

Lexicographic ordering: intuitive multicriteria optimization for IMRT

Kyung-Wook Jee, Daniel L McShan and Benedick A Fraass

Department of Radiation Oncology, University of Michigan, UH-B2C432, Box 0010,
1500 E. Medical Ctr. Dr., Ann Arbor, MI 48109, USA

E-mail: wook@umich.edu

Received 27 October 2006, in final form 10 January 2007

Published 7 March 2007

Online at stacks.iop.org/PMB/52/1845

Abstract

Optimization problems in IMRT inverse planning are inherently multicriterial since they involve multiple planning goals for targets and their neighbouring critical tissue structures. Clinical decisions are generally required, based on tradeoffs among these goals. Since the tradeoffs cannot be quantitatively determined prior to optimization, the decision-making process is usually indirect and iterative, requiring many repetitive optimizations. This situation becomes even more challenging for cases with a large number of planning goals. To address this challenge, a multicriteria optimization strategy called lexicographic ordering (LO) has been implemented and evaluated for IMRT planning. The LO approach is a hierarchical method in which the planning goals are categorized into different priority levels and a sequence of sub-optimization problems is solved in order of priority. This prioritization concept is demonstrated using two clinical cases (a simple prostate case and a relatively complex head and neck case). In addition, a unique feature of LO in a decision support role is discussed. We demonstrate that a comprehensive list of planning goals (e.g., ~23 for the head and neck case) can be optimized using only a few priority levels. Tradeoffs between different levels have been successfully prohibited using the LO method, making the large size problem representations simpler and more manageable. Optimization time needed for each level was practical, ranging from ~26 s to ~217 s. Using prioritization, the LO approach mimics the mental process often used by physicians as they make decisions handling the various conflicting planning goals. This method produces encouraging results for difficult IMRT planning cases in a highly intuitive manner.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Intensity modulated radiation therapy (IMRT) plans are typically determined using a planning technique termed inverse planning (Barth 1990, Boyer *et al* 2001). Currently, the usual inverse planning paradigm consists of (1) defining the anatomical structures to be used for optimization of the plan, (2) creating beams divided into small beamlets or fluence bixels and (3) optimizing the beamlet fluences driven by an objective or cost function. Although many different types of cost functions have been described in the radiotherapy literature (Bortfeld 1999, Kessler *et al* 2005, Mohan 1996, Niemierko *et al* 1992, Wu *et al* 2000, 2003), virtually all inverse planning has been performed using a single cost function.

However, there is general consensus that while these techniques may lead to acceptable or even excellent IMRT plans, there are a number of deficiencies associated with this type of approach: (1) it is extremely difficult to select the appropriate cost function (type, parameters, etc) which will lead to what the physician considers to be the 'optimal' plan, (2) most planners continue to use a trial and error approach to choosing or adjusting cost function type and/or parameters in an effort to improve the plans obtained from the inverse planning system, (3) since there are fundamental conflicts between the need to increase dose to the tumour while decreasing dose to each of the normal tissues, the clinical tradeoffs between complication probability and tumour control probability are very difficult and not unique, especially since so few tissues have accurately known dose–volume–response relationships and (4) due to all the uncertainties described above, there may be many solutions (plans) which are very different in how they treat particular tissues while at the same time being potentially very similar in the overall cost function value. In practice, optimizations are performed in a multiple trial-and-error manner until an acceptable solution is found.

Recently, in order to address some of these issues, *multicriteria* optimization strategies have begun to be applied to inverse planning problems. One such strategy involves preparation of many alternative solutions that are optimal in terms of the *Pareto optimality* concept (Pareto 1906). A pool of solutions that encompasses the wide range of tradeoff space is presented to a planner. Then, the planner navigates the solutions and selects the best quality plan based on a personal preference and intuition. This Pareto approach is often called an *a posteriori articulation of preference* method (Marler and Arora 2004, Messac 2002) in which the planner avoids time-consuming optimization trials while making a decision with the pre-computed alternative solutions. Research efforts are being made towards the efficient generation and navigation of Pareto optimal solutions (Bortfeld *et al* 2003, Cotrutz *et al* 2001, Craft *et al* 2005, 2006, Lahanas *et al* 2003).

For IMRT problems with the planning goals exhibiting distinctive levels of importance or priority, *a priori articulation of preference* methods such as lexicographic ordering (LO) (Marler and Arora 2004, Miettinen 1999) can provide an intuitive and efficient way of generating a plan solution. In this approach, the planning goals are categorized in order of priority, and a sequence of small problems with selected goals are optimized one level of priority at a time until the entire planning goals are addressed. By subdividing a large multicriteria problem, the complex space of tradeoffs can be significantly simplified, thereby reducing the need for iterative optimization trials. Preemptive goal programming is a similar method that also benefits from the intuitiveness of planning goal/objective prioritization; this concept was introduced to the radiation oncology community by Langer *et al* (2003). The present study demonstrates the first successful implementation of LO for clinical IMRT planning.

In this paper, the implementation of the LO method is described and its efficacy for IMRT planning is demonstrated using a straightforward prostate case and a relatively complex

head and neck IMRT case. Various implications of using the LO method for direct inverse planning as well as its use for a decision-supporting role in plan optimization are described and discussed.

2. Methods and materials

2.1. Lexicographic ordering method

The conventional method used to solve IMRT optimization problems minimizes a mathematical function (the *objective* or *cost function*) whose minimum hopefully corresponds to the desired planning result, i.e., the ‘optimal’ solution. For a given planning case, since multiple planning criteria associated with many tissue structures are to be optimized, the cost function in the plan optimization, $F(x)$, is a collection (i.e., vector) of individualized functions, $f_i(x)$ corresponding to each planning criterion. The general plan optimization problem thus can be posed as a multicriteria optimization (MCO) problem:

$$\begin{aligned} &\text{Find } x \in \mathfrak{R}^n \text{ minimizing } F(x) = [f_1(x), f_2(x), \dots, f_N(x)]^T, \\ &\text{subject to} \\ &g_j(x) \leq 0; \quad j = 1, 2, \dots, m \quad \text{and} \\ &h_k(x) = 0; \quad k = 1, 2, \dots, e, \end{aligned} \quad (1)$$

where N , m and e are the total numbers of objective functions, inequality and equality constraints, respectively. The vector $x \in \mathfrak{R}^n$ represents n design or control variables that are optimized during the planning. In the case of IMRT, those variables are the intensities of beamlets. Although a plan solution that simultaneously optimizes all $f_i(x)$ is most desirable, except for trivial cases, such a solution is not usually attainable. Instead, the best-compromised solution based on the planner’s implicit *preferences* towards individual planning criteria is sought through tradeoffs. The need for a decision-maker’s intervention in the optimization process is known to be a primary characteristic of MCO problems (Steuer 1985).

In IMRT planning, the conventional approach for dealing with the MCO aspect of the problem is to use a *scalarization* approach, in which multiple objective functions are combined to form a single function. Due to their simplicity and effectiveness, these methods are predominantly used not only for RT problems, but also for general engineering and scientific MCO problems (Marler and Arora 2004). One of the most common scalarization approaches is the weighted-sum method:

$$\min f_{\text{total}}(x) = \sum_{i=1}^N w_i f_i(x). \quad (2)$$

where the w_i are weight factors assigned to each objective function (‘costlet’, Kessler *et al* (2005)) in order to articulate the planner’s preferences. The best values of the weight factors cannot be easily determined since (a) the numerical quantities of all $f_i(x)$ are typically not based on a uniform scale; (b) the number of trading objective functions can be large; and (c) the consequences of a given tradeoff cannot be quantitatively known prior to the optimization.

Lexicographic ordering (LO) is a MCO method that can potentially avoid the use of the weight factors by incorporating priorities of the individual planning criteria (objective functions) explicitly in the optimization process. The LO method solves a sequence of small optimization problems in order of priority as follows:

$$\begin{aligned} &\min f_i(x) \\ &\text{subject to } f_j(x) \leq f_j(x_j^*) \\ &\text{where } i = 1, 2, \dots, N, \quad j = 1, 2, \dots, i - 1 \quad \text{if } i > 1. \end{aligned} \quad (3)$$

The first step in this method is to categorize the objective functions into different levels based on their importance, the highest level ($i = 1$) being the most important. Then, a conventional search algorithm (a nonlinear constrained optimization algorithm in our implementation) is used to solve one level of optimization problem after another, starting at level 1 ($i = 1$). While the method progresses down from level 1 to level N (the final level), the preceding objective functions are converted to inequality constraints, with boundary values, $f_j(x_j^*)$, set by the *a priori* attained solutions, $x_j^* = \arg \min f_j(x)$ subject to the constraints from the upper level. Accordingly, the number of constraints increases up to $N - 1$ at the final level of optimization. When more than one planning objectives are identified as a similar priority, it is possible to combine them into the same level by a linear combination (typically with unity weights), as in equation (2). After minimizing the composite function, the individual objective functions are turned into constraints for the subsequent levels of optimization. This approach allows the total number of priority levels (hence the total number of optimization runs) to be reduced for a given planning case. The LO method is often referred to as a *feasible space reduction* method, since the feasible space is gradually reduced with gradually added constraints as the LO method progresses.

2.2. Optimization system and implementation

All the work presented here, including the LO-based optimization, has been performed using our in-house developed IMRT plan optimization system ('UMPlan/UMOpt') (Kessler *et al* 2005, Vineberg 2002, Jee *et al* 2005). UMOpt supports a general framework of IMRT inverse planning for external beam radiotherapy including various types of cost functions, a number of search algorithms and other features. For routine IMRT planning, including all work shown here, a convolution/superposition dose calculation algorithm, originally based on the work of Mackie *et al* (1985) but heavily modified for IMRT (McShan *et al* 2006), is used. During optimization, the doses to the regions of interest, D_v , are calculated as follows:

$$D_v = \sum_{l=1}^n T_{l,v} x_l \quad (4)$$

where x_l is the intensity of the l th beamlet (from a total of n beamlets) and $T_{l,v}$ represents a pre-computed coefficient matrix that quantifies a dose contribution from the l th beamlet with unity intensity to the v th dose calculation point. Typically, the dose calculation points are randomly distributed throughout the patient tissue structures. Based on the dose distribution, UMOpt uses a comprehensive set of planning metrics to compute objective function values, $f_i(D_v(x))$. These metrics include those related to straightforward dose statistics (min, max, mean), dose–volume relationships (DVH, uniformity) and biological model-based metrics (TCP, NTCP, EUD, V_{eff}). A detailed description of the planning metrics in UMOpt can be found in Kessler *et al* (2005).

Implementation of the LO method in UMOpt required two important additional software components: an efficient search algorithm that can explicitly handle nonlinear constraints and a flexible method of calculating high quality Jacobians for interactively created objective and constraint functions.

Since IMRT planning generally poses large-scale optimization problems (>300 control variables, i.e., beamlets) with nonlinear planning metrics, a large-scale nonlinear programming (NLP) algorithm called sequential quadratic programming (SQP) (Nocedal and Wright 1999, Gill *et al* 2002) has been used. The general idea of SQP is based on modelling (or approximating) nonlinear-constrained optimization problems by quadratic programming (QP) sub-problems and searching with directions obtained from the minimization of sub-problems.

The particular SQP implementation used in this study employs a BFGS reduced-Hessian algorithm (Broyden 1970, Shanno 1970) eliminating complex computations of the Hessian matrixes and a globalization strategy called the *line search* method (see Gill *et al* (2002) for a detailed theory of SQP).

Since the numerical performance of SQP depends significantly on the accuracy of Jacobians for objective and constraint functions, they are computed exactly using an automatic differentiation (AD) algorithm called ADOL-C (Griewank *et al* 1996, Jee *et al* 2005). AD algorithms automatically compute derivatives (in any order) of mathematical functions given as a computer program. Since any function in a program, no matter how complicated, executes a sequence of elementary arithmetic operations, by applying the chain rule repeatedly to these operations, derivatives can be computed automatically and accurately to machine-level precision (Griewank *et al* 1996). Specifically, ADOL-C was chosen to facilitate and expedite implementation of complex optimization metrics free of errors in derivatives by taking advantage of the operator-overloading mode of ADOL-C. Early results indicate that highly accurate derivatives provided by the AD algorithm are essential to obtain both feasible and optimal solutions for each level of LO optimization in a practical optimization time (Jee *et al* 2005) (see also table 3).

2.3. IMRT cases

The LO approach is demonstrated using IMRT planning for two cases: (1) a simple prostate case and (2) a clinical head and neck case. The simple case is used merely to demonstrate the basic behaviour of the LO method, and consists of a prostate structure (representing a target volume) and a bladder and a rectum (representing critical structures) surrounded by four oblique-angled 6 MV fields. These fields are subdivided into ~ 570 beamlets with each having a dimension of $5 \times 5 \text{ mm}^2$. The planning goals used in this example are arbitrarily chosen to best illustrate the basic functionality of the LO method and are described in section 3.

The clinical head/neck case is shown here as an example of a full complex clinical case. The goals of treatment are described in detail in section 3 while illustrating the LO method, but include multiple competing goals, including sparing non-involved tissues (especially the parotids) while also delivering very conformal, homogeneous and high dose to the bilateral neck targets (Eisbruch *et al* 2001, Eisbruch 2002a, 2002b). The prescribed dose for the primary targets (denoted as PTV70) is 70 Gy. In the contralateral neck, low risk nodal chains (PTV60) were to be treated to 60 Gy, while higher risk nodal areas (PTV64) were prescribed to 64 Gy. Detailed dosimetric requirements for target volume homogeneity and other requirements are described in the example. Organs-at-risk (OAR) or critical normal structures included the brainstem, spinal cord, mandible and parotids. The skin surface and tissue outside of all OARs were designated as unspecified normal tissues.

Both normal and target definitions were based on contours drawn on images from CT and MR with the patient immobilized using an alpha cradle and thermoplastic mask. The IMRT plan developed for this case included five coplanar beam fields that were arranged axially around the patient with equal spacing. Each 6 MV field was subdivided into $1 \times 1 \text{ cm}^2$ beamlets (a total of 1023 beamlets). Approximately 30 000 dose calculation points were randomly distributed throughout the various target and normal tissue structures. Dose calculations were performed using the aforementioned convolution/superposition algorithm with density corrections based on CT number.

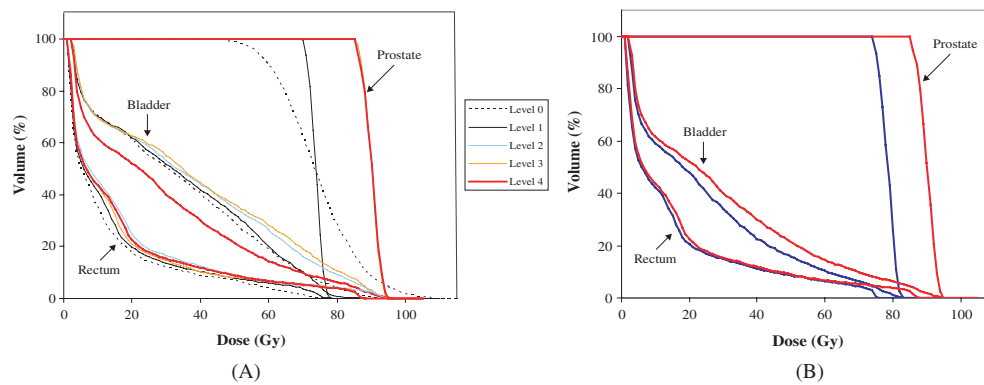


Figure 1. Illustrations of the basic LO method using a prostate example. (A) DVHs for PTV, rectum, bladder are shown for the four levels of optimization. The DVH plot from the last level (level 4) represents a final planning solution. PTV curves from levels 2, 3 and 4 are almost identical and shown overlapped. (B) The final solution from (A) (plotted in red) is compared to a solution obtained using the LO method based on a new priority order which describes the rectal wall dose as the most important issue, rather than target coverage (plotted in blue).

Table 1. (A) LO prioritization for the prostate example (for figure 1(A)) and (B) alternative LO prioritization (for figure 1(B)).

Planning criteria for a prostate IMRT case	
(A) LO prioritization (for figure 1(A))	
Level 1	Uniform irradiation of prostate PTV to $\pm 5\%$
Level 2	Escalation of PTV mean dose to 90 Gy
Level 3	Minimization of rectum dose using conventional five dose–volume criteria
Level 4	Minimization of overall bladder dose
(B) Alternative LO prioritization (for figure 1(B))	
Level 1	Limiting the maximum rectal dose below 75 Gy in addition to satisfying five point dose–volume criteria
Level 2	Uniform irradiation of prostate PTV to $\pm 5\%$
Level 3	Escalation of PTV mean dose to 90 Gy
Level 4	Minimization of overall bladder dose

3. Results

3.1. Basic functionality of the LO method

The basic behaviour of LO-based inverse planning is illustrated in figure 1 for a simple prostate example involving a PTV (prostate) and two neighbouring critical organs (rectum and bladder). For this first example, the physician identified the planning goals and prioritized them to four different levels (table 1(A)).

Figure 1(A) shows five sets of DVH plots obtained from each level of optimization. Prior to each optimization, the dose distributions were initialized with randomly chosen beamlet intensities and the corresponding DVH is shown as level 0. The first level optimization problem simply applied a *uniformity* planning goal for the PTV (i.e., make the target uniformity $\leq \pm 5\%$ of a *nominal* mean dose). Since optimization was exclusively performed for the PTV without having competing structures or dosimetric goals, this uniformity criterion could be readily achieved. Consequently, the bladder and rectum showed slight changes in DVHs. However,

these changes were considered irrelevant since their doses were not optimized at this level. Before proceeding to the subsequent optimization level, the achieved PTV uniformity was constrained. At the second level, the target dose distribution was further optimized, aiming to escalate the target (mean) dose to 90 Gy while the first level result ($\leq \pm 5\%$ uniformity) was kept unchanged. The PTV DVH from the second level optimization shows that the goals of the dose escalation as well as the uniformity were also fully achieved. The third level optimization constrained the results for the PTV and also minimized the rectal dose by means of five dose–volume planning criteria, i.e., the rectal volume receiving greater than 65, 70, 75, 80 and 85.2 Gy of dose should be less than 50, 35, 25, 15 and 0%, respectively. Without compromising the prior PTV results, the third level optimization achieved all of the dose–volume criteria except the (85 Gy, 0%) criterion for which a slight improvement was also observed—a decrease in volume from 3.4% to 2.8% receiving 85 Gy or greater. After converting the rectum results into five new individual constraints, finally, a general goal of reducing dose to the bladder was pursued by minimizing its mean dose. The mean dose of bladder before and after the final optimization was found to be ~ 37.9 and ~ 28.6 Gy, respectively, showing significant dose sparing. A solution generated after addressing all planning goals for all levels represents the final planning solution (level 4).

It is important to point out that when multiple conflicting goals are pursued in the LO-based optimization, the achievement levels of the higher priority goals always dominate those of the lower priority goals. This is due to the fact that optimized results from the higher levels set hard constraint boundaries that limit the solution pool searchable by the subsequent low-level optimizations. For example, the final solution shown in figure 1(A) will be different if the most important clinical issue were keeping the anterior rectal wall dose < 75 Gy. The consequence of this new prioritization can be demonstrated by adding a new highest-level goal that keeps the anterior rectal wall dose < 75 Gy, as in table 1(B). Figure 1(B) shows the final DVHs based on this new priority order. The result is plotted together with that following the original order (table 1(A)) for comparison. The result from the new priority order shows that the rectum max dose goal was fully achieved (~ 75 Gy). However, this achievement was followed by an expensive tradeoff for the lower-priority target dose goals. For example, when the target mean dose was escalated with a set goal of 90 Gy while constrained by the rectal dose limit of 75 Gy, the highest mean dose achievable was found to be ~ 78.4 Gy. These two goals are clearly competing, and the one placed on a higher priority will dominate the achievement level for the other. If some planning goals do not show distinctive priorities, they should be grouped together in a single level and optimized simultaneously, seeking for a middle ground in terms of the goal achievements.

3.2. Head and neck IMRT case

The intuitive nature of the LO approach is further demonstrated with the planning of a realistic head/neck IMRT case.

The main treatment protocol is the delivery of high doses to the various planning target volumes (PTVs) with the exclusion of non-involved tissue structures as much as possible. This case involves primary tumours that are confined to one side of the head/neck (see figure 2). The ipsilateral jugular nodes, containing clinical or radiological evidence of metastases, are also included in the primary target volumes, which are prescribed a dose of 70 Gy (this planning target volume is denoted as PTV70). Due to the strong link between local recurrence and the first echelon nodes with perineural invasion (Eisbruch *et al* 2004), these nodal areas are considered high risk and treated with 64 Gy (hence, PTV64). In the contralateral neck, left and right lymph node chains at risk of subclinical metastases were

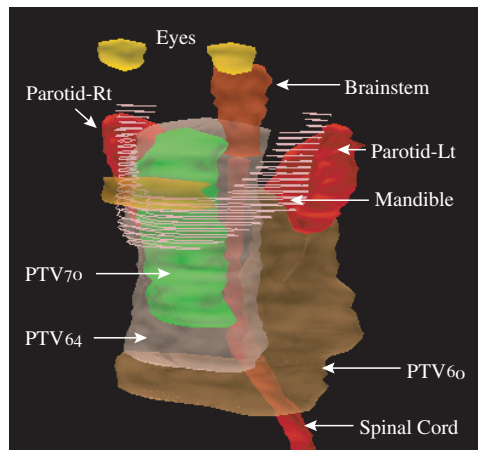


Figure 2. Patient anatomy used in the example head/neck case.

Table 2. Protocol criteria used in the head/neck example.

Dose–volume criteria for PTVs	
Prescription dose for PTVs	Minimum doses of 70, 64, 60 Gy should be given to PTV70, PTV64, PTV60, respectively 110% of the Rx dose of PTV should be given to <0.5 cc
Dose constraints for organs-at-risk	
Brainstem	0% of the volume is allowed to receive >54 Gy
Spinal cord	0% of the volume is allowed to receive >45 Gy
Lips	0% of the volume is allowed to receive >60 Gy
Mandible	0% of the volume is allowed to receive >70 Gy or 105% of the Rx dose to PTV70
Unspecified tissue outside PTVs (oral cavity, normal tissues)	Maximum doses should be $\leq 105\%$ of the Rx dose to PTV70 Maximum doses should be $\leq 105\%$ of the Rx dose to PTV70
Planning criteria for salivary (parotid) glands	
Parotids	Mean dose should be <26 Gy

identified as secondary targets and prescribed to 60 Gy (PTV60). For each of these clinical target volumes (CTVs), a PTV was defined using the current CTV-PTV margin applied in our clinic (3 mm for patients treated with daily on-line imaging and repositioning (Marsh *et al* 1997)). Other target volume dose homogeneity and minimum/maximum dose limitations were also specified, as described in table 2. For this example, a number of organ-at-risk (OAR) structures were included in the optimization, including brainstem, spinal cord, mandible and parotids. Based on clinical correlations of dose to parotid versus saliva production (Eisbruch *et al* 2004), limiting the mean dose to parotid to 26 Gy is considered an important goal. The skin surface and all normal tissue outside all other OARs were designated as unspecified normal tissues. The maximum tolerable and other required dosimetric constraints are also summarized in table 2.

In spite of the simplification of this planning protocol for the current example, there are still approximately 23 separate planning goals for targets and normal tissues. In order to illustrate the application of the LO method, it is important to prioritize the planning goals,

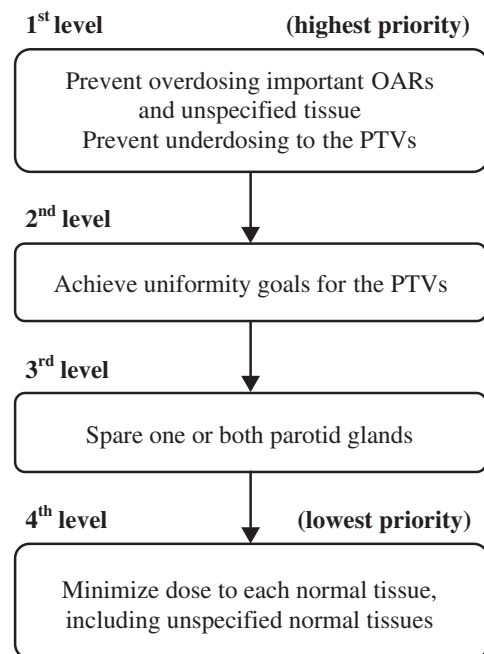


Figure 3. A flowchart showing the priority levels used in the LO-based optimization for the head/neck example. Specific goal levels for the planning criteria shown in the chart are detailed in table 2.

just as the physician typically does in the clinic. For this example, we use our present clinical head/neck protocol guidelines, as they clearly delineate the priority of the various goals. As illustrated in figure 3, the prevention of overdosing of OARs and/or underdosing of PTVs are considered the most important goals to achieve, therefore constituting the first priority level. Maintaining the uniformity of dose within each of the PTVs is pursued in the second level. The third priority level goal is to force the mean dose to one or both parotids below the 26 Gy threshold dose, if possible. Finally, although only an implicit part of the planning protocol, the final level of priority is to minimize the dose to each normal tissue, including the non-specified normal tissue.

Figure 4 shows the DVHs obtained with the LO-based optimization for the head and neck IMRT example. A total of 23 planning goals specified for this particular case are categorized into four different priority levels. Since these goals are primarily based on dose statistics and dose–volume metrics (table 2), optimizations were performed using dose-based (specifically, mean dose) and dose–volume objectives only. (Note that the use of the LO method is not limited to dosimetric metrics and can be used with a wide range of planning metrics including biological-models, as demonstrated in section 3.4.) The DVH plots as a function of optimization level illustrate the gradual process of the LO-based planning. In all of the plots, DVH curves highlighted in thicker lines correspond to particular structures that are specifically involved in optimization. In figure 5, changes in the plan’s dosimetric metrics that are directly involved in optimization are illustrated as the LO-method proceeds from the initial down to the final levels.

The first optimization level searched for a solution that makes the entire tissue volume (100%) receive doses below the maximum tolerable limits for the OAR and the normal tissue structures as well as fully delivering the prescription doses to the three target volumes. Since

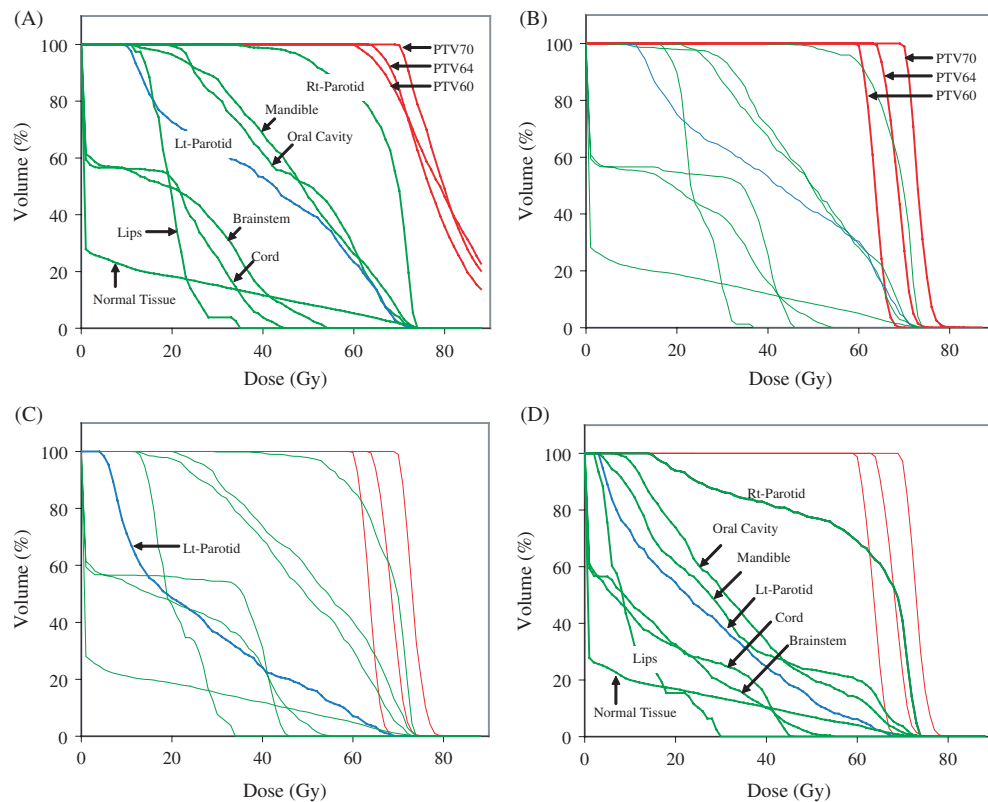


Figure 4. Plots of DVH curves generated by the LO-based optimization for the head/neck case shown in figure 2. Plots (A), (B), (C) and (D) correspond to DVH curves obtained from optimization levels 1, 2, 3 and 4, respectively, illustrating a gradual change of DVH curves as a function of the LO level. Particular structures involved in optimization are plotted in the thicker lines.

the target doses are prescribed to the minimum (min) doses of the PTVs in this particular protocol, the DVH objectives were minimized to assure that 100% of the target volumes receive doses greater than the prescribed doses. In this level, the multiple (11 goals in this case) objectives were summed together with each having a unity weight and then minimized until an optimal solution was found. Although the target and OAR goals could potentially compete with each other, the dosimetric results in figure 4(A) show that a plan solution that satisfies all of the first level goals exists. All OAR and normal tissues receive dose below the limit (73.5 Gy), with doses of 44.4, 53.7 and 34.8 Gy for the spinal cord, brainstem and lips, respectively. Min doses to PTV60, PTV64 and PTV70 are found to be 60, 64.1 and 70 Gy, respectively. In order to keep the achieved results unchanged for the subsequent levels of optimization, the objective functions were then turned into hard constraints with each boundary set by its optimized function value before proceeding to the second level.

The second level optimization addresses uniformities of the target dose distributions. Uniform distributions are achieved by penalizing high doses given to the PTVs above the maximum limits while keeping the min dose results from the first level constant. The specified uniformity in the present protocol is below 10% of the prescription (min) doses given to the less than 0.5 % of the target volumes. The second level optimization results (figure 4(B)) show that approximately 3, 21 and 11% of PTV70, PTV64 and PTV60 are given higher than the

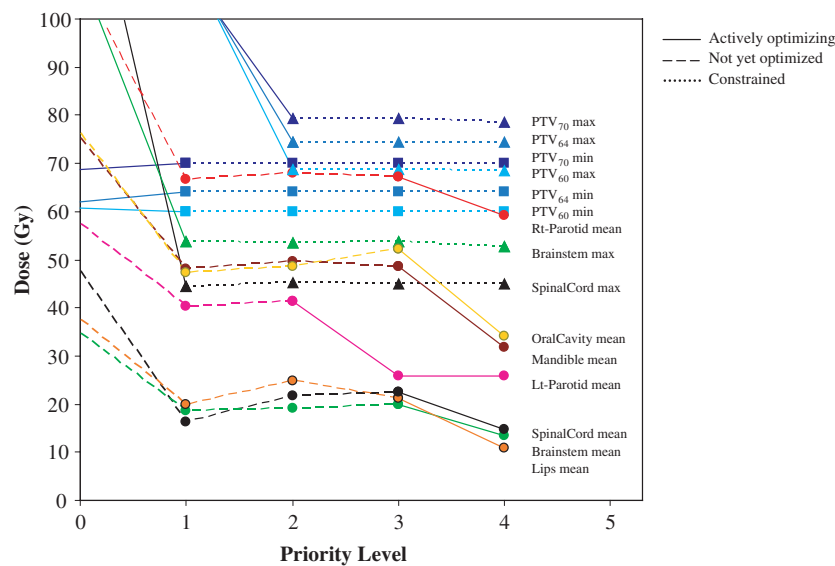


Figure 5. Dosimetric metrics plotted as a function of optimization level for the head/neck example. Symbols indicate the optimized results obtained from each designated priority level. Triangles, squares and circles represent maximum, minimum and mean dose values, respectively. Solid lines show changes forced by optimization that directly tried to change metrics, while dashed and dotted lines show changes during level optimizations that did not directly involve the structure. (All maximum and minimum dose values were calculated by averaging the highest and lowest doses received by $\sim 1\%$ of volume, respectively.)

upper dose limits (77, 70.4 and 66 Gy) and the actual max doses were found 79.4, 74.5 and 68.7 Gy, respectively. Although the prescribed uniformity could not be fully achieved, the results were accepted given that *these results cannot be further improved without compromising the higher priority results*. It is worth noticing from figures 4(B) and 5 that the optimization results from the first level are tightly constrained while the second level planning goals are being optimized.

After converting the uniformity objectives into constraints, organ sparing for the contralateral (patient-left) parotid is attempted by minimizing the mean parotid dose below 26 Gy (figure 4(C)). Figure 5 also indicates that the mean dose has been reduced significantly from ~ 41.4 Gy to ~ 25.9 Gy before and after the third level optimization. The constant level dotted lines in figure 5 again indicate that the dosimetric results for all other structures are tightly held unchanged.

Finally, at the fourth optimization level, a general goal of dose reduction was pursued for all non-target structures (figure 4(D)). This goal was achieved by minimizing the mean doses for the non-target tissues (eight structures in this case). As illustrated in figure 5, decreases in the mean doses are most significant (over ~ 10 Gy) for oral cavity, mandible and lips, moderate (over ~ 5 Gy) for spinal cord, brainstem and rt-parotid, and relatively small for Lt-parotid and unspecified normal tissues. The final dose distributions achieved are shown in figure 6.

3.3. Algorithmic performance

In the present system, the LO method is implemented with a gradient-based search algorithm (SQP). This gradient-based algorithm exhibits an extremely high convergence speed and

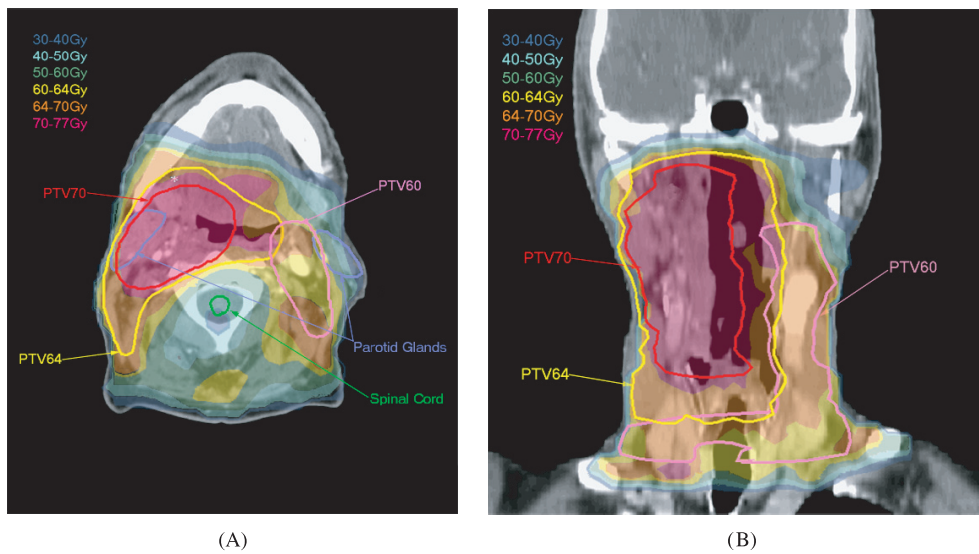


Figure 6. The final dosimetric distributions for the head/neck example are shown as dose colourwashes in (A) (axial) and (B) (coronal) CT cuts. The three target volumes—PTV60, PTV64 and PTV70—are tightly bounded by highly conformal and uniform doses. In addition, steep dose gradients are evident near the peripheral region of the targets, especially the contralateral parotid gland and the spinal cord.

Table 3. Algorithmic performances for LO-based optimization of the example head/neck case. The total SQP time represents time used by the SQP algorithm excluding function and Jacobian calculation times.

	Level 1	Level 2	Level 3	Level 4
Number of planning goals	11	3	1	8
Number of objective functions	1	1	1	1
Number of constraints	0	11	14	15
Number of major iterations	60	30	36	98
Number of total function calls	62	37	48	194
Single function calc. time* (ms)	110	110	110	110
Single Jacobian calculation time* (ms)	210	380	400	460
Total SQP time* (s)	4	23	28	84
Total optimization time* (s)	26	42	54	217

Timing benchmark* was performed on a Digital Alphastation with 1.1 GHz CPU and 16 GByte of memory.

therefore produces an IMRT plan solution in a practical timeframe. Algorithmic performances measured for the head/neck case (table 3) clearly demonstrate the reasonable time required to use the LO method for clinical cases. The table shows optimization time needed for each level ranges from 26 to 217 s and increases as the method proceeds down to lower levels. This is due to increases in the number of constraints and in the time required to calculate the corresponding Jacobian matrixes. With increasing level number, it was also observed that SQP required more iterations as well as more function calls before it converges and finds an optimal and feasible solution.

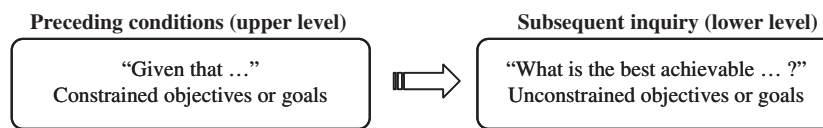


Figure 7. Diagram illustrating an example use of lexicographic ordering for decision support based on a bi-level approach.

3.4. Clinical decision support

In radiotherapy planning, many clinical questions can be posed as a combination of the preceding constraints and the subsequent inquiry. The LO method can provide an efficient way of addressing them since the method uses both constraints and objectives to implement the *objective prioritization*. For example, one may choose a bi-level prioritization strategy to assess clinical questions by placing the preceding constraints at the upper level and the subsequent inquiry at the lower level, as illustrated in figure 7.

A decision-support role for the LO method can be illustrated using the head and neck case described earlier. In this case, the goal of sparing the parotid glands was attempted in order to prevent xerostomia. Since a large amount of irradiation to the parotid glands typically cannot be avoided, how parotid sparing affects target coverage or critical tissue toxicity can be very useful information. A series of clinical questions can be posed to help assess parotid involvement. For example, one may inquire: (i) while the local tumour control is pursued, can both parotids can be spared? (ii) if not, can at least one parotid be protected? and (iii) if not, what would be the least damaging tradeoff of other important planning criteria that improves the parotid condition?

For this example, it has been already shown in figure 4(C) that the third level optimization can result in adequate sparing of the (contralateral) lt-parotid (<26 Gy mean dose), which answers question (ii). In order to find a definitive answer for question (i), one can perform the following optimization in a bi-level LO setting: the preceding constraints (e.g., the min and max dose constraints for the tumour volumes and the OAR, respectively) optimized at the upper priority and the subsequent inquiry performed using the mean dose criteria for *both* parotid glands. This bi-level LO optimization resulted in mean doses of 56.5 and 27.8 Gy for rt-parotid and lt-parotid, respectively (as shown in figure 8), a failure of organ sparing for both parotid glands. Based on this information, one can easily reach a conclusion that, *for this particular patient and beam geometry*, it is sensible to pursue sparing of one (the contralateral) gland only and look for potential benefits of sparing other OAR tissues.

Another example of the LO-based decision support can be shown using a dose escalation protocol. Since escalating the target doses is directly competing against sparing the OAR and normal tissue doses, information about the highest achievable target doses that concurrently provide adequate protections for all critical tissues can be useful in determining the treatment dose. The LO method can provide this information by addressing dose constraints for OAR and normal tissues at the upper level and maximizing of the target doses at the lower level.

Results in figures 9(A) and (B) illustrate an example use of LO to investigate dose escalation for the head/neck case. The safety of the critical structures (lt-parotid, cord and brainstem in this particular example) was first ensured by nominal NTCP limits (30, 5 and 5%, respectively) at the upper level. In addition, homogeneity criteria for the three target volumes (−5%/+10% deviations from the mean doses of the targets) were included at the same level. In the absence of the dose escalation criteria for the targets, the upper level optimization readily found a solution that satisfies all of the NTCP goals for the lt-parotid,

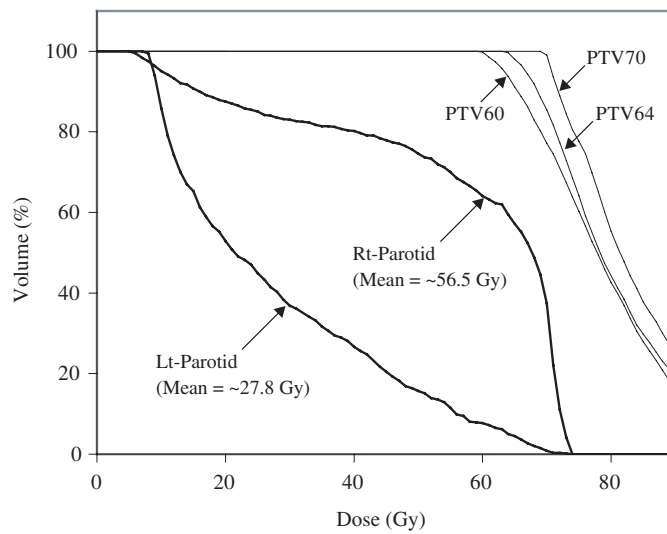


Figure 8. Optimization results for the head/neck example showing the lowest achievable mean doses for the ipsilateral (right) and contralateral (left) parotid glands (~ 56.5 and ~ 27.8 Gy, respectively) with the max and min dose constraints imposed to the normal tissues and target structures, respectively. Although included in optimization, DVH curves for other normal structures are not plotted in this figure for clarity.

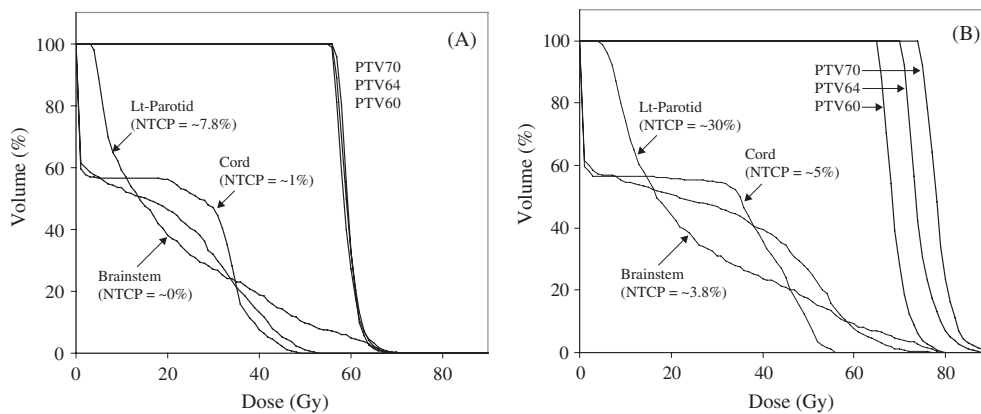


Figure 9. Illustration of LO-based planning for target dose escalation using the example head and neck case. (A) DVH curves for the upper level optimization for the critical structures (e.g., Lt-parotid, cord and brainstem) based on nominal NTCP limits (30, 5 and 5%, respectively, determined by conventional dose–volume criteria in head and neck IMRT planning protocols) and a homogeneity goal for targets. (B) Lower level optimization for the (mean) dose escalation.

cord and brainstem (resulting in ~ 7.8 , 1.0, and 0.0%) and the homogeneity goal with values well below the $-5\%/+10\%$ limits. Subsequently, the mean target doses were escalated by maximizing them at the lower level. A comparison between target results in figures 9(A) and (B) shows that the mean doses can be significantly increased up to ~ 69.0 , 74.3 and 78.4 Gy for PTV60, PTV64 and PTV70, respectively from their initial values (~ 60 Gy). As expected, the NTCP values for Lt-parotid, cord and brainstem (30.0, 5.0 and ~ 3.8 %, respectively) as

well as the target dose uniformity were kept constrained below the prescribed limits during the mean dose maximization.

4. Discussion

The task of the physician attempting to optimize a treatment plan for an individual patient is complex, often requiring clinical decisions that balance and trade off the various potential benefits and toxicities which may be associated with any particular plan of radiation therapy. One of the main approaches (used in our clinic and many others) to cope with the large number of competing clinical issues in a given problem is to prioritize the clinical issues so that the most important issues are dealt with first, attempting to make compromises only on the less important issues while completely satisfying the more important ones. This reduces the complex problem from what to do about many different decisions to several smaller ones in which it is relatively easy (or easier, at least) to state the relative importance of the various issues. Given this kind of decision-making process, it becomes intuitive to use the lexicographic ordering method to optimize the plan, since one can directly translate the priority-based decision-making into the ordered consecutive optimizations.

In LO, as illustrated in section 3.1, planning issues or goals placed in a higher level always dominate the achievements of others placed in lower levels—i.e., the higher priority goals are extremely favoured over the lower ones. Although this feature of LO helps make difficult decision problems manageable, it is important not to *overachieve* the high priority goals. One must set the optimization parameters (e.g., PTV min/max doses, OAR max doses and various DVH points) to values that are *clinically relevant*, thereby leaving enough search space for subsequent levels of optimization. When optimization results in the lower priority levels are not acceptable, a planner may choose to compromise the plan quality that is dominated by the higher priority goals. In such cases, quantitative information characterizing the tradeoff space for goals at different priorities may be highly desirable. In order to obtain such information, perhaps, a hybrid approach of the Pareto (described in section 1) and LO methods could be utilized. For example, a progressive relaxation of the high priority constraint boundaries could sample different solution points on the tradeoff space curve—a similar concept is called the *e-constraint* approach in multicriteria optimization. The feasibility of the constraint boundary relaxation has been demonstrated for IMRT planning by Alber *et al* (2002). Unlike the present Pareto approach (Craft *et al* 2005, 2006), the constraint relaxation has a potential of generating evenly distributed solution points in design (or function) parameter space. Our experiences with LO have shown us that (a) constraints in general IMRT problems can be either *actively or inactively binding* and (b) some constraints are so important that a planner may prefer not to trade with other goals. Incorporating local sensitivity (e.g., Lagrange multiplier (Alber 2002)) and planner's preferences in a strategic manner, one should be able to reduce the number of constraints/objectives to ones that are only relevant to clinical decisions. Consequently, this will limit the dimension of the tradeoff space, hence making the computation of the space more practical.

Direct comparison of the LO method to results of weighted sum cost function methods is a difficult task, and is not included in this report. The biggest problem with such a comparison is the usual difficulty for comparison of different optimization methods: since the stated goals of the optimization are determined mainly by the priority scheme (for LO) and by the cost function (for the weighted sum method), the goals of the two different methods are in general different, and so the plans resulting from the two optimizations ought to be different. The point of this LO report is not to show that it achieves 'better' plans or converges 'faster' than use of weighted sum cost functions. Rather, this is another method which allows statement of

the optimization problem in a different way, one which may be more intuitive to the physician and planner, and may more easily allow them to define the problem in such a way that they more easily achieve a plan which is acceptable to them. Study of this issue is on-going, and will be addressed in another manuscript.

5. Conclusions

IMRT optimization problems are multicriteria in nature. To address this issue, a multicriteria optimization strategy called lexicographic ordering has been implemented. Our experiences show that this LO method can facilitate an efficient articulation of individual planning goal preferences (or priorities) for IMRT planning and make complex multicriteria problems manageable by explicitly prohibiting objective tradeoffs between different priority levels.

The objective prioritization exhibited by the LO method generally reflects our intuitive mental process of making decisions to handle multiple and/or potentially conflicting planning goals. This study demonstrates that the method can be successfully used for planning clinical IMRT cases with clinically relevant metrics such as DVH, dose statistics and biological models. Although further study is necessary, it appears that this LO method can provide significant help to the planner who needs to make a sequence of decisions progressively *en route* toward an optimal plan solution in an intuitive and efficient manner.

Acknowledgments

We would like to thank Dr Edward W Larsen and Karen A Vineberg for valuable discussions and assistance. This work has been partially supported by NCI grant P01-CA59827.

References

- Alber M, Birkner M and Nusslin F 2002 Tools for the analysis of dose optimization: II. Sensitivity analysis *Phys. Med. Biol.* **47** N265–70
- Barth N H 1990 An inverse problem in radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **18** 425–31
- Bortfeld T 1999 Optimized planning using physical objectives and constraints *Semin. Radiat. Oncol.* **9** 20–34
- Bortfeld T R *et al* 2003 Intensity-modulated radiotherapy: a large scale multi-criteria programming problem *Berichte des Fraunhofer ITWM 43* (Kaiserslautern: Fraunhofer Insitut Techno- und Wirtschaftsmathematik)
- Boyer A L *et al* 2001 Intensity-modulated radiotherapy: current status and issues of interest *Int. J. Radiat. Oncol. Biol. Phys.* **51** 880–914
- Broyden C G 1970 The convergence of a class of double rank minimization algorithms *J. Inst. Math. Appl.* **6** 76–90
- Cotrutz C, Lahanas M, Kappas C and Baltas D 2001 A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy *Phys. Med. Biol.* **46** 2161–75
- Craft D, Halabi T and Bortfeld T 2005 Exploration of tradeoffs in intensity-modulated radiotherapy *Phys. Med. Biol.* **50** 5857–68
- Craft D, Halabi T, Shih H and Bortfeld T 2006 Approximating convex pareto surfaces in multiobjective radiotherapy planning *Med. Phys.* **33** 3399–407
- Eisbruch A 2002a Intensity-modulated radiotherapy of head-and-neck cancer: encouraging early results *Int. J. Radiat. Oncol. Biol. Phys.* **53** 1–3
- Eisbruch A 2002b Clinical aspects of IMRT for head-and-neck cancer *Med. Dosim.* **27** 99–104
- Eisbruch A *et al* 2001 Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer *Int. J. Radiat. Oncol. Biol. Phys.* **50** 695–704
- Eisbruch A *et al* 2004 Recurrences near base of skull after IMRT for head-and-neck cancer: implications for target delineation in high neck and for parotid gland sparing *Int. J. Radiat. Oncol. Biol. Phys.* **59** 28–42
- Gill P E, Murray W and Saunders M A 2002 SNOPT: An SQP algorithm for large-scale constrained optimization *SIAM J. Optim.* **12** 979–1006
- Griewank A *et al* 1996 ADOL-C: a package for the automatic differentiation of algorithms written in C/C++ *ACM Trans. Math. Softw.* **22** 131–67

- Jee K-W, McShan D and Fraass B 2005 Implementation of automatic differentiation tools for multicriteria IMRT optimization *Automatic Differentiation: Applications, Theory, and Tools (Lecture Notes in Computational Science and Engineering vol 50)* (Berlin: Springer) pp 225–234
- Kessler M L *et al* 2005 Costlets: a generalized approach to cost functions for automated optimization *Optim. Eng.* **4** 421–48
- Lahanas M, Baltas D and Zamboglou N 2003 A hybrid evolutionary algorithm for multi-objective anatomy-based dose optimization in high-dose-rate brachytherapy *Phys. Med. Biol.* **48** 399–415
- Lahanas M, Schreiber E and Baltas D 2003 Constrained free gradient-based optimization algorithms for multiobjective inverse planning in intensity modulated radiotherapy *Phys. Med. Biol.* **48** 2843–71
- Langer M, Lee E K, Deasy J O, Rardin R L and Deye J A 2003 Operations research applied to radiotherapy: an NCI-NSF-sponsored workshop February 7–9, 2002 *Int. J. Radiat. Oncol. Biol. Phys.* **57** 762–8
- Mackie T R, Scrimger J W and Battista J J 1985 A convolution method of calculating dose for 15-MV x-rays *Med. Phys.* **12** 188–96
- Marler R T and Arora J S 2004 Survey of multi-objective optimization methods for engineering *Struct. Multidisc. Optim.* **26** 369–95
- Marsh R *et al* 1997 Design and analysis of an immobilization and repositioning system for treatment of neck malignancies *Med. Dosim.* **22** 293–7
- McShan D L, Kessler M L, Vineberg K and Fraass B A 2006 Inverse plan optimization accounting for random geometric uncertainties with a multiple instance geometry approximation (MIGA) *Med. Phys.* **33** 1510–21
- Messac A and Mattson C 2002 Generating well-distributed sets of Pareto points for engineering design using physical programming *Optim. Eng.* **3** 431–50
- Miettinen K 1999 *Nonlinear Multiobjective Optimization* (Boston, MA: Kluwer)
- Mohan R, Wang X and Jackson A 1996 Optimization of 3-D conformal radiation treatment plans *Front. Radiat. Ther. Oncol.* **29** 86–103
- Niemierko A, Urie M and Goitein M 1992 Optimization of 3D radiation therapy with both physical and biological end-points and constraints *Int. J. Radiat. Oncol. Biol. Phys.* **23** 99–108
- Nocedal J and Wright S J 1999 *Numerical Optimization* (New York: Springer)
- Pareto V 1906 *Manual of Political Economy* (New York: Augustus M Kelley) (1971 translation of 1927 edition)
- Shanno D F 1970 Conditioning of quasi-Newton methods for function minimization *Math. Comput.* **24** 647–656
- Steuer R E 1985 *Multiple Criteria Optimization: Theory, Computation and Application* (New York: Wiley)
- Vineberg K A, Eisbruch A, Coselmon M M, McShan D L, Kessler M L and Fraass B A 2002 Is uniform target dose possible in IMRT plans in the head and neck? *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1159–72
- Wu Q, Mohan R and Niemierko A 2000 IMRT optimization based on the generalized equivalent uniform dose (EUD) *Proc. of the 13th Int. Conf. on the Use of Computers In Radiotherapy* ed T Bortfeld and W Schlegel (Heidelberg: University of Heidelberg) pp 17–9
- Wu Q, Djajaputra D, Wu Y, Zhou J, Liu H H and Mohan R 2003 Intensity-modulated radiotherapy optimization with gEUD-guided dose-volume objectives *Phys. Med. Biol.* **48** 279–91