Interactive Decision Support in Radiation Therapy Treatment Planning

Matthias Ehrgott
Department of Engineering Science
The University of Auckland
Private Bag 92019, Mail Centre, Auckland 1042, New Zealand
email: m.ehrgott@auckland.ac.nz

Ines Winz
School of Geography and Environmental Sciences
The University of Auckland
Private Bag 92019, Mail Centre, Auckland 1042, New Zealand
email: i.winz@auckland.ac.nz

Abstract

This paper proposes the use of an interactive decision support system to guide the treatment planning process for external beam radiation therapy. Based on multicriteria optimisation our research treatment planning software CARINA calculates efficient (also called Pareto optimal) treatment plans. These are stored in a database and accessed for evaluation by the treatment planner. The interactive component consists of navigation among the precalculated plans using free search, fine search and exact search as well as sensitivity analysis, which extracts dose dependence information for all structures from the plan database. As a result, plan quality is improved by finding advantageous trade-offs in competing treatment plans, trial-and-error is avoided, and effectiveness of treatment planning is increased.

Keywords: Multicriteria optimisation; decision support system; radiation therapy; treatment planning.

1 Introduction

Cancer is one of the most significant health problems worldwide with respect to its incidence and mortality alike. The main treatment form besides surgery and chemotherapy is radiation therapy. Ionising radiation is used to damage the DNA and interfere with division and growth of cancer cells. An estimated 50% of all patients diagnosed with cancer would currently benefit from radiotherapy, either to cure the disease or to palliate symptoms [47]. A curative treatment is focused on generating a tumouricidal dose, while the total dose in a palliative treatment is comparatively lower and is aimed at achieving temporary relief of symptoms.

Radiation therapy exploits the fact that cancerous cells are more susceptible to radiation than healthy cells. This difference in susceptibility is called the therapeutic ratio [54]. Treatment planning is concerned with improving the therapeutic ratio by appropriately choosing beam directions, beam intensities and other treatment parameters. Due to recent improvements in medical imaging (e.g. magnetic resonance imaging) and radiotherapy technology (e.g. intensity modulation by multi-leaf collimation) [43], an increase in applicability and treatment success of radiation therapy has been noted [53, 54]. However, narrow therapeutic ratios are still widely observed and have to be dealt with: They may result either in a lethal dose deposited in the tumour leading to unacceptable damage to one or more healthy structures or, conversely, in the avoidance of any damage to healthy structures implying ineffective treatment in terms of tumour control.

Before commencing treatment an often lengthy and complex planning process takes place. Complexity arises from patient geometry, i.e. the shape and site of the tumour and surrounding tissues, and the fact that a large number of plan parameters, such as number of beams, their directions, and intensities form interdependent, non-intuitive relationships [18] that influence the final radiation dose distribution. The optimisation of the radiation intensity delivered by given beams is managed by the treatment planning system (TPS). The optimised treatment plan is then evaluated by the radiation therapist and/or radiation oncologist.

A treatment plan consists of the treatment configuration for all equipment used to irradiate the patient, resultant dose distribution in the patient, and a set of treatment instructions [54]. This includes beam modality (photons, protons, etc.) and energy (measured in MeV), irradiation geometry (number of beams and their angles of incidence), the point of incidence (isocentre), beam intensities, and patient set-up (treatment table positioning, immobilisation devices).

Many advances were observed in areas of patient immobilisation [42], imaging [4, 43], and dose distribution calculation [22, 30] over the last years. However, plan optimisation, though widely investigated, is still often not satisfactory [23, 38]. The role of plan optimisation is to decide on a final treatment plan which is the best possible plan for the individual patient.

Often optimal solutions of mathematical models underlying optimisation are not clinically acceptable. This invariably results in a trial-and-error process where the planner changes input parameters in the search for a better optimisation output. The search may be very time-consuming, depending on the experience of the planner and the complexity of the case. Thus treatment planning can form bottlenecks and aggravate the waiting lists problem in oncology centres around the world [31].

In this paper we will assume that beam modality and geometry are given and focus on the intensity problem (also referred to as fluence map optimisation or beam weight optimisation).

This problem consists in deciding beam intensities that generate a dose distribution in the patient which achieves the curative or palliative primary objective of the treatment. Dose is measured as absorbed radiation in units of Gray (Gy). For planning purposes the primary objectives are realised by specifying a varying number of patient dependent treatment goals. These include avoidance, conformity, homogeneity, and simplicity goals [21]. Guidelines and protocols usually recommend values for each of these goals with respect to treatment site and progress of the disease [37]. Despite this, the level of importance of each of these goals may vary and different radiation therapists may set goals differently for the same patient based on their training, experience, and understanding of the patient's situation [40].

In the following section we outline the development of treatment planning strategies from forward to inverse (that is, optimisation based) planning, focusing on the most recent multicriteria based optimisation models. We argue that multicriteria optimisation models are most appropriate for radiotherapy treatment planning, but must be accompanied by decision support for effective use in practice.

In Section 3 we introduce the method adopted in our research treatment planning system Carina and illustrate it with an example planning session. Section 4 concludes the paper and points out directions for future research.

2 Forward, inverse, and multicriteria based treatment planning

Several planning strategies have been developed over the years, with improvements following advances in treatment equipment.

Early radiotherapy treatment used *forward* treatment planning, which is conducted iteratively. The treatment planner specifies all parameter values, after which the dose calculation software computes the dose distribution. If this is judged not acceptable by the oncologist, the initial parameter values are then adapted by trial-and-error until the dose distribution is satisfactory. In 2002, this planning strategy was still used in about 90% of the 5,500 cancer centres worldwide [49]. Despite the trial-and-error process, forward planning was acceptable for early forms of radiotherapy, because the use of open fields and simple wedge filters to modify intensities meant that only relatively few parameters needed to be specified.

Inverse planning has been introduced in the 1970's and has become popular during the 1990's. The idea of inverse planning is to specify the desired outcome, such as dose distribution, and compute beam intensities that produce this outcome, thereby eliminating the trial-and-error process of forward planning. The inverse planning paradigm requires optimisation models to mathematically formulate the relationship between beam intensities and dose distribution and to judge the quality of a treatment plan.

The introduction of intensity modulated radiotherapy (IMRT) made the inverse planning approach indispensable. IMRT uses multileaf collimators to allow a decomposition of a beam into a large number of beamlets. Since the intensity of the beamlets can be chosen independent of one another, the number of planning parameters has increased by orders of magnitude. It has become quite impossible for even experienced planners to consistently produce high quality plans in every

case [52].

A large number of different optimisation models have been suggested to optimally solve the intensity problem including linear programming models [5, 8, 29, 38, 41], mixed integer programming models [6, 21, 35, 51], and non-linear programming models [2, 10, 14]. [44] and [12] have each reviewed several optimisation models. We show three examples of intensity optimisation models: A linear model by [28], the most widely used non-linear model based on linear least squares (see e.g. [9, 45]) and a so-called biological model [11]. We do not present a formulation for a mixed integer programming model, because these models are usually extensions of linear programming formulations that use binary variables to include additional constraints, such as dose-volume constraints or homogeneity constraints, or include beam angle selection in the model. Such issues are beyond the scope of the paper.

Mathematical models for the intensity optimisation problem are based on the discretisation of the body and the beams. The body is divided into volume elements (voxels) or dose points. Voxels are cubic and their edge length may be defined by the slice thickness of the patient images and is often 3 or 5 mm. Deposited dose is calculated for one dose point in every voxel and assumed to be the same throughout the voxel. A beam is discretised into beam elements (bixels or beamlets or sub-beams). Their size depends on the number of leafs of the collimator and the number of stops for each leaf as it moves across the beam. The number of voxels may be tens or hundreds of thousands, and the number of bixels can be up to 1000 per beam (for a collimator with 120 leaves and stops every 5 mm for a 40 × 40 cm field [50]). The relationship between intensity and dose can then be formulated by a linear map

$$x \to d = Ax,$$

where x is a vector of bixel intensities indexed by (a, i), i.e. beam number a and beamlet i of beam a. The entries $a_{(j,a,i)}$ of A represent the rate at which dose is deposited in voxel j by beamlet i of beam a. Finally, d is a dose vector indexed by voxels j that represents the discretised dose distribution in the patient. The calculation of the values $a_{(j,a,i)}$ is referred to as dose calculation. We assume that A is given.

In the linear model of [28], bixels are assigned to each of three tissue types: tumour, critical structures, and remaining normal tissue. This results in a decomposition of A by rows into A_T, A_C, and A_N. Accordingly, let TUB and TLB be a vector of upper, respectively lower, bounds on dose for the tumour voxels, CUB a vector of upper bounds for the critical structure voxels, and NUB a vector of upper bounds for the remaining normal tissue. The model is as follows:

$$\begin{array}{rrcccl}
\text{minimise} & \omega_T l^T \alpha + \omega_C u_C^T \beta + \omega_N u_N^T \gamma & & & & \\
\text{subject to} & TLB - L\alpha & \leq & A_T x & \leq & TUB \\
 & & & A_C x & \leq & CUB + U_C \beta \\
 & & & A_N x & \leq & NUB + U_N \gamma \\
 & 0 & \leq & L\alpha & \leq & TLB \\
 & -CUB & \leq & U_C \beta & & \\
 & 0 & \leq & U_N \gamma & & \\
 & 0 & \leq & x. & &
\end{array} \tag{1}$$

Here, α, β and γ measure the underdose and overdose of tumour, critical structures, and remaining tissue compared to TUB, TLB, CUB and NUB, respectively. Nonnegative and full (column) rank matrices with positive row sums L, U_C, U_N and positive vectors l, u_C, u_N define how deviations are measured (e.g. the maximum or average over voxels). Through penalty factors ω_T, ω_C, and ω_N associated with these deviations the objective function penalises deviations from desired dose levels. The first three constraints place an upper, or in the case of the tumour, a lower elastic bound on the dose received. Elastic means that the bounds are allowed to vary with α, β, and γ, respectively. The remaining inequalities constrain the elastic bound values, and enforce non-negativity. It can be shown that the feasible region of this model is never empty for any L, U_C or U_G. This is unlike other linear programming formulations, where constraints may be set so tight that they become infeasible, a main criticism of linear programming models for intensity optimisation in the past. For more details we refer the reader to [28, 29].

Most non-linear programming formulations of the intensity problem are quadratic programmes of the least-squares type, i.e. they minimise the weighted squared differences between actual and prescribed doses summed over all structures, see e.g. [14, 45]:

$$\begin{array}{rl} \text{minimise} & \omega_T f_T^q(x) + \sum_{k=1}^K \omega_k f_{C_k}^q(x) + \omega_N f_N^q(x) \\ \text{subject to} & x \geq 0, \end{array} \tag{2}$$

where f_T, f_{C_k} and f_N are defined as

$$\begin{array}{rcl} f_T^q(x) & = & \|(TLB - A_T x)_+\|_2^2, \\ f_{C_k}^q(x) & = & \|(A_{C_k} x - CUB_{C_k})_+\|_2^2 \text{ for } k = 1, \ldots, K, \\ f_N^q(x) & = & \|(A_N x - NUB)_+\|_2^2. \end{array}$$

Here the critical structure voxels C are partitioned into individual structures C_k, $k = 1, \ldots, K$.

Negative values of $TLB - A_T x$, $A_{C_k} - TUB_k$, and $A_N x - NUB$ should be encouraged, hence $(\cdot)_+ := \max\{\cdot, 0\}$. The parameters ω_T, ω_{C_k}, and ω_N represent weights or "importance factors" of the tumour, critical structures and normal tissue. There are many variations of this model, e.g. $TG - A_T x$ may be used instead of $(TLB - A_T x)_+$ with a target dose TG. Then f_T is a measure of dose variability that is to be minimised. Setting $\omega_N = 0$ indicates that normal tissue is not considered. Values of $CUB_k = 0$ may be used in order to minimise dose to critical structures.

Whatever the variation, it has to be pointed out that the quadratic model (2) is convex, hence there is only one (global) minimum. These quadratic programmes can be solved with several techniques, e.g. sequential quadratic programming, gradient methods, quasi-Newton methods [10], or special techniques for nonnegative least squares problems [7] (see also [44] and references therein).

In biological models mathematical expressions for the dose response relationship of different tissues are used to formulate probabilities for tumour control (TCP) and complications in normal tissue ($NTCP$) as functions of dose (and therefore of intensity), see e.g. [3, 11]. This approach leads to nonlinear optimisation models to maximise, e.g. the complication free tumour control probability

$$\max TCP(1 - NTCP). \tag{3}$$

We observe that both the linear (1) and quadratic (2) models include "importance weights" as parameters in the objective functions, namely $(\omega_T, \omega_C, \omega_N)$ ($\omega_C = \omega_N = 1$ and ω_T large

are used in [28, 29]). These parameters have to be specified *before* optimisation. Therefore, albeit optimisation based inverse planning removes the trial-and-error process of choosing x and calculating Ax until a satisfactory treatment plan is found, a new trial-and-error process may be necessary. If the values chosen for the weights $(\omega_T, \omega_C, \omega_N)$ do not yield a satisfactory plan even with optimal x, then they have to be adjusted in an iterative fashion.

A second observation is that in forward and inverse planning alike, the fundamental problem is that minimising the dose to healthy tissues conflicts with generating a sufficiently high dose to the tumour. As a result, difficult decisions have to be made regarding the overdosing of organs and/or the underdosing of the tumour. These decisions will always have to balance the perceived risk of unsuccessful tumour control with the possibility of complications in healthy tissues.

The models (1) – (3) reflect this dilemma by including components that correspond to the different goals of treatment planning in the objective functions. In order to deliver a radiation dose to the tumour that achieves the curative or palliative intent of the treatment they penalise low dose to the tumour either directly or via low TCP. In order to limit the dose to healthy tissue they penalise high dose to the critical structures and normal tissue, again, either directly or via high $NTCP$. However, with these single objective models treatment plans are judged on the basis of one value representing the plan quality (objective function value or figure of merit). In this way they hide the possible trade-offs between the conflicting goals and hinder the effective use of that valuable information.

With this observation it becomes apparent that the formulations commonly used in inverse planning are not adequate models of the problem. Instead, the intensity problem should be formulated as multicriteria optimisation problem that keeps the conflicting objectives separate, i.e. the objective functions of (1) – (3) should be replaced by

$$\begin{aligned}
\min F^l(x) &= (f_T(x), f_C(x), f_N(x)) = (l^T\alpha, u_C^T\beta, u_N^T\gamma), \\
\min F^q(x) &= (f_T(x), f_{C_1}(x), \ldots, f_{C_k}(x), f_N(x)), \\
\min F^b(x) &= (f_T(x), f_{CN}(x)) = (TCP, 1 - NTCP),
\end{aligned}$$

respectively.

A feasible solution x of a multicriteria optimisation problem is called *efficient* (or Pareto optimal) if there is no other feasible solution x', such that x' is at least as good as x for all components of F^l (or F^q, or F^b) and strictly better for at least one. If x is efficient we say that $F(x)$ is a nondominated point. Multicriteria optimisation problems have a set of efficient solutions with incomparable nondominated points, rather than optimal solutions with a unique optimal objective value.

A number of multicriteria optimisation models have already been proposed in the literature. Multicriteria models with quadratic objectives can be found in [13, 25, 34], multiobjective linear programming formulations are given in [26, 27] and models using the concept of equivalent uniform dose are used in [32, 33, 48]. A survey of possible objective functions is given in [39].

We present the multicriteria linear model of [26], which has been used in CARINA. The model

is essentially a special case of the multicriteria version of (1).

$$\begin{array}{rrcl} \text{minimise} & (\alpha, \beta_1, \ldots, \beta_K) & & \\ \text{subject to} & TLB - A_T x & \leq & \alpha e \\ & A_T x - CUB_k & \leq & \beta_k \text{ for } k = 1, \ldots, K \\ & x & \geq & 0 \\ & \alpha, \beta_k & \geq & 0 \quad k = 1, \ldots, K, \end{array} \tag{4}$$

where e denotes a vector of ones of appropriate dimension. Thus we have $f_T^l(x) = \alpha = \max_{j \in T}((TLB - A_T x)_j)_+$ and similarly $f_{C_k}^l = \beta_k = \max_{j \in C_k}((A_{C_k} x - CUB_k)_j)_+$.

The objective is to minimise the maximum deviation from specified bounds in all structures and the constraints ensure that only overdose in critical structure voxels and underdose in tumour voxels is considered. Intensities x have to be nonnegative. Ideally, one would want a solution with $F^l(x) = (0, \ldots, 0)$, but due to the problem of narrow therapeutic ratios this does not usually exist, and we have to find a compromise solution.

The set of efficient solutions/nondominated points of a multicriteria model of the intensity problem represents the available trade-offs between the conflicting goals of radiotherapy treatment. Thus the solution which offers the best compromise for the individual patient should be chosen from this set. This approach guarantees that there is no plan that achieves better results against all objective functions, but guarantees enough flexibility for the planner to take case-specific information into account.

Since the set of efficient solutions (nondominated points) is a continuum it is not possible to compute all efficient solutions. In this context two results are important. [39] have shown that many of the objective functions used in intensity optimisation (including radiobiological ones) can be transformed into convex functions through strictly monotone mappings. [39] have also demonstrated that this does not change the set of efficient solutions (note that (1), (2), and (4) have convex objective and constraints). Thus, the intensity problem can be formulated as convex multicriteria optimisation problem. For such problems the following fundamental theorem of multicriteria optimisation is known.

Theorem 1 *[24] Let f and G be convex function. The set of properly efficient solutions of the multicriteria optimisation problem*

$$\min\{F(x) = (f_1(x), \ldots, f_p(x)) : G(x) \geq 0\}$$

is identical to the union of the sets of optimal solutions of weighted sum problems $\min\{\sum_{i=1}^p \omega_p f_p(x) : G(x) \geq 0\}$ *for all* $\omega \in \Omega = \{(\omega_1, \ldots, \omega_p) : \omega_k > 0$ *for all* $k = 1, \ldots, p$ *and* $\sum_{k=1}^p \omega_k = 1\}$.

Properly efficient solutions exclude those efficient solutions which have an infinite trade-off between at least two of the objectives: A deterioration in one can only achieve an infinitesimally small improvement in another. Since such solutions are clearly not relevant in the radiotherapy context, this qualification is of no importance for intensity optimisation models, on the contrary, it is an additional advantage.

Theorem 1 illustrates the close relationship between the traditional models like (1) – (3) and multicriteria models such as (4). The iterative process of optimisation and adjustment of weights

can be interpreted as the search for an efficient solution that offers an adequate compromise between the treatment goals for the case at hand.

Some researchers have tried to "optimise" importance factors [55] or to calculate them during optimisation [56] to speed up this process. The alternative that multicriteria models offer is to use a multi-plan strategy for treatment planning. This is addressed in the next section.

3 Decision support-based treatment planning

Some authors have suggested to generate a set of importance weights and solve the single objective version of their multicriteria model with a weighted sum of the objectives for all of these weights. The treatment planner can then select one of the generated plans that best suits the individual patient [13, 34]. This strategy is based on the idea that if a wide variety of weights is used, the resulting plans should also cover the set of efficient solutions.

However, it is known in multicriteria optimisation that the relationship between weights and efficient solutions/nondominated points is quite counterintuitive: very similar weights might produce widely differing efficient solutions, whereas very different weights might produce the same efficient solution. It is also known that given a set of weight vectors that are equally distributed over the set Ω there is no guarantee that the resulting nondominated points are well distributed too, as has been demonstrated e.g. by [15].

It is thus necessary to avoid using importance weights altogether in order to compute a set of treatment plans that truly reflect the available treatment options for a specific patient.

The preceding discussion shows that the problem shifts from generating a mathematically optimal plan to determining the one plan that is most suitable for the individual patient from among the set of efficient solutions. Multicriteria models make this possible because they allow a decoupling of plan optimisation and final plan determination. Once a set of efficient solutions is determined, the question is *How can the treatment planner determine the one plan that is best for the patient from a large number of efficient plans?*

As a result, the trial-and-error process, that iterates between human action (specifying weights) and plan optimisation, is abandoned in favour of guided search among pre-computed efficient treatment plans (Figure 1). A decision support system provides the planner with the necessary guidance in selecting the final treatment plan and making trade-off decisions between a set of pre-computed efficient treatment plans. Ideas for such DSS based treatment planning systems can be found in [33, 40].

With the model (4) our approach is aimed at generating a well-balanced trade-off between the over- and underdose of organs and the tumour, respectively. We have addressed this question with the software Carina.

3.1 Generating efficient solutions

Let us denote the feasible set of (4) by X. We first estimate the range of values that f_T^l and $f_{C_k}^l$ in (4) can take. This means calculating the ideal and nadir points of (4). The ideal point is 0, because all values are by definition nonnegative and there always exist solutions such that $f_T = 0$ (with large x) or $f_{C_k} = 0$ (with $x = 0$). The nadir point (the point defined by the maximal values

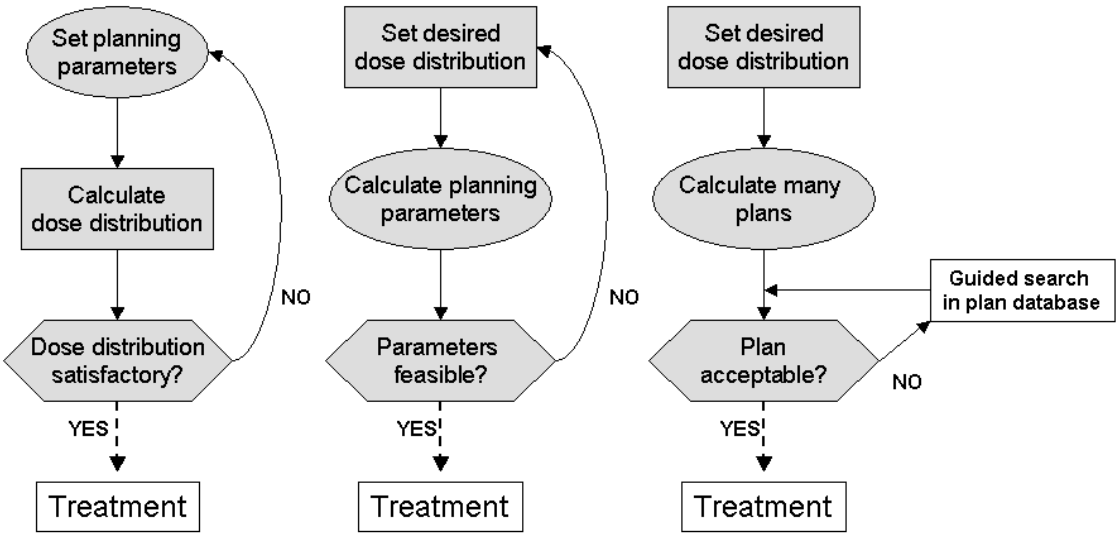

Figure 1: Changes in the treatment planning paradigm from forward planning (left) to inverse planning (middle) to decision support-based planning (right).

the objectives can attain over the efficient set) is hard to compute [20]. We estimate it by a lower bound by solving the following problems:

$$\begin{array}{l} \min f_T^l(x) + \varepsilon \sum_{k=1}^{K} f_{C_j}(x) \\ \min f_{C_k}^l(x) + \varepsilon \left(f_T(x) + \sum_{j \neq k} f_{C_k}(x) \right) \text{ for } k = 1, \ldots, K. \end{array} \tag{5}$$

Let x^T and $x^k, k = 1, \ldots, K$ be optimal solutions. By Theorem 1 these solutions are efficient. Furthermore, positive (but small) ε implies that $f_T(x^T)$ respectively $f_{C_k}(x^k)$ is small. Thus $\hat{f}^T = \max_{k=1,\ldots,K}\{f_T(x^k)\}$ and $\hat{f}^k = \max\{f_{C_k}(x^T), \max_{j \neq k}\{f_{C_k}(x^j)\}\}$ give the desired lower bounds on the nadir values.

Next, we find a solution $\hat{x}$ with the values of $f_T(\hat{x})$ and $f_{C_k}(\hat{x})$ as equal as possible. This is called the lexicographic max-ordering solution [16] or nucleolar solution [36] of (4). It can be interpreted as a balanced or unbiased solution of the problem and is also guaranteed to be efficient, see [16].

We define a grid of points that covers the area defined by the range of values $[0, \hat{f}^1] \times \ldots \times [0, \hat{f}^K]$ and is centred at $F^l(\hat{x})$. The grid is homogeneous, in that the distances between all neighbouring solutions are equal. One exception to this is the division of the grid into an inner and outer grid. The *inner grid* surrounds the balanced solution, and is surrounded by the outer grid. The points in the *outer grid* will be less densely spaced than in the inner grid (Figure 2). This is reasonable because the inner grid is of higher clinical importance, as the values of objectives are less extreme. In CARINA the user can decide on either the number of grid points or the maximum distance between two neighbouring points.

In order to find efficient solutions, the improved ε-constraint method [19] is used and an LP is

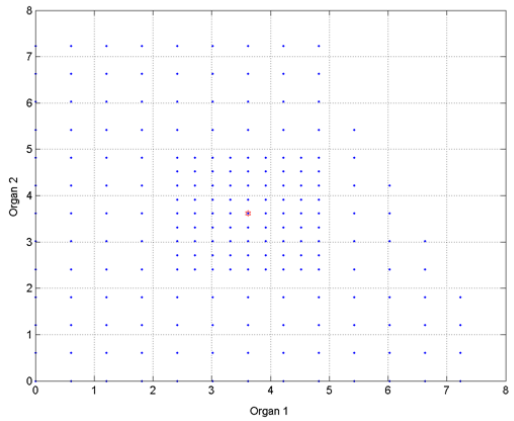

Figure 2: The grid of efficient solutions divided into inner (densely spaced points) and outer grid. The balanced solution is situated in the exact centre of the inner grid. The grid may not be complete if the boundary of the solution space has been reached.

solved for every gridpoint $g = (g_1, \ldots, g_K)$.

$$\begin{array}{rrcl} \text{minimise} & f_T(x) - \sum_{k=1}^{K} \lambda_k s_k + \sum_{k=1}^{K} \mu_k u_k & & \\ \text{subject to} & f_{C_k}(x) + s_k - u_k & = & g_k \text{ for } k = 1, \ldots K \\ & x & \in & X \\ & s_k, l_k & \geq & 0 \text{ for } k = 1, \ldots, K \end{array} \tag{6}$$

With positive λ_k and μ_k an optimal solution of (6) is always efficient [19]. The formulation aims at finding efficient solutions that match the values of the grid points, but allows for corrections in case a grid point is not efficient. In addition, for every grid point, the best possible result for the tumour is obtained. Note that unlike the importance factors ω_k the weights λ_k and μ_k here are technical parameters and can be handled automatically during optimisation.

Efficient treatment plans are calculated by solving the problems (5) – (6). The number of problems to be solved depends on the number of critical structures (i.e. objectives) and the coarseness of the grid and can be large. Note, however, that no interaction with the treatment planner is required during this computation, so it can be done unsupervised (at night), requiring only computing time. All efficient solutions are stored in a database. Each solution consists of the beamlet values x and the objective values $f_T(x)$ and $f_{C_k}(x)$. These are necessary for plan evaluation/navigation. The structure of the database is such that there is one table containing the efficient candidate plans. Each solution in this set is recognised by a unique plan identifier. The parameters used to calculate the plans, are stored in a different table and can be recognised by their unique parameter set identifiers.

During navigation, plan evaluation and comparison take place, and the radiation therapist decides on the final treatment plan that best fulfils individual treatment goals. This requires close interaction between the software and the radiation therapist. During the evaluation of the balanced solution the radiation therapist decides which treatment goals are met or which deviation value has to be improved (i.e. decreased). For this improvement another criterion value has to be traded-off, i.e. the radiation therapist determines which structure's deviation will be deteriorated

(i.e. increased). Given these requirements for improvement and deterioration, a free search or a fine search can be initialised. In the *free search*, the radiation therapist specifies and inputs the exact values for improvement and deterioration for each structure. Another option is to fix a number of structures at their current deviation levels. CARINA writes an SQL string according to these inputs, queries the database, and returns the plan which best fulfils these requirements. In this process, the deviation between a specified value and the actual value in the database will be minimised. The free search is cumbersome when simply a neighbouring solution is sought, i.e. a slight increase in one structure traded-off against a slight decrease in another structure. In this case, a *fine search* is more appropriate. The only inputs necessary are which structure to improve and which to deteriorate. CARINA will return a unique neighbouring solution if one exists. Otherwise it informs the planner that his requests cannot be met and resets the last user inputs. A third option is an *exact search*, where the plan identifier is input and the corresponding plan is immediately displayed. This is useful when solutions are revisited for comparison purposes or after sensitivity analysis (see below).

Each iteration of a free or fine search constitutes one search process. Navigation may consist of several of these search processes until a satisfying treatment plan is found. Searches will be very fast and query times instantaneous due to the query optimisation engine in the database management system. The navigation log will store the plan identifiers, deviation values, and search requirements for each search process. Evaluation of single solutions or the comparison of several solutions is possible by means of dose distribution/difference diagrams, dose-volume histograms, and bar charts of the deviation values [46]. The polygon approach by Küfer et al. [33] is another elegant concept for comparing deviation values.

During each successive search process CARINA provides further information based on the plan data available in the database. It provides the user with the number of available solutions in the database, which is bound to decrease as only a limited number of plans can satisfy set requirements. CARINA also outputs permissible value ranges for each structure, as they also change with specified requirements. However, most important is the use of *sensitivity analysis* to retrieve information on how the dose to one structure is dependent on the dose to another structure. The usefulness of sensitivity analysis has been noted elsewhere [1, 40]. For example, sensitivity analysis can yield information on how much a critical structure could be spared if a tumour dose reduction is accepted.

This information is not provided by current TPS, because multicriteria methods are necessary to obtain and exploit it. Additionally, sensitivity analysis has the power to reveal advantageous trade-offs where the total improvement greatly exceeds the total deterioration when comparing two rival treatment plans. Table 1 gives a convenient overview of the several navigation options CARINA provides.

3.2 Example treatment planning session

This simple example demonstrates the use of sensitivity analysis and compares it with the fine search. In this example, the tumour is situated in close proximity to three organs: spinal cord, left and right kidney. The treatment parameters specified before optimisation are summarised in Table 2.

Option	*Input*	*Explanation*
Free search	*Exact value desired for each structure*	*Minimisation of deviation between specified value and value saved in the database*
Fine search	*One structure desired to improve and one structure desired to deteriorate*	*Neighbouring solution will be found*
Exact search	*Plan identifier*	*Straight plan retrieval*
Available solutions	*Automatic after each search process*	*The number of plans that can satisfy the currently set requirements*
Value ranges	*Automatic before each search process*	*Possible value range of each structure available for the next search*
Sensitivity analysis	*One structure that may improve and one structure that may deteriorate*	*Shows how the two structure doses are dependent from one another and can thus reveal advantageous trade-offs*

Table 1: Navigation options and information provided by CARINA.

1581 efficient solutions were stored in the database.

The radiation therapist's task is to trade-off over- and underdose to organs and tumour, respectively, with regard to the treatment goals. Figure 3 shows the dose distribution diagram and dose-volume histogram for the balanced solution. The deviation values in all structures are equal with $\alpha = \beta_k = 2.97$ for $k = 1, \ldots, K$. Doses are already close to the bounds, which is a direct consequence of the large number of treatment fields. However, there is still room for navigation. Treatment goals are such that the dose to the radio-sensitive spinal cord ($k = 3$) should be decreased further, while accepting increased irradiation of the left kidney ($k = 1$). The right kidney ($k = 2$) should be spared as well, so that the main radiation burden is on the left kidney.[1] If possible the dose deposited in the tumour should not be more than $3\,Gy$ below the lower bound, i.e. $\alpha \leq 3$.

With this in mind, an improved plan would have the right kidney and the spinal cord irradiated at their set limits, i.e. $\beta_2 = 0$ and $\beta_3 = 0$. In order to obtain information on dose dependence between the left kidney and the tumour, a sensitivity analysis is performed which fixes deviation levels β_2 and β_3 at $0\,Gy$. The result is given in Table 3.

Opting for solution 631 instead of the balanced solution represents an advantageous trade-off. The total improvement is $6.22\,Gy$, which represents the combined improvements of the right kidney ($2.97\,Gy$), the spinal cord ($2.97\,Gy$), and the tumour ($0.28\,Gy$). This has been traded-off against a total deterioration of only $0.6\,Gy$ (in the left kidney only). Plans 727 and 823 represent even higher advantageous trade-offs, with improvements of $7.32\,Gy$ and $8.43\,Gy$, and deteriorations of $1.2\,Gy$ and $1.79\,Gy$, respectively. However, when single radiation doses become too large, they

[1] It is a common approach to completely protect one to the detriment of the other when dealing with paired organs.

Equipment specifications:	*Dose distribution specifications:*
$\succ$ 10 *treatment fields*	$\succ$ 80 Gy *tumour (min)*
$\succ$ 15 *degrees equispaced angles*	$\succ$ 33 Gy *left kidney (max)*
$\succ$ 16 MV *beam energy*	$\succ$ 33 Gy *right kidney (max)*
	$\succ$ 25 Gy *spinal cord (max)*
Grid specifications:	
$\succ$ 9 *points inner grid*	
$\succ$ 4 *points outer grid*	

Table 2: Example parameter settings.

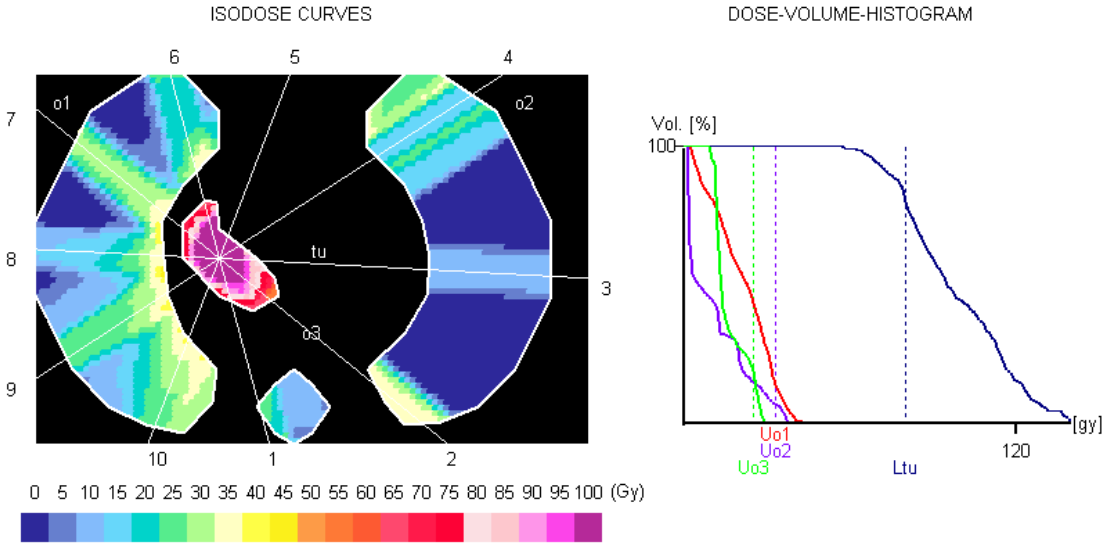

Figure 3: The isodose (left) and DVH (right) plots for the balanced solution. The objective function values are $F^l = \{2.97, 2.97, 2.97, 2.97\}$.

cause streaks of high radiation through the body. Hence, solution 631 is made the new plan of choice (Figure 4).

Note that a free search with $\alpha < 3$ and $\beta_2 = \beta_3 = 0$ would have given the same result, albeit without dose dependence information.

The next step is to search within the immediate neighbourhood of plan 631 in order to achieve a better deviation value for the left kidney. A fine tuning search is done with $\alpha < 3$, $\beta_1 < 3.57$, $\beta_2 = 0$ and $\beta_3 > 0$. Querying the database brings forward that such a plan does not exist. Repeating the query with $\beta_3 = 0$ and $\beta_2 > 0$ shows that β_2 must be increased to at least $4.76\,Gy$. This is not acceptable. The remaining options are to either increase the tumour deviation, keep the β_1 value at $3.57\,Gy$, or decrease both β_2 and β_3. The last approach was tested by sensitivity analysis. The result is given in Table 4.

Unfortunately, the doses to the right kidney and spinal cord would have to be increased quite dramatically in order to decrease the dose to the left kidney even slightly. This is not acceptable.

planID	α	β_1	$\Delta\,\alpha$	$\Delta\,\beta_1$	$\Delta\,\beta_1 - \Delta\,\alpha$
631	2.69	3.57	0.28	.60	.32
727	1.59	4.17	1.38	1.19	−.19
823	0.48	4.76	2.49	1.79	−.70

Table 3: Information extracted on the dose dependence between the tumour (T_1) and the left kidney (T_2).

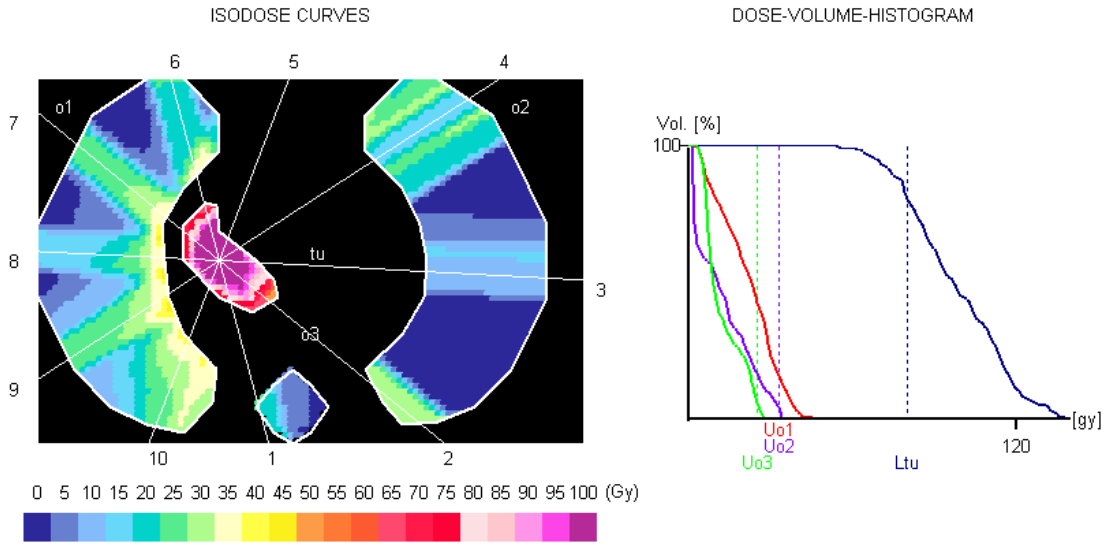

Figure 4: The isodose (left) and DVH (right) plots for solution 631. The objective function values are $F^l = \{2.69,\ 3.57,\ 0,\ 0\}$.

Hence, plan 631 is chosen as the final treatment plan for this patient. The corresponding beam intensities are stored in the database and can be retrieved for implementation,

4 Conclusions and directions for future research

Based on the reasonable assumption that oncology personnel want as much influence on the final treatment plan as possible, CARINA's navigation process accomplishes this using versatile options for navigation searches. Moreover, it provides information to support all search processes. Sensitivity analysis has emerged as an extremely useful option to find treatment plans having advantageously traded-off deviation values.

Interactivity in treatment planning is a rather new concept. One condition for interactivity is certainly that plan optimisation and evaluation is instantaneous. However, most TPS have re-optimisation times of several minutes. [40] recently presented a new research TPS that interactively explores optimal treatment plans based on objective parametric programming. A limited number of optimal treatment plans with varying parameters are pre-calculated and subsequently interactively evaluated. They also conclude that interactivity leads to more efficiency in treatment planning and increases plan quality. It has to be noted, however, that the set of solutions to

planID	α	β_2	β_4	$\Delta\,\alpha$
1213	2.998	2.68	4.17	−.306
1217	2.997	2.97	2.68	−.304
588	2.991	3.57	1.19	−.298
1224	2.983	3.27	2.08	−.290
1218	2.977	2.97	2.97	−.285

Table 4: Sensitivity analysis for an improvement in the left kidney and a deterioration in the right kidney, the spinal cord, and the tumour.

choose from is biased with respect to the choice of initial parameters and hence does not exhaust the whole efficient set. As a result, the best patient-tailored solution may not be identified.

Multicriteria optimisation on the other hand provides a cover of the whole relevant solution space. The only limitations of this approach are the need for storage space and the large plan calculation time [39]. The advantages however are convincing. The decision making task performed by the radiation therapist and oncologist is supported, not replaced, by a decision support system. The reason for achieving or not achieving patient specific treatment goals becomes more transparent and directly influenceable. As a result of avoiding trial-and-error and re-optimisation, planning times are drastically shortened. In addition, plan quality is improved by finding and exploiting advantageous trade-offs.

Radiation therapy treatment planning is a very complex process. It is important to direct attention to not only the intensity problem itself, but also to factors influencing it. Consequently, radiation therapists should be supported in their choice of initial parameters. Here, beam direction optimisation is most critical, as the influence of beam directions on the resulting dose distribution is substantial and the choice of beam directions often non-intuitive [17].

The support through sensitivity analysis should be expanded. The idea is to initialise a search program that extracts advantageous trade-offs from the database of pre-computed treatment plans. As a result, CARINA could propose treatment plans based on such a search outcome.

Acknowledgements

This research was supported by a postgraduate scholarship for the second author from The University of Auckland's Vice Chancellor's Development Fund, Project 23102. The implementation of CARINA was based on an implementation of (4) at the Fraunhofer Institute for Industrial Mathematics, Kaiserslautern, Germany. We are grateful for the permission to use parts of this programme.

References

[1] M. Alber, M. Birkner, and F. Nüsslin. Tools for the analysis of dose optimisation: II. Sensitivity analysis. *Physics in Medicine and Biology*, 47:N265–N270, 2002.

[2] M. Alber and F. Nüsslin. An objective function for radiation treatment optimisation based on local biological measures. *Physics in Medicine and Biology*, 44:479–493, 1999.

[3] M. Alber and F. Nüsslin. A representation of an ntcp function for local complication mechanisms. *Physics in Medicine and Biology*, 46:439–447, 2001.

[4] S. Apisarnthanarax and K.S.C. Chao. Current imging paradigms in radiation oncology. *Radiation Research*, 163:1–25, 2005.

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

[6] G. Bednarz, D. Michalski, P.R. Anne, and R.K. Valicenti. Inverse treatment planning using volume-based objective functions. *Physics in Medicine and Biology*, 49:2503–2514, 2004.

[7] S. Bellavia, M. Macconi, and B. Morini. An interior point Newton-like method for nonnegative least squares problems with degenerate solution. Technical report, Dipartimento di Energetica, University of Florence, 2005.

[8] S.C. Billups and J.M. Kennedy. Minimum-support solutions for radiotherapy planning. *Annals of Operations Research*, 119:229–245, 2003.

[9] T. Bortfeld. Optimized planning using physical objectives and constraints. *Seminars in Radiation Oncology*, 9:20–34, 1999.

[10] T. Bortfeld, J. Stein, and K. Preiser. Clinically relevant intensity modulated optimization using physical criteria. In D.D. Leavitt, editor, *XIIth International Conference on the Use of Computers in Radiation Therapy*, pages 1–4, Salt Lake City, Utah, U.S.A., 1997. Madison Medical Physics Publishing.

[11] A. Brahme and A. K. Agren. Optimal dose distribution for eradication of heterogeneous tumours. *Acta Oncologica*, 26:377–85, 1987.

[12] Y. Censor. Mathematical optimization for the inverse problem of intensity-modulated radiation therapy. In J.R. Palta and T. R. Mackie, editors, *Intensity-Modulated Radiation Therapy: The State of The Art*, Medical Physics Monograph No. 29, pages 25–49. Medical Physics Publishing, Madison, Wisconsin, USA, 2003.

[13] C. Cotrutz, M. Lahanas, C. Kappas, and D. Baltas. A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. *Physics in Medicine and Biology*, 46:2161–2175, 2001.

[14] S. M. Crooks and L. Xing. Application of constrained least-squares techniques to IMRT treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, 54(4):1217–1224, 2002.

[15] I. Das and J.E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14:63–69, 1997.

[16] M. Ehrgott. *Multicriteria Optimization*. Springer Verlag, Berlin, 2nd edition, 2005.

[17] M. Ehrgott, A. Holder, and J. Reese. Beam selection in radiotherapy design. Technical Report 95, Department of Mathematics, Trinity University , USA, 2005. Available online at `http://lagrange.math.trinity.edu/tumath/research/techreport.shtml`.

[18] M. Ehrgott and R. Johnston. Optimisation of beam directions in intensity modulated radiation therapy planning. *OR Spectrum*, 25(2):251–264, 2003.

[19] M. Ehrgott and S. Ruzika. An improved ε-constraint method for multiobjective programming. Report in Wirtschaftsmathematik http://kluedo.ub.uni-kl.de/volltexte/2005/1893/, Department of Mathematics, Technical University of Kaiserslautern, 2005.

[20] M. Ehrgott and D. Tenfelde-Podehl. Computation of ideal and nadir values and implications for their use in MCDM methods. *European Journal of Operational Research*, 151(1):119–131, 2003.

[21] M.C. Ferris, R.R. Meyer, and W.D. D'Souza. Radiation treatment planning: Mixed integer programming formulations and approaches. Technical Report 02-08, Computer Sciences Department, University of Wisconsin, USA, 2002.

[22] M. Fippel, M. Alber, M. Birkner, W. Laub, F. Nüsslin, and I. Kawrakow. Inverse treatment planning for radiation therapy based on fast Monte-Carlo dose calculation. In A. Kling, F. Barao, M. Nakagawa, L. Travora, and P. Vaz, editors, *Monte Carlo 2000 Conference*, Lisbon, 2000. 217–222.

[23] J.M. Galvin, G. Ezzell, A. Eisbrauch, C. Yu, B. Butler, Y. Xiao, I. Rosen, J. Rosenman, M. Sharpe, L. Xing, P Xia, T. Lomax, D.A. Low, and J. Palta. Implementing IMRT in clinical practice: A joint document of the American Society for Therapeutic Radiology and Oncology and the American association of physicists in medicine. *International Journal of Radiation Oncology, Biology, Physics*, 58(5):1616–1634, 2004.

[24] A.M. Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications*, 22:618–630, 1968.

[25] O.C.L. Haas. *Optimization and control systems modelling in radiotherapy treatment planning.* PhD thesis, Coventry University, UK, 1997.

[26] H.W. Hamacher and K.-H. Küfer. Inverse radiation therapy planing – A multiple objective optimization approach. *Discrete Applied Mathematics*, 118(1-2):145–161, 2002.

[27] A. Holder. Partitioning multiple objective optimal solutions with applications in radiotherapy design. Technical Report 54, Department of Mathematics, Trinity University, San Antonio, USA, 2001. To appear in OPTE.

[28] A. Holder. Designing radiotherapy plans with elastic constraints and interior point methods. *Health Care Management Science*, 6:5–16, 2003.

[29] A. Holder. Radiotherapy treatment design and linear programming. In M.L. Brandeau, F. Sainfort, and W.P. Pierskalla, editors, *Operations Research and Health Care*, pages 741–774. Kluwer Academic Publishers, Norwell MA, 2004.

[30] R. Jeraj, P.J. Keall, and J.V. Siebers. The effect of dose calculation accuracy on inverse treatment planning. *Physics in Medicine and Biology*, 47:391–407, 2002.

[31] L. Kenny and M. Lehman. Sequential audits of unacceptable delays in radiation therapy in Australia and New Zealand. *Australasian Radiology*, 48(1):29–34, 2004.

[32] K.-H. Küfer and H.W. Hamacher. A multicriteria optimization approach for inverse radiotherapy planning. In W. Schlegel and T. Bortfeld, editors, *XIIIth International Conference on the Use of Computers in Radiation Therapy*, pages 26–28, Heidelberg, Germany, 2000. Springer Verlag, Berlin.

[33] K.-H. Küfer, A. Scherrer, M. Monz, F. Alonso, H. Trinkaus, T. Bortfeld, and C. Thieke. Intensity-modulated radiotherapy – A large scale multi-criteria programming problem. *OR Spectrum*, 25:223–249, 2003.

[34] M. Lahanas, E. Schreibmann, and D. Baltas. Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Physics in Medicine and Biology*, 48:2843–2871, 2003.

[35] E.K. Lee, T. Fox, and I. Crocker. Integer programming applied to intensity-modulated radiation therapy treatment planning. *Annals of Operations Research*, 119:165–181, 2003.

[36] E. Marchi and J.A. Oviedo. Lexicographic optimality in the multiple objective linear programming: The nucleolar solution. *European Journal of Operational Research*, 57:355–359, 1992.

[37] Radiation Therapy Oncology Group. Active protocols. Website accessed 28/01/2006. `http://www.rtog.org/members/active.html`.

[38] 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. *Physics in Medicine and Biology*, 48:3521–3542, 2003.

[39] H.E. Romeijn, J.F. Dempsey, and J.G. Li. A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, 49:1991–2013, 2004.

[40] I. Rosen, H.H. Liu, N. Childress, and Z. Lia. Interactively exploring optimized treatment plans. *International Journal of Radiation Oncology, Biology, Physics*, 61(2):570–582, 2005.

[41] I.J. Rosen, R.G. Lane, S.M. Morrill, and J. Belli. Treatment planning optimisation using linear programming. *Medical Physics*, 18(2):141–152, 1991.

[42] W. Schlegel. New technologies in conformal radiation therapy. In H.J. Feldmann, P. Kueschaurek, and M. Molls, editors, *Three-dimensional Radiation Treatment. Technical Innovations and Clinical Results*, volume 34 of *Frontiers in Radiation Therapy and Oncology*, pages 26–39. Karger, Basel, 2000.

[43] W. Schlegel and A. Mahr. *3D-Conformal Radiation Therapy: A Multimedia Introduction to Methods and Techniques*. Springer Verlag, Heidelberg, 2002. Springer Verlag, Berlin.

[44] D.M. Shepard, M.C. Ferris, G.H. Olivera, and T.R. Mackie. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, 41(4):721–744, 1999.

[45] S. V. Spirou and C.-S. Chui. A gradient inverse planning algorithm with dose-volume constraints. *Medical Physics*, 25(3):321–333, 1998.

[46] R.E. Steuer. *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley & Sons, New York, NY, 1985.

[47] B.W. Stewart and M.D. Kleihues. *World Cancer Report.* IARC Press, Lyon, 2003.

[48] C. Thieke. *Multicriteria optimisation in inverse radiotherapy planning.* PhD thesis, Ruprecht–Karls–Universität Heidelberg, Germany, 2003.

[49] Varian Medical Systems: Annual Report 2002. IMRT - targeting cancer. Report available online, 2002. `http://www.varian.com/comp/2002/var_02_imrt.pdf`.

[50] Varian Medical Systems, Inc. The advantages of millenium multileaf technology. Website accessed 28 January 2006, 2003. `http://www.varian.com/orad/prd056.html`.

[51] C. Wang, J. Dai, and Y. Hu. Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Physics in Medicine and Biology*, 48:4065–4076, 2003.

[52] S. Webb. *The Physics of Conformal Radiotherapy.* Iinstitute of Physics Publishing, 1997.

[53] S. Webb. *Intensity-modulated radiation therapy (Series in Medical Physics).* Institute of Physics Publishing, 2001.

[54] J.R. Williams and D.I. Thwaites. *Radiotherapy Physics: In Practice.* Oxford University Press, UK, 2nd edition, 2000.

[55] L. Xing, J. G. Li, S. Donaldson, Q. T. Le, and A. L. Boyer. Optimization of importance factors in inverse planning. *Physics in Medicine and Biology*, 44:2525–36, 1999.

[56] Y. Yang and L. Xing. Clinical knowledge-based inverse treatment planning. *Physics in Medicine and Biology*, 49:5101–5117, 2004.