of a collection of candidate beams. Thus, the resulting plan returns the optimal beam geometry as well as beam intensities.

Let x_i be a binary variable denoting the use or non-use of beam i. The following constraints limit the total number of beams used in the final plan and ensure that beamlet intensities are zero for beams not chosen:

$$\sum_{i \in \mathcal{B}} x_i \leq B_{\max} \qquad \text{and} \qquad w_{ij} \leq M_i x_i. \tag{3}$$

Here, $B_{\max}$ is the maximum number of beams desired in an optimal plan, and M_i is a positive constant that can be chosen as the largest possible intensity emitted from beam i.

For each voxel in each anatomical structure, we associate one binary variable and one continuous variable to capture whether or not desired dose level is achieved and the deviation of received dose from desired dose. We also impose additional constraints into our treatment plan design, as discussed below.

Clinically, it may be desirable to incorporate coverage constraints within the model. For example, the clinicians may consider that it is acceptable if, say, 95% of the PTV receives the prescription dose, $PrDose$. Such a coverage requirement can be modelled as follows.

$$\sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{N}_i} D_{P,ij} w_{ij} - r_P = PrDose, \; P \in PTV \tag{4}$$

$$r_P \leq D^{OD}_{PTV} v_P \quad (5)$$

$$r_P \geq D^{UD}_{PTV}(v_P - 1) \quad (6)$$

$$\sum_{P \in PTV} v_P \geq \alpha |PTV|. \quad (7)$$

Here, r_P is a real-valued variable that measures the discrepancy between prescription dose and actual dose; v_P is a 0/1 variable that captures whether voxel P is above or below the prescription dose bounds or not; α corresponds to the minimum percentage of coverage required (*e.g.*, $\alpha = 0.95$); D^{OD}_{PTV} and D^{UD}_{PTV} are the maximum overdose and maximum underdose levels tolerated for tumor cells; and $|PTV|$ represents the total number of voxels used to represent the planning target volume. The values D^{OD}_{PTV} and D^{UD}_{PTV} can be chosen according to the homogeneity level desired by the clinician for the resulting plan. If $r_P > 0$, then voxel P receives sufficient radiation dose to cover the prescribed dose. In this case, $v_P = 1$ and the amount of radiation for voxel P above the prescribed dose is controlled by the maximum-allowed-overdose constant, D^{OD}_{PTV}. Similarly, when $r_P < 0$, voxel P is underdosed, and the amount of underdose is limited by D^{UD}_{PTV}. In this case, $v_P = 0$.

By design, constraints (5) and (6) serve two purposes: 1) they capture the number of PTV voxels satisfying the prescription dose, and 2) they provide a means of controlling underdose, overdose, and dose homogeneity in the tumor. For the latter, the ratio $(PrDose + D^{OD}_{PTV})/(PrDose - D^{UD}_{PTV})$ can be viewed as an implied PTV homogeneity constraint associated with the model. Using a model with a smaller homogeneity constraint can be expected to result in a more homogeneous plan. Constraint (7) corresponds to the coverage level desired by the clinician.

Recently Equation (4) has been used to capture dose gradient fall-off when 100% tumor coverage is demanded. This was achieved by minimizing the dose surrounding the tumor region [34]. For IMRT planning optimization, it alone was used to model the deviation from prescribed dose for the PTV [13, 72, 17]. In these studies, a nonlinear objective function was formulated to steer the gradient-based optimization engine towards achieving the prescribed dose for the target volume; specifically, the objective was to minimize the sum of dose deviation across the target volume: $\|r\|_q = (\sum_P |r_P|^q)^{1/q}$ (with no imposed constraints). When $q = 2$, this is a least-squares problem.

It is desirable that dose received by radiation sensitive organs/tissues other than the tumor volume should be controlled to reduce the risk of injury. Thus, for other anatomical structures involved in the planning process, along with the basic dose constraints given in (2), additional binary variables are employed for modeling the dose-volume-tolerance relationships. To incorporate this concept into the model, let α_k, $\beta_k \in (0, 1]$ for k in some index set K. (In our implementations, the cardinality of the index set K is typically between 3 and 10 but could be larger.) The following set of constraints ensures that at least $100\beta_k\%$ of the voxels in an organ-at-risk, OAR, receive dose less than or equal to $\alpha_k\, PrDose$. The symbols $y^{\alpha_k}_P$ and z^{OAR}_P denote binary variables.

$$\sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{N}_i} D_{P,ij} w_{ij} \leq [\alpha_k\, PrDose] y^{\alpha_k}_P + \mathcal{D}_{max} z^{OAR}_P, \; P \in OAR \quad (8)$$

$$\sum_{P \in OAR} y^{\alpha_k}_P \geq \beta_k |\, OAR\,| \quad (9)$$

$$y^{\alpha_k}_P + z^{OAR}_P = 1 \quad (10)$$

$$y^{\alpha_{k_1}}_P \leq y^{\alpha_{k_2}}_P \text{ for } \alpha_{k_1} \leq \alpha_{k_2}. \quad (11)$$

Here, $\mathcal{D}_{max}$ is the maximum dose allowed for OAR (often determined by the maximum dose thought to be well-tolerated), and α_k, β_k combinations are patient and tumor specific. When the total dose received by a voxel P is less than $\alpha_k\, PrDose$, $y^{\alpha_k}_P = 1$, and this contributes to a voxel count in Constraint (9). When it does not satisfy the dose bound $\alpha_k\, PrDose$, then $y^{\alpha_k}_P = 0$, and in this case the dose will be forced to be lower than the maximum dose tolerance allowed, $\mathcal{D}_{max}$, and $z^{OAR}_P = 1$. Note that by using discrete variables to represent each voxel and controlling the number of points satisfying a certain dose level, we can impose strict dose-volume criteria within the solution space. This is in contrast to the common approach of incorporating "soft" dose-volume criteria into a composite objective function [72]. Langer[31] was the first to apply MIP ideas to model dose-volume relationships in conventional radiation therapy. For IMRT, the challenge is that the resulting problem instances are large-scale (involving hundreds of thousands or even millions of in-

equalities), and are computationally taxing and difficult to solve without the development of specialized algorithms [35].

Besides the commonly used least-squares dose deviation objective function, other objective functions have been used, including: minimizing the squared radiation dose to OARs, maximizing the minimum dose to tumor target, maximize/minimize weighted sum of doses to target and OARs. Other more complex biological objective functions — involving equivalent uniform dose (the p-norm or generalized mean value), tumor control probability, and normal tissue complication probability — have also been proposed [37, 38, 60, 59, 47, 50, 48, 49, 46].

The MIP treatment planning models for real patient cases involve tens to hundreds of thousands of binary variables and constraints. Our experience is that the resulting MIP instances are intractable via commercial MIP solvers. However, we have observed that, by using specialized algorithms [35], clinically superior treatment plans can be obtained [36].

2.4 Computational results for a real patient case

We briefly describe a patient study. Input data includes 3D images of tissue to be treated. On these images, the planning target volume (PTV) is delineated, and contours of organs-at-risk (OAR) and normal tissue are outlined. In addition to these structures, a tissue ring of 5 mm thickness is drawn around the PTV. We call this ring the *critical-normal-tissue-ring*. In [31] it was demonstrated that this normal tissue construct can assist in obtaining conformal plans for radiosurgery. In [35, 36], we have shown its usefulness in designing conformal IMRT plans. For the results herein, depending on the volume of the anatomical structure, a 3–5 mm voxel size (for dose computation) is used for setting up the MIP model instances.

For each beamlet, the dose per monitor unit intensity to a voxel is calculated. The total dose per unit intensity deposited to a voxel is equal to the sum of dose per intensity deposited from each beamlet. For the results described here, 16-24 coplanar fields of size 10×10 cm^2 to 15×15 cm^2 are generated as candidate fields, each of which consists of 400-900 0.5×0.5 cm^2 beamlets. This results in a large set of candidate beamlets used for instantiating the treatment planning models.

In [35], we study the effect of maximum beam angles allowed on plan quality. In [36], five objective functions are considered and contrasted on three different tumor sites to compare plan quality and to gain understanding of the steering effects of clinical objectives. Below, we illustrate the results for a head-and-neck case obtained via multiple objectives.

Some common metrics for reporting quality of treatment plans include:

- Coverage — Coverage is computed as the ratio of the target volume enclosed by the prescription isodose surface to the total target volume. Coverage is always less than or equal to 1.
- Conformity — Conformity is a measure of how well the prescription isodose surface conforms to the target volume; it is computed as the ratio of the total volume enclosed by the prescription isodose surface to the target volume enclosed by this same surface. Conformity is always greater than or equal to 1.
- Homogeneity — The homogeneity index is defined as the ratio of the maximum dose to the minimum dose received by the tumor volume.
- Mean dose and maximum dose for each critical structure.
- Dose-volume histograms (volume receiving more than each given dose level) and isodose curves.

Observe that these metrics are not entirely independent. For example, while it is desirable to obtain a prescription isodose surface big enough to cover the target volume in order to ensure good coverage, it is also desirable to have this surface "small" in order to conform to the target volume. In addition, variations in conformity and coverage affect the amount of irradiation to nearby organs at risk, thus affecting dose distribution levels of these organs.

Head-and-neck tonsil cancer. We focus on a tonsil cancer case where the PTV is adjacent to the left submandibular salivary gland. The following structures with their respective clinical dose limits are considered. PTV should receive 68 Gy; left parotid: $30\% \leq 27$ Gy and $100\% \leq 68$ Gy; right parotid: $100\% \leq 15$ Gy; right submandibular gland:

100% $\leq$ 30 Gy; left submandibular gland: 10% $\leq$ 27 Gy and 100% $\leq$ 68 Gy; larynx: 80% $\leq$ 30 Gy and 100% $\leq$ 55 Gy; spinal cord and brainstem: 100% $\leq$ 45 Gy.

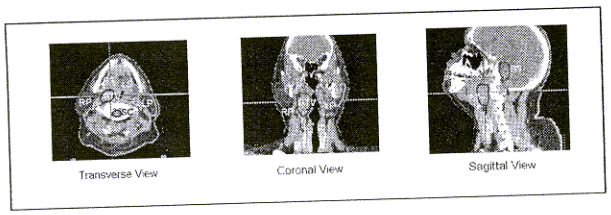

Figure 3: Anatomical structures for the head-and-neck. Notation: right parotid (RP), left parotid (LP), right submandibular gland (RS), left submandibular gland (LS), spinal cord (SP), brain stem (BS).

A total of 1501 PTV voxels, 406 critical-normal-tissue-ring voxels, 3247 voxels for the OARs and 6416 normal tissue voxels were used to instantiate the MIP treatment model.

Here, we report the results for a plan with a maximum of 7 beams in which the objectives include minimizing the total dose to the critical structures and optimizing the PTV conformity. The results are based on the utilization of a specialized branch-and-bound MIP solver for large-scale external beam radiation [35] that is built on top of a general-purpose mixed integer research code (MIPSOL) [33]. Figure 4 shows the dose volume histograms, and Figure 5 shows the isodose curves. Compared to the clinical plan, we observe the following:

a. For all critical structures, the mean dose and max dose received are drastically less than the clinical plan.

b. For OARs that are close to the tumor volume, namely the left parotid ($< 10mm$) and the left submandibular gland ($< 10mm$), the mean dose received is significantly reduced (70% and 50%, respectively). The spinal cord enjoys moderate dose reduction (33%).

c. The coverage constraint and the objective helped in achieving 98% coverage. Underdose and overdose constraints kept minimum and maximum dose to the tumor relatively uniform, with a homogeneity index of 1.24. And the conformity objective helped to achieve a superior conformity value of 1.34. These all improve over the clinical plan, which had 97% coverage, and scores of 1.4 for homogeneity, and 1.6 for conformity.

d. The total overall monitor units of radiation from the MIP optimized plan is less than that from the clinical plan, indicating that the plan uses less radiation but yet can still deliver the required prescription dose to the tumor, thus sparing excessive radiation dose to the critical structures and normal tissue.

It is noteworthy that in contrast to the typically equispaced beams chosen when beam configurations are pre-selected, the optimal 7-beam plans obtained herein (that are considered clinically acceptable) do not have equispaced beams. Indeed, the optimal beam angles returned appear to be non-intuitive, and to depend on PTV size and geometry and the spatial relationship between the tumor and the critical structures.

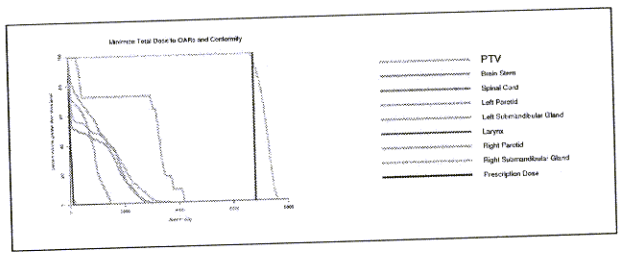

Figure 4: Dose-volume histogram for the head-and-neck for the MIP model with objective of minimizing the OARs dose and optimizing prescription dose conformity to tumor.

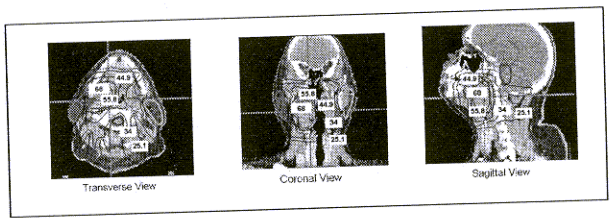

Figure 5: Isodose curves for the head-and-neck case. The critical-normal-tissue-ring is represented by the dotted curve.

2.5 Other mathematical programming approaches

As previously mentioned, linear and nonlinear programming have long been used for radiation therapy treatment optimization [5, 56, 55]. Lacking discrete variables, LP and NLP models typically use a pre-selected beam configuration, and focus on determining beam intensities. Below, we briefly outline some recent approaches in this area.

Simplified least-squares objective function and dose-volume constraints: In [55], using pre-selected beam angles, linear programming approaches were used to determine the associated optimal intensity map. The authors approximated the least-squares objective function measuring deviation of tumor voxel dose from prescribed dose via a piecewise linear function. They also utilized conditional value-at-risk (CVaR) constraints to control the mean dose received by subsets of voxels receiving the highest or lowest doses among all voxels in a given structure. Two forms of such constraints were used:
(i) lower α-CVaR: The average dose received by the subset of a target of relative volume 1-α receiving the lowest doses must be at least equal to L^α.
(ii) upper α-CVaR: The average dose received by the subset of a structure of relative volume 1-α receiving the highest doses may be no more than U^α.
CVaR constraints were originally proposed by Rockafellar and Uryasev [54] to formulate risk management constraints in terms of the tail means of distributions of financial risk. Mathematically, the upper α-CVaR constraint on a structure S is defined as

$$\bar{\zeta}_S^\alpha(w)+\frac{1}{(1-\alpha)|S|}\sum_{j\in S}\max\{0, D_P(w)-\bar{\zeta}_S^\alpha(w)\} \le U_S^\alpha, \tag{12}$$

where U_S^α is an upper bound target, $D_P(w)$ is the total dose from intensity vector w for voxel P, and $\bar{\zeta}_S^\alpha(w)$ denotes the smallest dose level with the property that no more than $100(1-\alpha)$ percent of the structure S receives a larger dose. The authors showed that including such partial-volume constraints to bound the tail averages of the differential dose-volume histograms of structures helps to improve dose homogeneity to the target and to spare dose to critical structures.

Nonlinear programming approach: Sheperd *et al.* [58] summarized several LP and NLP models for determining optimal intensity maps. To model dose-volume constraints, they applied a nonlinear error function approach. Their problem involved minimizing the standard objective of sum of the square differences between the prescribed and the actual doses over all of the voxels in the tumor, subject to two partial volume constraints — to OARs and to normal tissue. For an OAR S, the partial volume constraint defined on S was:

$$\sum_{P\in S} erf(D_P(w)-\Lambda_P) \le \alpha|S| \tag{13}$$

where Λ denotes a selected dose limit and α denotes the fraction of the volume allowed to exceed this limit. The error function *erf(x)* realizing the partial volume constraints is a nonlinear function. (See fig. 4.4 in [58].)

The authors also compared this with an MIP approach to model partial volume constraints on OARs, involving a simplified version of constraints (8)–(11) described above.

Direct aperture approaches via mixed integer programming: Preciado-Walters *et al.* [52] formulated the treatment planning problem as a mixed integer program over a coupled pair of column generation processes: the first designed to produce intensity maps for the IMRT beamlet grid, followed by the second to specify protected area choices aiding in reducing the computational burden of enforcing the dose-volume restrictions on tissues.

Instead of determining the beamlet intensity for each beam, and then applying leaf-sequencing to determine delivery patterns, the planning involved first selecting a fixed set of deliverable beams. For each of these beams, the authors pre-determined heuristically a set of delivery patterns. They then introduced continuous nonnegative decision variables x_{jq} to represent the assigned intensity to whole pattern q of beam j. Then the dose at any voxel P, is calculated by

$$D_P = \sum_{j=1}^{n}\sum_{q\in\mathcal{Q}_j} a_{Pjq}x_{jq} \tag{14}$$

where $x_{jq} \ge 0$, $\mathcal{Q}_j$ is the set of patterns for beam j, and a_{Pjq} is the implied dose coefficient of pattern q from beam j at voxel P when the pattern q is constructed.

The resulting MIP model for treatment planning employed the objective function of maximizing the minimum tumor dose. Similar to the above MIP models, the constraints include upper and lower dose bounds on tumor voxels, and upper dose bounds for healthy tissues. Dose-volume constraints are formulated just as constraints (8)–(11) above.

3. Summary and discussion

This article provides a brief overview of optimization issues in intensity-modulated radiation therapy, and summarizes our experience with an integer programming approach. The MIP model described allows simultaneous optimization over the space of beamlet intensity weights and beam angles. Based on experiments with clinical data, this approach can return good plans that are clinically acceptable and practical. This work is distinguished from recent IMRT research in several ways. First, in previous methods beam angles are selected prior to intensity map optimization. Herein, we employ 0/1 variables to model the set of candidate beams, and thereby allow the optimization process itself to select optimal beams. Second, instead of incorporating dose-volume criteria within the objective function as in previous work, herein, a combination of discrete and continuous variables associated with each voxel provides a mechanism to strictly enforce dose-volume criteria within the constraints. The challenge of using MIP modeling for IMRT is that the resulting instances are very large-scale, and since general MIP is NP-hard, specialized algorithms designed to solve IMRT instances are required. Third, incorporating the critical-normal-tissue-ring can improve conformity in general tumor sites, without addition of other dose-shaping structures. In general, our MIP approach uses constraints to control a variety of clinical criteria (coverage, homogeneity, underdose to PTV, overdose to PTV, dose-volume limits on organs-at-risk and normal tissue), while assigning an objective to help with the solution search. The model can also be expanded to incorporate energy selection, couch angles and other treatment parameters.

Patient studies indicate that using the MIP approach, one can produce good clinical plans that aggressively lower OAR dose below pre-imposed levels without compromising local tumor control [36]. This is appealing since lower OAR dose should translate to lower normal tissue complication probability.

Computationally, the specialized optimization engine returns good feasible solutions within 30 minutes. We have performed standard leaf-sequencing techniques on the resulting optimal intensity map, and showed that returned plans are deliverable. The results provide evidence that the MIP approach is viable in producing good treatment plans that can potentially lead to significant improvement in local tumor control and reduction in normal tissue complication.

With pre-selected beam angles, other approaches such as linear programming [55] and nonlinear programming [72, 58] can be used for intensity map optimization. Comparisons are needed to gauge the quality of these plans versus those from MIP approaches. Direct aperture optimization [52] is appealing, since resulting segments are implementable directly. Again, comparisons are needed to determine the effectiveness, advantages and tradeoffs among different planning optimization methods.

Other computational challenges actively pursued by medical physics experts include image segmentation, planning under uncertainties, biological modeling, leaf-sequencing and treatment outcome analysis.

4. Acknowledgement

This research was partially supported by grants from the National Science Foundation, the National Institute of Health, and the Charles Edison Foundation.

REFERENCES

[1] American Cancer Society, *Source: Cancer Facts and Figures – 2006*, Atlanta, Georgia 2006.

[2] *Intensity-Modulated Radiotherapy: Current Status and Issues of Interest*, Intensity Modulated Radiation Therapy Collaborative, Working Group, Int. J. Radiat. Oncol. Biol. Phys., 51 (2001), pp. 880–914.

[3] A. Ahnesjo, *Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media*, Med. Phys., 16 (1989), pp. 577–592.

[4] A. Ahnesjo and M. M. Aspradakis, *Dose calculations for external photon beams in radiotherapy*, Phys. Med. Biol., 44 (1999), pp. 99–155.

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

[6] J. J. Battista and M. B. Sharpe, *True three-dimensional dose computations for megavoltage x-ray therapy: A role for the superposition principle*, Aust. Phys. Eng. Sci. Med., 15 (1992), pp. 159–178.

[7] T. Bortfeld, *Optimized planning using physical objectives and constraints*, Semin. Radiat. Oncol., 9 (1999), pp. 20–34.

[8] T. Bortfeld, K. Jokivarsi, M. Goitein, J. Kung, and S. B. Jiang, *Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation*, Phys. Med. Biol., 47 (2002), pp. 2203–2220.

[9] J. D. Bourland and E. L. Chaney, *A finite-size pencil beam model for photon dose calculations in three dimensions*, Med. Phys., 19 (1992), pp. 1401–1412.

[10] A. Boyer and E. Mok, *A photon dose distribution model employing convolution methods*, Med. Phys., 12 (1985), pp. 169–177.

[11] A. Brahme, *Development of radiation therapy optimization*, Acta Oncol., 39 (2000), pp. 579–595.

[12] M. P. Carol, *Integrated 3-D conformal multivane intensity modulation delivery system for radiotherapy*, Proceedings of the 11th International Conference on the Use of Computers in Radiation Therapy, Madison, WI, 1994.

[13] Y. Chen, D. Michalski, C. Houser, and J. M. Galvin, *A deterministic iterative least-squares algorithm for beam weight optimization in conformal radio therapy*, Phys. Med. Biol., 47 (2002), pp. 1647–1658.

[14] R. E. Cooper, *A gradient method of optimizing external-beam radiotherapy treatment plans*, Radiology, 128 (1978), pp. 235–243.

[15] C. Cotrutz and L. Xing, *Using voxel-dependent importance factors for interactive DVH-based dose optimization*, Phys. Med. Biol., 47 (2002), pp. 1659–1669.

[16] C. Cotrutz and L. Xing, *Segment-based dose optimization using a genetic algorithm*, Phys. Med. Biol., 48 (2003), pp. 2987–2998.

[17] S. M. Crooks and L. Xing, *Linear algebraic methods applied to intensity modulated radiation therapy*, Phys. Med. Biol., 46 (2001), pp. 2587–2606.

[18] S. Das and L. Marks, *Selection of coplanar and non-coplanar beams using three-dimensional optimization based on maximum beam separation and minimized nontarget irradiation*, Int. J. Radiat. Oncol. Biol. Phys., 38 (1997), pp. 643–655.

[19] C. De Wagter, C. O. Colle, L. G. Fortan, B. B. Van Duyse, D. L. Van den Berge, and W. J. De Neve, *3D conformal intensity-modulated radiotherapy planning: interactive optimization by constrained matrix inversion*, Radiot. Oncol., 47 (1998), pp. 69–76.

[20] J. O. Deasy, *Multiple local minima in radiotherapy optimization problems with dose-volume constraints*, Med. Phys., 24 (1997), pp. 1157–1161.

[21] J. O. Deasy, E. K. Lee, T. Bortfeld, M. Langer, K. Zakarian, J. Alaly, Y. Zhang, H. Liu, R. Mohan, R. Ahuja, A. Pollack, J. Purdy, and R. Rardin, *A collaboratory for radiation therapy treatment planning optimization research*, Ann. Oper. Res., to appear.

[22] D. Djajaputra, Q. Wu, Y. Wu, and R. Mohan, *Algorithm and performance of a clinical IMRT beam-angle optimization system*, Phys. Med. Biol., 48 (2003), pp. 3191–3212.

[23] B. A. Fraass, M. L. Kessler, D. L. McShan, L. H. Marsh, B. A. Watson, W. J. Dusseau, A. Eisbruch, H. M. Sandler, and A. S. Lichter, *Optimization and clinical use of multisegment intensity-modulated radiation therapy for high-dose conformal therapy*, Semin. Radiather. Oncol., 9 (1999), pp. 60–77.

[24] C. L. Hartmann Siantar, P. M. Bergstrom, W. P. Chandler, *et al.*, *Lawrence Livermore National Laboratory's PEREGRINE Project*, Proceedings of the XII International Conference on the Use of Computers in Radiation Therapy, Salt Lake City, Utah, 1997.

[25] M. Hilbig, R. Hanne, P. Kneschaurek, F. Zimmermann, and A. Schweikard, *Design of an inverse planning system for radiotherapy using linear optimization*, Med. Phys., 12 (2002), pp. 89–96.

[26] T. Holmes and T. R. Mackie, *A comparison of three inverse treatment planning algorithms*, Phys. Med. Biol., 39 (1994), pp. 91–106.

[27] D. H. Hristov and B. G. Fallone, *An active set algorithm for treatment planning optimization*, Med. Phys., 24 (1997), pp. 1455–1464.

[28] F. Khan, *The Physics of Radiation Therapy*, Williams and Wilkins, second edition, Baltimore, 1992.

[29] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Optimization by simulated annealing*, Science, 220 (1983), pp. 671–680.

[30] H. Kooy, L. Nedzi, J. Loeffler, E. Alexander, C. Cheng, E. Mannarino, E. Holupka, and R. Siddon, *Treatment planning for stereotactic radiosurgery of intracranial lesions*, Int. J. Radiat. Oncol. Biol. Phys., 21 (1991), pp. 683–693.

[31] M. Langer, R. Brown, M. Urie, J. Leong, M. Stracher, and J. Shapiro, *Large scale optimization of beam weights under dose-volume restrictions*, Int. J. Radiat. Oncol. Biol. Phys., 18 (1990), pp. 887–893.

[32] M. Langer, S. Morrill, R. Brown, O. Lee, and R. Lane, *A comparison of mixed integer programming and fast simulated annealing for optimizing beam weights in radiation therapy*, Med. Phys., 23 (1996), pp. 957–964.

[33] E. K. Lee, *Computational Experience with a General Purpose Mixed 0/1 Integer Programming Solver (MIPSOL)*, Software Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, 1997.

[34] E. K. Lee, T. Fox, and I. Crocker, *Optimization of radiosurgery treatment planning via mixed integer programming*, Med. Phys., 27 (2000), pp. 995–1004.

[35] E. K. Lee, T. Fox, and I. Crocker, *Integer programming applied to intensity-modulated radiation therapy treatment planning*, Ann. Oper. Res., 119 (2003), pp. 165–181.

[36] E. K. Lee, T. Fox, and I. Crocker, *Simultaneous beam geometry and intensity map optimization in intensity-modulated radiation therapy*, Int. J. Radiat. Oncol. Biol. Phys., 64 (2006), pp. 301–320.

[37] J. T. Lyman, *Complication probability as assessed from dose volume histograms*, Radiat. Res., 104 (1985), pp. 13–19.

[38] J. T. Lyman and A. B. Wolbarst, *Optimization of radiation therapy. III. A method of assessing complication probabilities from dose-volume histograms*, Int. J. Radiat. Oncol. Biol. Phys., 13 (1987), pp. 103–109.

[39] C.-M. Ma, E. Mok, A. Kapur, *et al.*, *Clinical implementation of a Monte Carlo treatment planning system*, Med. Phys., 26 (1999), pp. 2133–2143.

[40] T. R. Mackie, P. Reckwerdt, T. McNutt T, *et al.*, *Photon beam dose computations*, J. Palta and T. R. Mackie, editors, Teletherapy: Present and Future, Advanced Medical Publishing, College Park, MD (1996) pp. 103–136.

[41] T. R. Mackie, J. W. Scrimger, and J. J. Battista, *A convolution method of calculating dose for 15-MV x-rays*, Med. Phys., 12 (1985), pp. 188–196.

[42] G. S. Mageras and R. Mohan, *Application of fast simulated annealing to optimization of conformal radiation treatments*, Med. Phys., 20 (1993), pp. 639–647.

[43] R. Mohan, C. Chui, and L. Lidofsky, *Differential pencil beam dose computation model for photons*, Med. Phys., 13 (1986), pp. 64–73.

[44] R. Mohan, G. S. Mageras, B. Baldwin, *et al.*, *Clinically relevant optimization of 3-D conformal treatments*, Med. Phys., 19 (1992), pp. 933–944.

[45] R. Mohan, X. Wang, A. Jackson, *et al.*, *The potential and limitations of the inverse radiotherapy technique*, Radiot. Oncol., 32 (1994), pp. 232–248.

[46] A. Niemierko, *Reporting and analyzing dose distributions: A concept of equivalent uniform dose*, Med. Phys., 24 (1997), pp. 103–110.

[47] A. Niemierko and M. Goitein, *Calculation of normal tissue complication probability and dose-volume histogram reduction schemes for tissues with a critical element architecture*, Radiot. Oncol., 20 (1991), pp. 166–176.

[48] A. Niemierko and M. Goitein, *Modeling of normal tissue response to radiation: The critical volume model*, Int. J. Radiat. Oncol. Biol. Phys., 25 (1992), pp. 135–145.

[49] A. Niemierko and M. Goitein, *Implementation of a model for estimating tumor control probability for an inhomogeneously irradiated tumor*, Radiot. Oncol., 29 (1993), pp. 140–147.

[50] A. Niemierko, M. Urie, and M. Goitein, *Optimization of 3D radiation therapy with both physical and biological end points and constraints*, Int. J. Radiat. Oncol. Biol. Phys., 23 (1992), pp. 99–108.

[51] O. Z. Ostapiak, Y. Zhu, and J. Van Dyk, *Refinements of the finite size pencil beam model of three-dimensional photon dose calculation*, Med. Phys., 24 (1997), pp. 743–750.

[52] F. Preciado-Walters, R. Rardin, M. Langer, and V. Thai, *A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer*, Math. Program., 101 (2004), pp. 319–338.

[53] A. B. Pugachev, A. L. Boyer, and L. Xing, *Beam orientation optimization in intensity-modulated radiation treatment planning*, Med. Phys., 27 (2000), pp. 1238–1245.

[54] R. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Banking Finance, 26 (2002), pp. 1443–1471.

[55] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, A. Kumar, and J. G. Li, *A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning*, Phys. Med. Biol., 48 (2003), pp. 3521–3542.

[56] I. I. Rosen, R. G. Lane, S. Morrill, and J. A. Belli, *Treatment plan optimization using linear programming*, Phys. Med. Biol., 18 (1991), pp. 141–152.

[57] J. Sempau, S. J. Wilderman, and A. F. Bielajew, *DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations*, Phys. Med. Biol., 45 (2000), pp. 2263–2291.

[58] D. M. Shepard, M. C. Ferris, G. H. Olivera, and T. Rockwell Mackie, *Optimizing the delivery of radiation therapy to cancer patients*, SIAM Rev., 41 (1999), pp. 721–744.

[59] A. R. Smith and C. C. Ling, editors, *Implementation of three dimensional conformal radiotherapy*, Int. J. Radiat. Oncol. Biol. Phys., 33 (1995), pp. 779–976.

[60] A. R. Smith, J. A. Purdy, editors, *Three-dimensional photon treatment planning*, Report of the Collaborative Working Group on the Evaluation of Treatment Planning for External Photon Beam Radiotherapy, Int. J. Radiat. Oncol. Biol. Phys., 21 (1991), pp. 3–268.

[61] S. V. Spirou and C. S. Chui, *A gradient inverse planning algorithm with dose-volume constraints*, Med. Phys., 25 (1998), pp. 321–333.

[62] G. Starkschall, *A constrained least-squares optimization method for external beam radiation therapy treatment planning*, Med. Phys., 11 (1984), pp. 659–665.

[63] J. Stein, R. Mohan, X. Wang, T. Bortfeld, Q. Wu, K. Preiser, C. C. Ling, and W. Schlegel, *Number and orientations of beams in intensity-modulated radiation treatments*, Med. Phys., 24 (1997), pp. 149–160.

[64] H. Szu and R. Hartley, *Fast simulated annealing*, Phys. Lett. A., 122 (1987), pp. 157–162.

[65] U. Treuer, H. Treuer, M. Hoevels, P. Muller, and V. Sturm, *Computerized optimization of multiple isocenters in stereotactic convergent beam irradiation*, Phys. Med. Biol., 43 (1998), pp. 49–64.

[66] S. Webb, *Optimisation by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator*, Phys. Med. Biol., 36 (1991), pp. 1201–1226.

[67] S. Webb, *The Physics of Three-Dimensional Radiation Therapy: Conformal Radiotherapy, Radiosurgery and Treatment Planning*, Institute of Physics Publishing, Philadelphia, 1993.

[68] S. Webb, *Intensity-Modulated Radiation Therapy*, Institute of Physics Publishing, Bristol, 2000.

[69] S. Webb, D. J. Convery, P. M. Evans, *Inverse planning with constraints to generate smoothed intensity-modulated beams*, Phys. Med. Biol., 43 (1998), pp. 2785–2794.

[70] A. E. S. von Wittenau, L. J. Cox, P. M. Bergstrom, *et al.*, *Correlated histogram representation of Monte Carlo derived medical accelerator photon-output phase space*, Med. Phys., 26 (1999), pp. 1196–1211.

[71] M. K. Woo, J. R. Cunningham, and J. J. Jerioranski, *Extending the concept of primary and scatter separation to the condition of electronic disequilibrium*, Med. Phys., 17 (1990), pp. 588–595.

[72] Q. Wu and R. Mohan, *Algorithms and functionality of an intensity modulated radiotherapy optimization system*, Med Phys., 27 (2000), pp. 701–711.

[73] Y. Yu, M. Schell, and J. B. Zhang, *Decision theoretic steering and genetic algorithm optimization: Application to stereotactic radiosurgery treatment planning*, Med. Phys., 24 (1997), pp. 1742–1750.