

Reduced-order Parameter Optimization for Simplifying Prostate IMRT Planning

**Renzhi Lu¹, Richard J. Radke¹‡, Laura Happersett², Jie Yang²
Chen-Shou Chui², Ellen Yorke², and Andrew Jackson²**

¹Electrical, Computer, and Systems Engineering Department,
Rensselaer Polytechnic Institute, Troy, NY 12180

²Department of Medical Physics, Memorial Sloan-Kettering Cancer Center,
New York, NY 10021

E-mail: lur@rpi.edu, rjradke@ecse.rpi.edu, happersl@mskcc.org,
yangj@mskcc.org, chuic@mskcc.org, yorkee@mskcc.org,
jacksona@mskcc.org

PACS numbers: 87.53.Tf

Keywords: importance factors, IMRT optimization, prostate IMRT planning

Submitted to: *Phys. Med. Biol.*

‡ Please address correspondence to Richard Radke.

Abstract.

Intensity-modulated radiotherapy (IMRT) has become an effective tool for cancer treatment with radiation. However, even expert radiation planners still need to spend a substantial amount of time manually adjusting IMRT optimization parameters such as dose limits and costlet weights in order to obtain a clinically acceptable plan. In this paper, we describe two main advances that simplify the parameter adjustment process for five-field prostate IMRT planning. First, we report the results of a sensitivity analysis that quantifies the effect of each hand-tunable parameter of the IMRT cost function on each clinical objective and the overall quality of the resulting plan. Second, we show that a recursive random search over the 6 most sensitive parameters as an outer loop in IMRT planning can quickly and automatically determine parameters for the cost function that lead to a plan meeting the clinical requirements. Our experiments on a 10-patient dataset show that for 70% of the cases, we can automatically determine a plan in 10 minutes (on the average) that is either clinically acceptable or requires only minor adjustment by the planner. The outer loop optimization can be easily integrated into a traditional IMRT planning system.

1. Introduction

Intensity-modulated radiotherapy (IMRT) has been called “one of the most important technical advances in radiation therapy since the advent of the medical linear accelerators” (Ling et al., 2004). This computer-controlled method of delivering radiation can precisely irradiate a target with complex 3-D shape while simultaneously protecting normal radiation-sensitive tissues (Ling et al., 2004, Palta et al., 2004). The IMRT planning problem is usually solved by dividing each radiation beam into subcomponent pencil beams (or “beamlets”), and applying numerical optimization algorithms to determine the beamlet intensities such that the resultant radiation dose distribution best matches the requirements specified by the physician. The clinical goals of planning are encapsulated in an objective function that assigns a numerical score to each plan.

A basic difficulty is the formulation of this objective function. IMRT planners typically have to make compromises between competing clinical objectives, e.g., delivering as high and as uniform a dose as possible to the planning target volume (PTV), while sparing organs at risk (OARs) and normal tissues as much as possible. However, the compromise desired in any given case is not easy to specify in terms of the parameters (e.g., dose limits, weights, importance factors) defining the objective function. Furthermore, the appropriate parameter values differ from case to case. In current implementations of IMRT, prior knowledge of these parameters is not available, and planners can spend a substantial amount of time adjusting parameters in order to get a clinically acceptable plan. Despite improvements in the speed of numerical optimization procedures, the start-to-finish inverse planning process of obtaining a clinically acceptable IMRT plan for a difficult site can take several hours, largely due to the manual trial-and-error process of adjusting parameters in the objective function (Bedford and Webb, 2003, Bortfeld et al., 2004, Spirou and Chui, 1998). A systematic and partly automated methodology would reduce the time wasted in exploring unfruitful parameter sets and aid in reaching the desired planning goals efficiently.

In this paper, we describe two main advances that simplify the parameter adjustment process: a sensitivity analysis that identifies important parameter/constraint combinations, and an outer-loop optimization algorithm over the sensitive parameter set that can quickly and automatically determine parameters for the cost function that lead to a plan meeting the clinical requirements. We focus on the clinical planning procedure applied at Memorial Sloan-Kettering Cancer Center (MSKCC), and study prostate cancer treatment planning to a prescription dose of 86.4 Gy with 5-field IMRT (Ling et al., 2004). The paper is organized as follows. In Section 1.1, we review the types of tunable parameters commonly encountered in IMRT objective functions, and methods for their determination. In Section 2.1, we describe our mathematical formulation for prostate IMRT planning, and introduce our framework for outer-loop parameter optimization. In Section 2.2, we describe methods for sensitivity analysis that quantify the effect of each hand-tunable parameter of the IMRT cost function on each clinical objective. In Section 2.3, we describe a score function that ranks a plan based on how well it satisfies multiple clinical constraints. Section 2.4 describes the parameter optimization algorithm, which uses a recursive random search algorithm for fast global optimization. Sections 3.1 and 3.2 present sensitivity analysis and parameter optimization results on a 36-patient dataset using simplified dose calculations. A 10-patient subset is then evaluated clinically for acceptability following full dose calculation in Section 3.3. Section 4 concludes the paper with discussion and ideas for future work.

1.1. Tunable Parameters in IMRT Objective Functions

A typical scalar objective function for IMRT planning is composed of several dose- and dose-volume-based terms (i.e., costlets (Kessler et al., 2005)). Each term includes a dose parameter (i.e., a minimum or maximum limit) or a pair of dose-volume parameters (i.e., a point on a dose-volume-histogram (DVH) curve), and reflects a clinical objective. The costlets are usually combined into a scalar value using a weighted sum that reflects the relative penalty imposed by the planner for not satisfying each objective. While in theory, inverse treatment planning is a one-step, automatic process once the cost function is fixed, in practice, a great deal of trial and error is required to arrive at parameter settings such that the resulting minimizer is clinically acceptable (Purdy et al., 2001). The weights (or “importance factors”) are used by the planner to try to “steer” the optimization algorithm to more clinically desirable solutions (Webb, 2001), but this can be difficult since these weights are inherently imprecise and unintuitive (Bedford and Webb, 2003). The role of dose limits in IMRT optimization is also confusing, since it has been observed that in a penalty-based objective function, dose limits more stringent than are clinically necessary are frequently required to obtain convergence to an acceptable plan (see, e.g., (Samuelsson and Johansson, 2003, Spirou and Chui, 1998, Starkschall et al., 2001)).

The influence of different sets or distributions of importance factors on the qualitative character of the resulting plans has been investigated (e.g., (Lian and Xing, 2004, Meyer et al., 2004)), but not on a very large scale. Typically, only a few parameter combinations are tested to demonstrate that a difference exists. In contrast, the sensitivity analysis and reduced-order

models described below integrate results and experience from tens of previously planned cases and thousands of parameter combinations.

Several researchers have investigated methods for easing the planner's task by automatically determining good importance factors. Wu and Zhu (Wu and Zhu, 2001) used a genetic algorithm to alternately optimize over importance factors and beam weights. Bedford and Webb (Bedford and Webb, 2003) eliminated the importance factors by introducing a constraint-based objective function, and optimized beam weights using simulated annealing. Both approaches were applied to 3D Conformal Radiotherapy (CRT); however, for IMRT optimization, using such stochastic methods would be extremely slow because of the large number of beamlets to be determined. One solution is to optimize the beamlet intensities using a fast, deterministic method in an inner loop, while iteratively searching for "good" parameters in an outer loop. Xing et al. (Xing et al., 1999a,b) suggested a ranking function based on a plan's deviation from an ideal DVH for each organ, and showed how the ranking function could be minimized to determine the optimal importance factors. However, assuming simultaneous ideal DVHs for all targets/OARs may not be very realistic from a clinical perspective, and it is not clear whether this approach can scale up to full IMRT planning problems.

Fewer publications have investigated the effect of dose limits in penalty-based objective functions. Hunt et al. (Hunt et al., 2002) summarized the influence of optimization parameters on the resulting dose distributions, and observed that dose limits generally have a greater influence than importance factors. They specified a procedure for changes to be made in optimization parameters given specific deficits in clinical plans. However, this method only suggested the direction of change, and was heuristically implemented by hand. Barbieri et al. (Barbieri et al., 2002) searched for the best optimization parameters via structured grid searches over historical plans, using a recipe based on empirical knowledge. However, such grid searches might not be suitable for a degenerate parameter space, i.e., one in which different combinations of parameters lead to similar dose distributions.

Penalty-based scalar objective functions are not the only option for IMRT planning. A recent focus of interest is multiobjective (MO) optimization, first introduced in (Cotrutz et al., 2001), which results in a representative set of Pareto-efficient solutions for various conflicting optimization objectives. Craft et al. (Craft et al., 2005) showed how to find the Pareto-efficient space of solutions for MO optimization, and developed an interface for the planner to examine the way the tradeoffs change when the optimization parameters are altered. Other recent approaches were described by Romeijn et al. (Romeijn et al., 2004) and Yu (Yu, 1997). While neither the trial-and-error tweaking process described above nor prior knowledge about the parameters is required, a major drawback of the MO approach is speed, since a large number of full-dose-calculated candidate plans need to be generated. Furthermore, given a set of solutions, the problem of choosing the "best" one from a clinical and patient-specific perspective still remains (Amols and Ling, 2002). A decision-making process that incorporates clinical reasoning was introduced by Meyer et al. (Meyer et al., 2004), but it only works when the trade-offs between different plans are very large. Another natural approach is to cast IMRT as a constrained minimization problem (see, e.g., (Romeijn et al., 2003, Shepard

et al., 1999)), but this is still not the clinical standard, since efficiently solving such problems with DVH-based constraints is very difficult. We also note that Starkschall et al. (Starkschall et al., 2001) described an approach to treatment planning based on an applying an automatic search algorithm to simply find a feasible solution to all the constraints. However, in case no feasible solution is found, manual trial-and-error relaxation of constraints is still necessary, and the approach seems difficult to scale from CRT to IMRT.

In this study, we concentrate on penalty-based objective functions consisting of dose and dose-volume terms, since these are the norm in clinical IMRT planning (Langer et al., 2003). However, we note that the tools developed below would be mathematically general enough to apply to parameter estimation problems in dose-response-based objective functions, e.g., those based on TCP, NTCP, the equivalent uniform dose (EUD) (Wu et al., 2002), or the effect-volume histogram (EVH) (Alber and Nüsslin, 2002). Furthermore, the sensitivity analysis we propose may also be useful in a multiobjective context (e.g., by indicating which combinations of objectives on the Pareto front are worth exploring).

2. Materials and Methods

2.1. IMRT Planning for the Prostate

To meet the clinical goals, we use the quadratic dose-based objective function implemented at MSKCC (Spirou and Chui, 1998, Ling et al., 2004), and determine the beamlet intensities by numerical optimization. For the k th target, the corresponding objective function term is:

$$F_{target_k} = \frac{1}{N_k} \left(\sum_{i=1}^{N_k} (D_i - D_{pres_k})^2 + w_{min_k} \sum_{i=1}^{N_k} (D_i - D_{min_k})^2 \cdot \Theta(D_{min_k} - D_i) + w_{max_k} \sum_{i=1}^{N_k} (D_i - D_{max_k})^2 \cdot \Theta(D_i - D_{max_k}) \right), \quad (1)$$

where N_k is the number of points in the target, and D_i is the dose to the i th point in the target. D_{pres_k} is the prescription dose, and D_{min_k} and D_{max_k} are the minimum and maximum dose allowed without penalty. w_{min_k} and w_{max_k} are the penalties (weights) for under- and over-dosing, and $\Theta(x)$ is the Heaviside function. The choice of the parameter set $P_k = \{D_{pres_k}, D_{min_k}, D_{max_k}, w_{min_k}, w_{max_k}\}$ completely specifies the objective function for target k . A similar objective function term is defined for each organ at risk (OAR), which also includes parameters D_{dv_k} and w_{dv_k} that define the dose-volume-histogram (DVH) constraints:

$$F_{OAR_k} = \frac{1}{N_k} \left(w_{max_k} \sum_{i=1}^{N_k} (D_i - D_{max_k})^2 \cdot \Theta(D_i - D_{max_k}) + w_{dv_k} \sum_{i=1}^{N_{dv_k}} (D_i - D_{dv_k})^2 \cdot \Theta(D_i - D_{dv_k}) \right), \quad (2)$$

where the sum in the second term is carried out over the lowest N_{dv_k} doses that are greater

Targets	D_{pres}	D_{min}	w_{min}	D_{max}	w_{max}
PTVE	100%	98%	50	102%	50
Overlap	84%	83%	10	85%	20
Mid	98%	97%	10	99%	10

Table 1: Default parameters for the targets.

OARs	D_{max}	w_{max}	D_{dv}	w_{dv}	Volume
RectOE	85%	20	30%	20	30%
BladO	96%	35	30%	20	30%

Table 2: Default parameters for the OARs.

than D_{dv_k} , and N_{dv_k} is the minimum number of point dose changes required to bring the k th organ into compliance with the DVH constraint (Spirou and Chui, 1998).

At MSKCC, a typical prostate plan objective function has 3 target structures (the expanded PTV, or PTVE; the overlap region between the PTV and rectum, or Overlap; and a cylindrical structure near the center of the PTV for the protection of the urethra, or Mid), as well as 2 OAR structures (the expanded outer rectal wall, or RectOE; and the outer bladder wall, or BladO).§ Hence we can formulate the overall intensity optimization problem as:

$$I^*(P) = \arg \min_I \left(\sum_{j=1}^{N_{target}} F_{target_j}(D(I), P) + \sum_{k=1}^{N_{OAR}} F_{OAR_k}(D(I), P) \right), \quad (3)$$

where $P = \{P_j, j = 1, \dots, N_{target}, P_k, k = 1, \dots, N_{OAR}\}$ is the set of about 20 dose limits and weights for all the optimization structures that defines the objective function.|| The current MSKCC clinical evaluation protocol requires that the plan for IMRT treatment of the prostate to 86.4 Gy satisfies the following conditions: (1) for the PTV, $V_{95} \geq 87\%$, and $D_{max} \leq 111\%$, (2) for the rectal wall, $V_{87} \leq 30\%$, $V_{54} \leq 53\%$, and $D_{max} \leq 99\%$, and (3) for the bladder wall, $V_{54} \leq 53\%$. All doses are expressed as percentage of the prescription dose, and the notation V_x means the structure volume percentage receiving at least x% of the prescription dose. The parameters P must be carefully selected, usually by an expert planner’s judgment and/or manual iterations, to obtain a clinically acceptable dose distribution that satisfies the constraints of the protocol, as illustrated in Figure 1a. As described above, determining viable parameter values is not an immediate, one-step process for several reasons, the main one being that the sum-of-penalties form of the objective function is not guaranteed to force the resulting solution to fully satisfy the clinical constraints. Tables 1 and 2 list the default parameter settings specified in the current protocol, which serve as the starting points for manual adjustment.

Our goal in this paper is to automatically find a set of parameters such that the minimizer of the corresponding IMRT objective function satisfies the clinical constraints. One option would be to perform an exhaustive search over the approximately 20-dimensional parameter space as in (Xing et al., 1999a,b), which would be computationally inefficient. Instead, we

§ The expanded structures are slightly larger than the corresponding anatomical structures. These “dummy” structures are used by MSKCC planners to guide their particular optimization algorithm to a more clinically desirable solution. Evaluation of the resulting plan is done on the basis of dose to the true anatomical structures (e.g. the PTV and the rectal wall).

|| Equation (3) reflects MSKCC’s clinical practice that the prescription dose weights for all targets are equal to 1. In theory, these weights could be independently adjusted, but in practice, the planners find that adjusting the weights on the minimum and maximum dose constraints allows them to achieve satisfactory clinical outcomes.

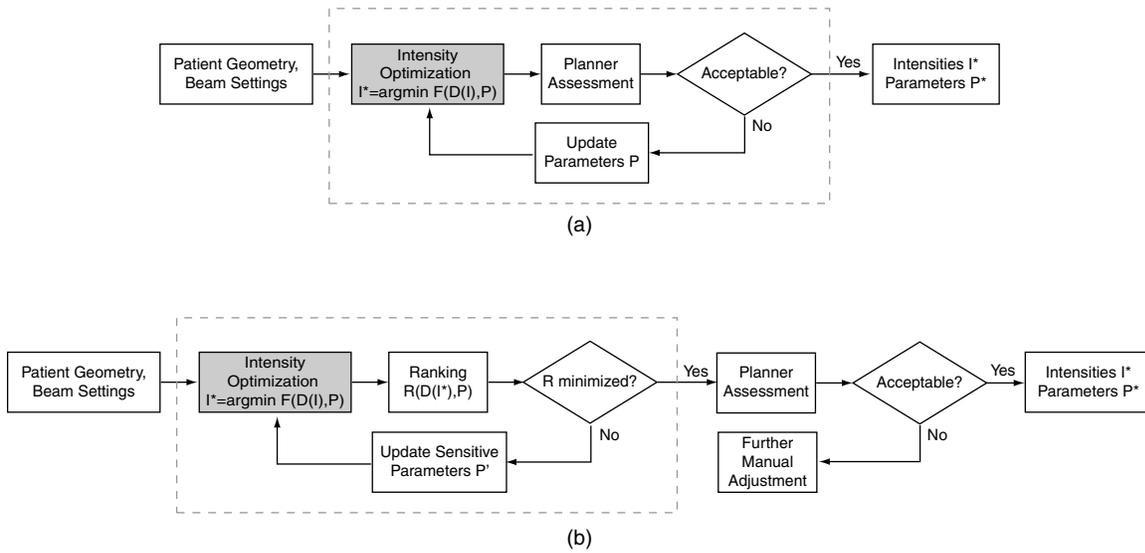


Figure 1: Clinical planning and our proposed outer-loop algorithm. (a) In the clinic, a planner manually adjusts a set of patient-specific parameters P defining a cost function, and optimizes the cost function using an automatic algorithm (shaded box). If all constraints of the protocol are satisfied and the plan is judged to be acceptable, the resulting beam intensities are used to treat the patient. (b) We propose to replace the dotted box with an algorithm that automatically determines a parameter set that will result in a clinically acceptable plan. A patient-independent ranking function R reflecting the constraints of the protocol is minimized in an outer loop over a small subset of the parameters, determined based on sensitivity analysis.

performed a sensitivity analysis to evaluate the way weights and dose limits affect the clinical constraints, and more importantly, to identify the critical parameters that really have an impact on inverse planning.

We obtained clinical five-field, 86.4 Gy IMRT plans for 36 prostate cancer patients from MSKCC, all created by experienced planners; the access to this data was approved by the MSKCC Privacy Review Board. These hand-tuned clinical plans were referred to as ground truth. Our automated parameter optimization used the same CT data and beam directions, but the tunable parameters P , and hence the intensities, were chosen according to the following procedure. We first performed the sensitivity analysis (discussed in the next section) on the first 5 patients, and the resulting sensitivity measures were averaged over these patients to determine the sensitive set. Since patients were numbered randomly, these five patients had no obvious connections with each other. Then we performed a two-loop optimization over all 36 patients, as illustrated in Figure 1b. The inner loop optimized the intensity distribution given a parameter set P , and the outer loop optimized the sensitive parameter set P' with respect to a patient-independent ranking function R that reflects how well a plan satisfies the clinical evaluation criteria. An approximate dose calculation algorithm was applied in the outer loop to increase speed while maintaining accuracy (see Section 2.3). The resultant dose statistics were compared with those from the clinical ground truth. Finally, 10 randomly

selected plans were evaluated by the same physicist with extensive experience with prostate planning at MSKCC.

The automatically generated plans (i.e., parameter sets and field intensities) might not coincide with the clinical plans, due to the mathematical degeneracies discussed in the next section. For the purpose of the experiment, we assume that the quality of the plan is defined entirely by the degree to which the clinical constraints described above are satisfied.

The protocol at MSKCC requires that, prior to evaluation of a treatment plan, the dose distribution is normalized so that the maximum dose to the PTV is equal to 110%. Since our goal is to be able to clinically evaluate the outer-loop plans and directly compare them to plans for the same patients produced in the clinic, all of our sensitivity analysis refers to the response variables of plans after this normalization.

2.2. Parameter/Constraint Sensitivity Analysis

The goal of sensitivity analysis is to automatically divide the parameter set defining the IMRT objective function into 3 subsets:

- (i) The fixed set, in which the parameters are held approximately constant by the planners in practice.
- (ii) The sensitive set. Varying any element of this set of parameters will greatly affect the plan evaluation.
- (iii) The insensitive set. The statistics of plan evaluation are not significantly influenced by this set of parameters. This leads to the degeneracy problem, i.e., that different sets of parameters can lead to similar dose distributions.

The first set is easy to identify; for prostate planning at MSKCC, it includes: (1) the prescription dose for PTVE (100% by definition); (2) the maximum and minimum dose for PTVE, Overlap and Mid, which are always of the form $D_{pres} \pm 1\%$ or $D_{pres} \pm 2\%$; and (3) the volume of RectOE constrained to a particular dose, which is always 30%. The planners have found empirically that they can arrive at acceptable solutions while keeping these parameters fixed. However, it is less obvious which of the remaining parameters are in the sensitive set and which are not.

We illustrate the idea of sensitivity analysis for prostate IMRT with a motivating example. We vary two parameters, the weights for PTVE Dmax and Dmin, and hold the other parameters of the objective function constant. The beamlet intensities are then optimized using each set of parameters and the clinical dose statistic V_{95}^{PTV} is evaluated. Subsequently, the RectOE Dmax, BladO Dmax and RectoE DVH doses were incremented by 10% and the process repeated. As shown in Figure 2, V_{95}^{PTV} coverage is very sensitive to PTVE Dmin weight. It is also influenced by the combination of RectOE Dmax, BladO Dmax and RectOE DVH doses. As we loosen the OAR requirements by increasing the dose limits in this combination, V_{95}^{PTV} improves significantly. However, V_{95}^{PTV} is relatively insensitive to PTVE Dmax weight.

To quantify this intuition for all parameters and planning goals, we carried out a systematic sensitivity analysis for the prostate. We note that in the context of constrained

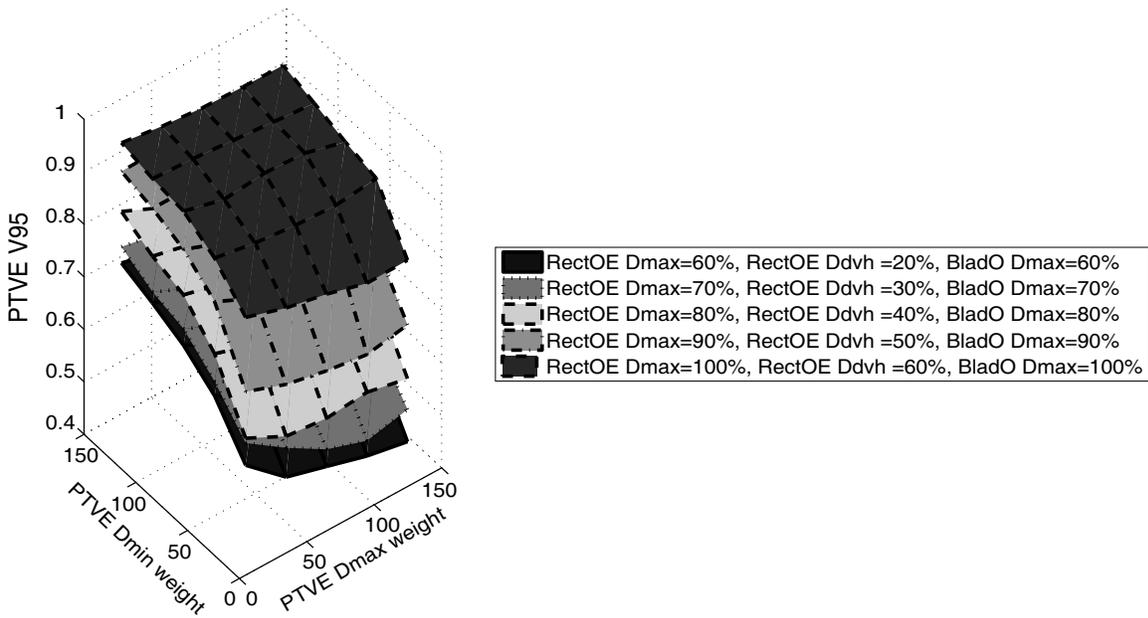


Figure 2: Example of sensitive/insensitive parameters. V_{95}^{PTV} is displayed as a function of the PTVE Dmax weight and Dmin weight. The relationship is also controlled by RectOE max dose, RectOE DVH dose and BladO max dose.

optimization, Alber et al. (Alber et al., 2002) reported the sensitivity of the objective function with respect to changes in dose limits using a Lagrange multiplier approach; however, such Lagrange information is not available in our unconstrained problem (3). We applied Monte Carlo simulation to measure the sensitivity of all clinical output variables (i.e., the dose and dose-volume constraints that determine the clinical acceptability of a plan), with respect to the tunable input parameters (i.e., the dose and dose-volume limits and penalties used in optimization). First, we determined the natural range of input parameters based on the values observed over all the training patients. Then, for each training patient, we generated 2000 full sets of input parameters from this range by Latin hypercube sampling (a particular case of stratified sampling that achieves a better coverage of the space of input parameters (Saltelli et al., 2004)). For each set of input parameters, we then minimized the corresponding IMRT cost function using the conjugate gradient method used in the clinic (Spirou and Chui, 1998), and obtained the values of the clinical output variables. Finally, we quantified the sensitivity of each input/output combination using three statistical measures. Using these measures, we can test whether the following hypothesis is true: “no correlation exists between IMRT outcome Y_j and IMRT parameter setting X_i at the 95% confidence level”. In the subsequent parameter adjustment (either by the planner’s manual adjustment or by an outer-loop optimization as we describe below), only a small subset of the parameters, i.e., the sensitive set, needs to be explored, which will decrease the dimensionality of the parameter space significantly. The other parameters are simply fixed to their defaults.

Suppose we generate M samples, each of which consists of a scalar clinical outcome Y , and N independent IMRT parameters X_1, \dots, X_N . From the M samples

$\{(Y_i, X_{i1}, \dots, X_{iN}), i = 1, \dots, M\}$, we compute the following statistics between Y and $X_j, j = 1, \dots, N$ (Kutner et al., 2004):

- The linear correlation coefficient, also known as the Pearson product moment correlation coefficient (PMCC):

$$PMCC(X_j, Y) = \frac{\frac{1}{M-1} \sum_{i=1}^M (X_{ij} - \bar{X}_j)(Y_i - \bar{Y})}{Std(X_j)Std(Y)}, \quad (4)$$

where $\bar{\cdot}$ and $Std(\cdot)$ represent the mean value and the standard deviation of a variable.

- The partial correlation coefficient (PCC):

$$PCC(X_j, Y) = PMCC(X_j - \hat{X}_j, Y - \hat{Y}), \quad (5)$$

where $\hat{X}_j = b_0 + \sum_{k \neq j} b_k X_k$ and $\hat{Y} = c_0 + \sum_{k \neq j} c_k X_k$ represent the linear regressions from $X_k, k \neq j$ onto X_j and Y . Thus, the PCC provides a measure of the strength of the linear relationship between two variables after a correction has been made for the linear effects of the other variables in the analysis.

- The standardized rank regression coefficients (SRRC):

$$SRRC(X_j, Y) = \frac{b_j Std(Rank(X_j))}{Std(Rank(Y))}, \quad (6)$$

where $Rank(\cdot)$ is the transformation that replaces the data points with their corresponding ranks, i.e., rank 1 is assigned to the smallest observation and rank M to the largest observation. b_j is the linear regression coefficient from $Rank(X_j)$ onto $Rank(Y)$. The problem associated with poor linear fits to nonlinear data can often be avoided with the use of rank transformations.

2.3. Scoring Candidate Plans

We next require a score function $R(P)$ that quantifies the degree to which the dose distribution provided by minimizing the optimization function with parameters P meets the clinical constraints. As described above, the treatment protocol is usually expressed in terms of several points on the DVH curves. Thus, we use a score function based on these key points, rather than an entire ideal DVH curve for each organ as in (Xing et al., 1999a).

Under the clinical protocol for 86.4 Gy prostate IMRT at MSKCC, we try to maximize the target dose coverage V_{95}^{PTV} and cold-spot D_{min}^{PTV} , while minimizing the rectal wall hot-spot $D_{max}^{rectwall}$, as well as certain points on the OAR DVH's: $V_{87}^{rectwall}$, $V_{54}^{rectwall}$, and $V_{54}^{bladwall}$. In addition, the planners usually try to push the middle constraint of the rectum (i.e., $V_{54}^{rectwall}$) to the protocol limit. Hence, we used a score function of the form:

$$\begin{aligned} R(D) = & \alpha_1 \cdot g(87 - V_{95}^{PTV}(\%)) + \alpha_2 \cdot g(75 - D_{min}^{PTV}(\%)) + \alpha_3 \cdot g(V_{87}^{rectwall}(\%) - 30) \\ & + \alpha_4 \cdot g(D_{max}^{rectwall}(\%) - 99) + \alpha_5 \cdot g(V_{54}^{bladwall}(\%) - 53) \\ & + \alpha_6 \cdot g(V_{54}^{rectwall}(\%) - 53) + \alpha_7 \cdot g(48 - V_{54}^{rectwall}(\%)), \end{aligned} \quad (7)$$

where $g(x) = \begin{cases} 1000 \cdot x & x \geq 0 \\ x & x < 0 \end{cases}$, which severely punishes the violation of clinical constraints. The parameters α_6 and α_7 work together to penalize deviation of $V_{54}^{rectwall}$ outside the interval [48%, 53%], to achieve the clinical goal of pushing $V_{54}^{rectwall}$ to the protocol limit.

The parameters α_i should reflect the emphasis placed by planners on individual clinical constraints, and differ from the weight terms in (3). For example, maximizing PTV coverage is usually considered more important than achieving stricter OAR protection than required by the clinical evaluation guidelines described in Section 2.1. Also, order statistics such as maximum doses are too noisy to assign a very large weight. The allocation that we used in our experiments in Section 3 below was $\alpha_1 = 1000, \alpha_2 = 10, \alpha_i = 1, i = 3, 4$, and $\alpha_i = 100, i = 5, 6, 7$. Any deviation from the planner’s expectation of a “good plan” results in a high positive score. Conversely, a negative score usually means that all the constraints are satisfied.

One natural question about the proposed approach has to do with the choice of ranking function in (7). Aren’t the coefficients α_i just different kinds of importance factors that still need to be tuned by a human? There are two key differences between the α_i in (7) and the dose limits and weights in (3). First, we show in Section 3.2 below that the same set of α_i was reasonably successful for the automatic outer-loop planning of all of the patients in our study (see Figure 6), whereas “good” weights and dose limits in (3) vary widely from patient to patient. This implies that the work of choosing a good ranking function would only need to be done once for a given clinical protocol and anatomic class, not once for every patient. Second, the outer-loop plan quality is relatively insensitive to the exact values of the α_i —only order-of-magnitude differences have an effect. The main function of the α_i is to reflect the clinical evaluation protocol and the expert planner’s opinion about which constraints are most important to satisfy. The goal is to produce a scalar function that roughly reflects the human evaluation process.

In our initial experiments, we set all α_i equal to 1. The optimization of (7) resembled a feasibility search, since the mere satisfaction of each constraint was emphasized over further improvement via $g(x)$. Most of the plans terminated with a negative ranking score, but the constraints (especially for the PTV) were barely satisfied. According to clinical evaluation feedback from MSKCC, planners preferred a further improvement in PTV coverage while saturating the rectum middle constraint. Hence, in our second attempt we increased α_1 and α_2 to reward improvement in the PTV constraints. Instead of changing the α_i on a fine scale, we only selected an order of magnitude for each α_i that seemed sufficient to reflect the clinical expectations. A sample of how the optimization result changes with different settings of α_i is listed in Table 3. Both Settings 2 and 3 emphasize PTV coverage, but with different magnitudes. Setting 4 puts more emphasis on OAR protection. We can see that the number of cases that met all the constraints (based on corrected-dose calculations, see below) is basically insensitive to the 4 different settings, although the α_i values are quite different. For Settings 2 and 3, moderate changes of α_i result in quite similar dose distributions. Comparing Settings 2 and 4, the differences in the orders of magnitude of the α_i related to the PTV and OARs determine which response variables are pushed beyond the protocol requirements. While we

Setting	α_1	α_2	α_3	α_4	α_5	α_6	α_7
1	1	1	1	1	1	1	0
2	1000	10	1	1	100	100	100
3	1500	100	1	1	50	50	50
4	1	1	100	100	100	100	0

Setting	# pass	\overline{V}_{95}^{PTV}	\overline{D}_{min}^{PTV}	\overline{V}_{54}^{blad}	\overline{V}_{87}^{rect}	\overline{V}_{54}^{rect}	$\overline{D}_{max}^{rect}$
1	33	87.8 (2.1)	74.1 (2.7)	33.2 (8.4)	10.7 (3.5)	43.8 (5.2)	96.1 (2.1)
2	34	92.7 (2.8)	78.1 (4.1)	37.0 (9.1)	11.3 (4.3)	51.1 (2.9)	96.8 (1.7)
3	35	92.9 (2.5)	78.7 (4.3)	38.4 (7.9)	12.2 (4.1)	51.2 (2.6)	97.0 (1.4)
4	33	87.2 (2.0)	72.6 (3.3)	31.3 (6.3)	8.2 (2.8)	40.5 (4.8)	95.3 (1.9)

Table 3: A summary of the outer-loop optimization results for all 36 patients, with respect to different settings of $\{\alpha_i\}$. “# pass” refers to the number of cases that satisfied all the clinical constraints, based on corrected dose calculations. All dose statistics (given in %) are the mean values over all 36 patients, with standard deviations in parentheses.

applied Setting 2 in the experiments reported in Section 3, Setting 4 might be a good choice for a more conservative protocol.

Both intensity optimization in the inner loop and plan ranking in the outer loop were performed with approximate dose calculations to reduce computation time. For optimization, MSKCC utilizes a truncated pencil beam method which accounts only partly for scattered dose (Spirou and Chui, 1998, Ling et al., 2004), although the calculation for clinical evaluation includes full scatter. We apply the same truncated method in the inner loop for speed. However, the full dose calculation is too time-consuming to be applied in each outer loop iteration. Siebers et al. (Siebers et al., 2002) proposed a hybrid method that leveraged the speed of a fast pencil-beam algorithm yet achieved the accuracy of optimization from superposition/convolution algorithms via the application of dose correction matrices. We adopt the same strategy, but keep the correction matrix constant during each outer loop iteration. That is, at the first outer-loop iteration, we compare the truncated and full dose calculation for each voxel in the treatment volume, and compute the local correction ratio $\frac{D_{full}}{D_{truncated}}$ for each voxel. In subsequent iterations, we only perform truncated dose calculations but scale the result by the correction ratio to improve the consistency with the true full dose (see the Appendix for more details). In summary, there are 3 different dose calculation methods being used: (1) the truncated dose kernel calculations (T-D) used by the optimization’s inner loop, (2) the “corrected” dose calculations (C-D) used in the outer loop, and (3) the full dose calculations (F-D) used for clinical planning and evaluation (and also to estimate the correction ratio).

2.4. Parameter Optimization Algorithm

The score function (7) is not differentiable with respect to the parameter vector, since the mapping between the parameters and the optimized intensities is not differentiable in (1) and

(2). Hence, we cannot use derivative-based local search methods such as quasi-Newton or steepest descent. Instead, to minimize over the sensitive set P' , we adopt the recursive random search (RRS) algorithm proposed in (Ye and Kalyanaraman, 2003), which is based on random sampling. The efficiency of random sampling is high at initial samples, but falls sharply at later samples. This observation led to the idea of restarting the sampling before its efficiency drops off, either by moving or resizing the sample space according to sample history. The RRS algorithm searches the parameter space in two recursive steps: exploration and exploitation. The exploration step examines the macroscopic features of the objective function (e.g., convex or valley-like structure) and attempts to identify promising areas in the parameter space P' that will be investigated in the exploitation step. RRS performs exploitation in two iterative steps: (i) random sampling within the current space, and (ii) realignment or shrinkage. In this way, most trivial subspaces will be excluded from exploitation, improving the overall efficiency of the search.

The time required for planning depends on the number of iterations in the outer-loop parameter selection algorithm. We define the following stopping criteria for the outer loop: (1) if the current ranking function is negative, and no more improvement is observed after the next 20 iterations; or (2) if the current ranking function is still positive, but no more improvement is observed after the next 100 iterations.

For comparison, we also performed global minimization using the controlled random search (CRS) algorithm (Price, 1983). CRS randomly generates a population of n points, and performs a downhill simplex move (e.g., reflections, expansions, and contractions) within a random subset of the n points. The population of n points is updated by removing the worst elements as better points are obtained during the downhill simplex moves.

3. Results

3.1. Sensitivity Analysis

For each of the five training patients, we randomly generated 2000 parameter combinations as described above, minimized the corresponding IMRT cost function, and obtained the values of six response variables: V_{95}^{PTV} , D_{min}^{PTV} , $V_{87}^{rectwall}$, $V_{54}^{rectwall}$, $V_{54}^{bladwall}$, and $D_{max}^{rectwall}$. For each objective function parameter/response variable combination, we computed the measures of sensitivity given by the Pearson product moment correlation coefficient (PMCC), the Partial Correlation Coefficient (PCC), and the Standardized Rank Regression Coefficient (SRRC), described in Section 2.2.

The sensitivity measures differ from patient to patient, since the optimized dose distributions and the compromises in the optimization depend strongly on the shape and position of the target and normal tissue (Hunt et al., 2002). To make the result applicable across a general population, we averaged the correlation measures across the first five patients to determine the parameter sensitivities. Table 4 summarizes the measures for V_{95}^{PTV} with respect to all of the parameters. The values of PMCC and SRRC are similar, indicating that linear vs. nonlinear regression models make little difference for this problem. PCC measures

Parameter Name	PMCC	PCC	SRRC
RectOE DVH Dose	0.45	0.66	0.47
RectOE Dmax Dose	0.44	0.65	0.50
PTVE Dmax Weight	0.39	0.60	0.37
BladO Dmax Dose	0.24	0.25	0.21
BladO DVH Dose	0.10	0.15	0.11
Mid Dpres Dose	0.08	0.13	0.06
Mid Dmin Weight	0.02	0.05	0.02
PTVE Dmin Weight	-0.02	-0.02	-0.03
Mid Dmax Weight	-0.02	-0.00	-0.02
BladO DVH Weight	-0.02	-0.04	-0.04
BladO Dmax Weight	-0.06	-0.17	-0.09
RectOE Dmax Weight	-0.08	-0.10	-0.05
Overlap Dmin Weight	-0.10	-0.18	-0.11
Overlap Dmax Weight	-0.12	-0.28	-0.16
RectOE DVH Weight	-0.27	-0.47	-0.28

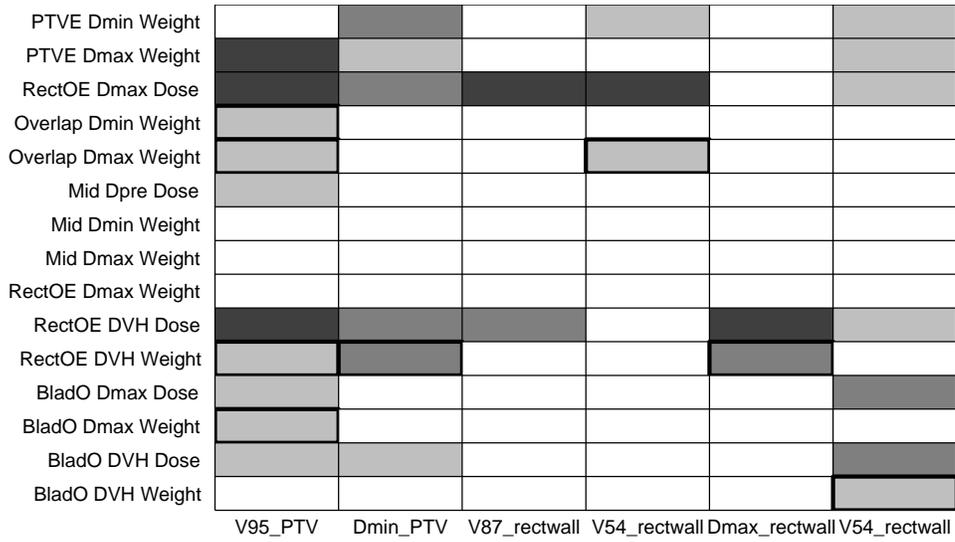
Table 4: PMCC, PCC and SRRC of V_{95}^{PTV} coverage with respect to all the parameters. The parameters are ordered from highest to lowest PMCC, which generally agrees with the ordering with respect to PCC and SRCC.

are larger in magnitude than PMCC or SRRC, which leads to the conclusion that the cost function parameters do not affect the IMRT evaluation objectives independently. The overall ranking of the parameters is generally consistent for all three measures.

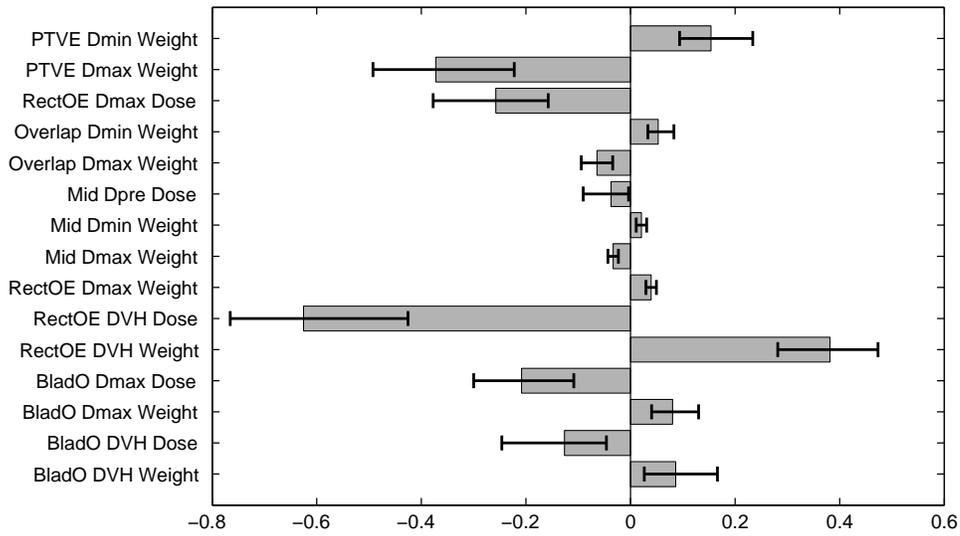
Figure 3a represents the PCC statistics for all the parameter/constraint combinations using a swatch chart, where a darker color signifies a higher correlation. In this view, the sensitive and insensitive sets can easily be visually separated. In Figure 3b, we illustrate the average PCC values for the overall outer-loop score function (7) with $\alpha_1 = 1000$, $\alpha_2 = 10$, $\alpha_i = 1$, $i = 3, 4$, and $\alpha_i = 100$, $i = 5, 6, 7$. From the averaged measures over the 5 patients, we can identify the 6 most influential factors: PTVE Dmin weight, PTVE Dmax weight, RectOE Dmax dose, RectOE DVH dose, RectOE DVH weight, and BladO Dmax dose. Although the relative ranks differed slightly for the 5 training patients (e.g., due to the variations in patients' geometry), the set containing the 6 most sensitive variables remained the same. Thus, the dimensionality of the sensitive parameter space can be reduced to 6. Based on the clinical plans in our experiment, we limit the range of variation in the subsequent parameter search as indicated in Table 5.

The influence of the different α_i settings discussed in Table 3 on the sensitivity analysis is shown in Table 6. Since the sensitivity of the ranking function is related to that of each constraint response variable, as well as the way they are combined, we expect changes in the values of the sensitivity measures. However, the sensitive sets only differ slightly among the 4 settings. In Settings 1 and 4 where PTV coverage is deemphasized, the two additional parameters BladO DVH dose and BladO DVH weight enter the sensitive set. Thus, an 8-dimensional parameter space would accommodate all 4 settings.

To determine the sensitive set above, we randomly chose 5 out of the 36 patients. The



(a)



(b)

Figure 3: (a) A swatch chart representing PCC values for all the input parameters, with respect to each output dose evaluation statistic. The gray level indicates the strength of the correlation as follows: empty indicates an absolute value of less than 0.1; a light tone indicates an absolute value between 0.1 and 0.3; a moderate tone indicates an absolute value between 0.3 and 0.5; and a dark tone indicates an absolute value greater than 0.5. Boxes with normal outlines indicate positive correlation; boxes with thicker outlines indicate negative correlation. (b) Average PCC values for the overall score function (7) with $\alpha_1 = 1000, \alpha_2 = 10, \alpha_i = 1, i = 3, 4,$ and $\alpha_i = 100, i = 5, 6, 7$. The range bars indicate the minimum and maximum values of the PCC over the 5 training patients. From these visualizations, the most sensitive parameters can easily be visually identified.

Parameter	Lower	Upper	Default	Clinical role
PTVE Dmax Weight	20	250	50	limit PTV max dose
PTVE Dmin Weight	20	250	50	ensure PTV V95
RectOE Dmax Dose	60%	100%	85%	limit rectwall Dmax,V87
RectOE DVH Dose	20%	60%	30%	limit rectwall V54
RectOE DVH Weight	10	100	20	limit rectwall V54 ,V87
BladO Dmax Dose	60%	100%	96%	limit bladwall Dmax

Table 5: The search range, default values, and clinical roles of the sensitive set of parameters.

Setting	Sensitive parameters
1	PTVE Max/Min weight; RectOE Max dose, DVH dose/weight; BladO DVH dose/weight
2	PTVE Max/Min weight; RectOE Max dose, DVH dose/weight; BladO Max dose
3	PTVE Max/Min weight; RectOE Max dose, DVH dose/weight; BladO Max dose
4	PTVE Max/Min weight; RectOE Max dose, DVH dose/weight; BladO Max dose, DVH dose/weight

Table 6: A summary of the sensitivity analysis result for the 5 training patients, with respect to different settings of $\{\alpha_i\}$. Settings 1 to 4 correspond to the values in Table 3.

purpose was to show that the sensitivities from a small subset of patients can be applied to a broader population without requiring patient-specific analysis. Prior information about the remaining patients did not play a role, since we feel it is important to preserve the separation of training data from testing data. However, Figure 4 illustrates an after-the-fact sensitivity analysis for all 36 patients, indicating the number of patients for which a particular parameter/constraint combination had an absolute PCC value of greater than 0.3. Compared to Figure 3, the patient-specific sensitivities are generally in accordance with what we learned from the 5 training patients. The only major difference is that a subset of patients (11 out of 36) are sensitive to BladO DVH dose and BladO DVH weight, which were not in the original sensitive set.

3.2. Parameter Optimization

We first tested the convergence performance of our parameter optimization method. In the inner loop for intensity optimization, we applied the conjugate gradient method from (Spirou and Chui, 1998), the same method used in the MSKCC planning system. The quadratic objective function (3) usually converged in 20 to 30 iterations. In the outer loop, we searched for the optimum parameters in the reduced parameter set using the RRS method, and as a comparison, using controlled random search (CRS). The parameter search in all cases started from the default values supplied by the clinical protocol (see Tables 1 and 2). Two examples of iterative optimization are shown in Figure 5 (top). For Patient 3, we found a set of parameters which satisfies all the constraints (based on the corrected dose calculations, or C-D) within 28 iterations using RRS, compared to 95 iterations using CRS. The optimized score function value $R(D)$ was -383.06, indicating that all the clinical constraints are satisfied. For

PTVE Dmin Weight	0	11	0	0	0	1
PTVE Dmax Weight	35	1	0	0	0	2
RectOE Dmax Dose	34	30	35	34	0	15
Overlap Dmin Weight	0	0	0	0	0	0
Overlap Dmax Weight	12	0	0	0	0	0
Mid Dpre Dose	0	0	0	0	0	0
Mid Dmin Weight	0	0	0	0	0	0
Mid Dmax Weight	0	0	0	0	0	0
RectOE Dmax Weight	0	0	0	0	0	0
RectOE DVH Dose	35	8	13	1	35	5
RectOE DVH Weight	18	12	0	0	23	0
BladO Dmax Dose	15	0	0	0	0	23
BladO Dmax Weight	0	0	0	0	0	1
BladO DVH Dose	2	8	0	0	0	11
BladO DVH Weight	0	2	0	0	0	11
	V95_PTV	Dmin_PTV	V87_rectwall	V54_rectwall	Dmax_rectwall	V54_rectwall

Figure 4: A sensitivity analysis for all 36 patients, indicating the number of patients for which a particular parameter/constraint combination had an absolute PCC value of greater than 0.3.

comparison, the ranking function value for the clinical plan actually used to treat Patient 3 was -2610, indicating some room for improvement. For Patient 8, RRS converged to -172.34 after 24 iterations, compared to -161.5 using CRS after 73 iterations. For comparison, the ranking function value for the clinical plan actually used to treat Patient 8 was -241, which is comparable to the RRS plan. For both RRS and CRS, we fixed the value of the objective function until the next move is downhill. RRS is able to escape from the unproductive sampling space more quickly than CRS.

We also examined the sensitivity of the optimization algorithm with respect to the dimensionality of the parameter space. As shown in Figure 5 (top), for Patient 3, the RRS in a full 13-dimensional space reached 1574 after 100 iterations (i.e., some constraints were left unsatisfied), and converged more slowly than the reduced-order search. Patient 8 has a similar result, indicating that the dimensionality reduction is highly beneficial for the random sampling technique.

To illustrate the improvement of plans gained by outer-loop parameter optimization, Figure 5 (bottom) also compares the dose-volume histograms (DVHs) from the RRS optimization with the result corresponding to the default parameter settings for Patients 3 and 8. While the rectum and bladder received lower doses using the default parameter settings, the clinical constraint for V_{95}^{PTV} was not satisfied. The reduced order parameter optimization achieved better coverage in the prostate, and the OAR protection was still acceptable. Figure 6 summarizes the results for all 36 patients based on 6 evaluation points on DVH curves: V_{95}^{PTV} , D_{min}^{PTV} , $V_{87}^{rectwall}$, $V_{54}^{rectwall}$, $V_{54}^{bladwall}$, and $D_{max}^{rectwall}$. Compared to the plans obtained using default settings, the parameter optimization significantly improved D_{min}^{PTV} and V_{95}^{PTV} . The mean value of D_{min}^{PTV} increased from 67.5% to 78.1%, and the mean value of V_{95}^{PTV} increased from 82.2% to 92.7%. RRS generally pushed $V_{54}^{rectwall}$ to the clinical limit of 53% as we expected, while still leaving enough margins for $V_{54}^{bladwall}$ and $V_{87}^{rectwall}$. All

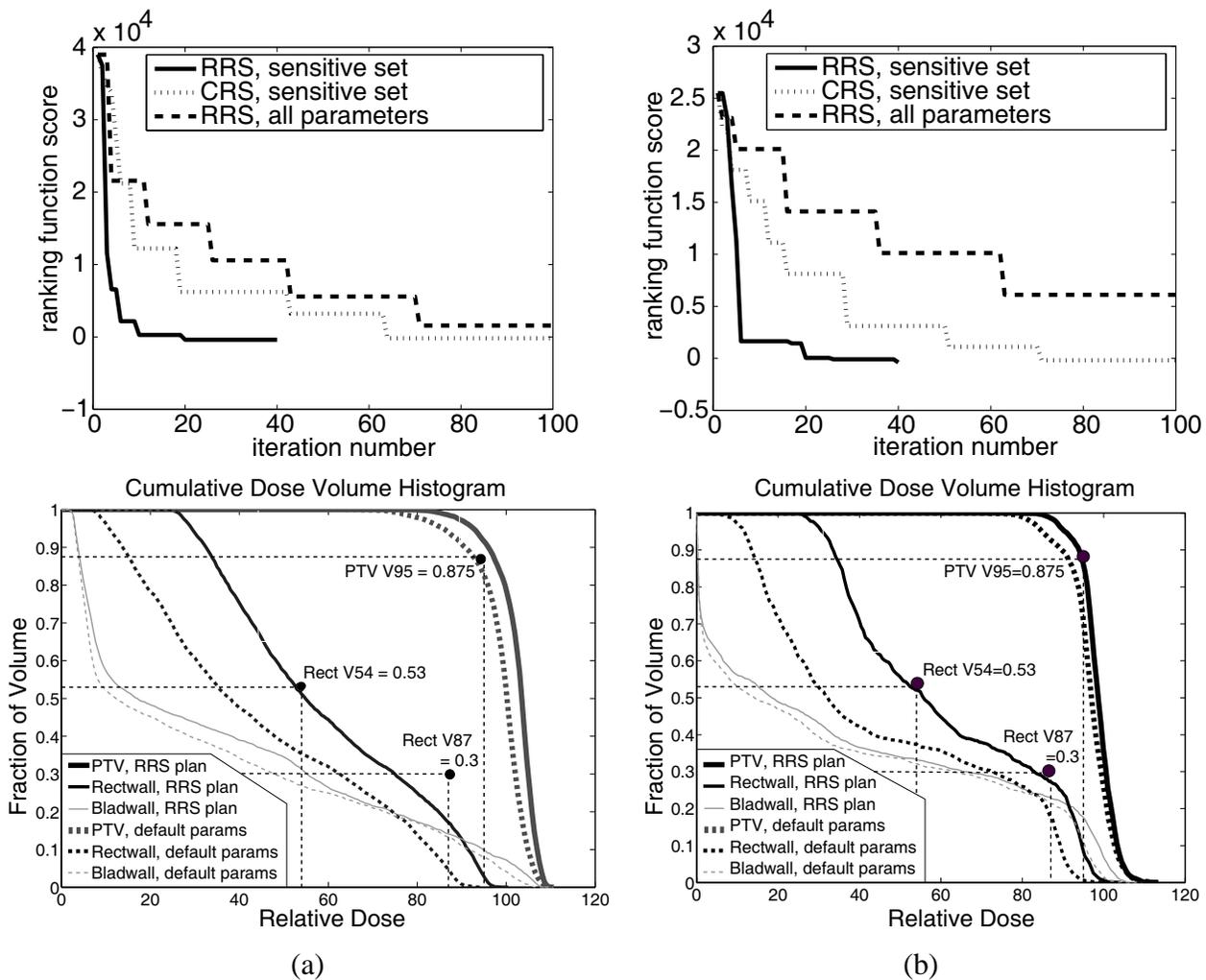


Figure 5: Parameter optimization for (a) Patient 3 and (b) Patient 8. The upper row illustrates the overall score as a function of the iteration number. The lower row compares the DVHs of the prostate, rectum wall and bladder wall for the RRS optimization result vs. the DVHs corresponding to the default parameter settings. The clinical constraints are marked as circled points.

6 statistics from RRS optimization were comparable to those of hand-tuned clinical plans evaluated by full dose calculation. Based on corrected dose calculations, the constraints for all patients were found to be satisfied (resulting in a negative ranking scores), except for slightly higher $D_{max}^{rectwall}$ in 2 of the 36 cases. This is not overly worrisome, since this kind of maximum/minimum statistic is inherently noisy. We expect the scores could be improved if the constraint is replaced with D_{05} or other more robust measures for hotspot.

It is reasonable to expect that in prostate IMRT, the major conflict is between PTV coverage and rectum hotspot, due to the overlap between the two structures. The two constraints V_{95}^{PTV} and D_{min}^{PTV} compete against the rectum constraint $D_{max}^{rectwall}$, which is the most difficult compromise to negotiate in the parameter search. For the “cooler” parts of

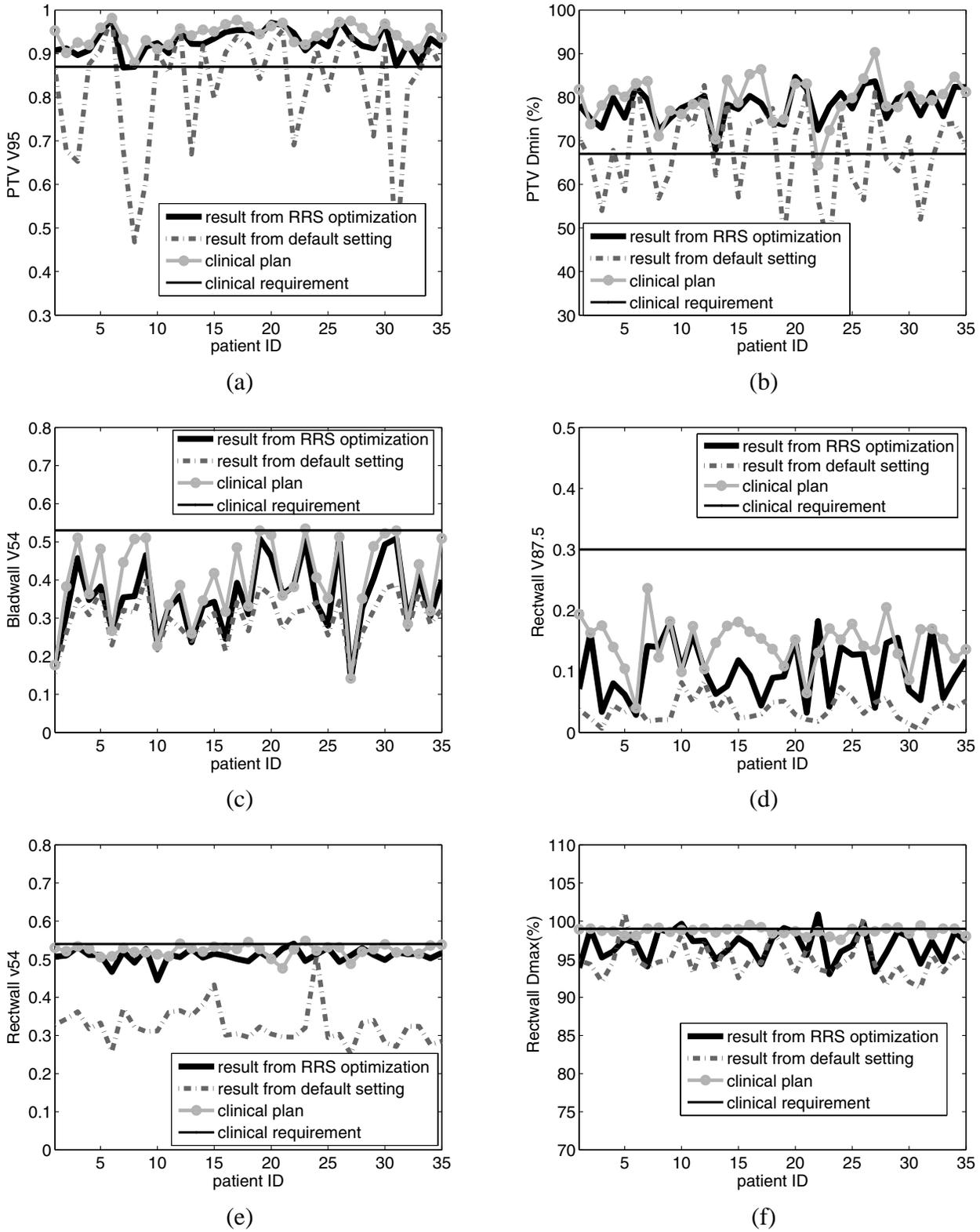


Figure 6: Comparisons of dose evaluation statistics for RRS, default, and clinical parameter settings. The clinical parameters come from the hand-tuned historical plans from MSKCC. (a): V_{95}^{PTV} . (b): D_{min}^{PTV} . (c): $V_{54}^{bladwall}$. (d): $V_{87}^{rectwall}$. (e): $V_{54}^{rectwall}$. (f): $D_{max}^{rectwall}$.

the rectum and bladder (which are also farther away from the PTV), the constraints for $V_{87}^{rectwall}$, $V_{54}^{rectwall}$ and $V_{54}^{bladder}$ are relatively easy to satisfy, even using the default parameter settings. Since these OAR constraints are also negatively correlated with V_{95}^{PTV} and D_{min}^{PTV} , the outer-loop search achieves the compromise by increasing $V_{54}^{rectwall}$ and $V_{54}^{bladder}$ to the clinical limits, while maximizing the PTV coverage.

The number of iterations required for outer-loop convergence ranged from 40 to 180, and for each iteration, the inner-loop intensity optimization takes about 5 seconds on a Pentium 4 1.6 GHZ, 1GB RAM PC. The average time for outer-loop convergence for one clinical case is about 10 minutes. For a typical patient, the inner-loop optimization involves about 10000 total sample points in the PTV and OARs, and the optimization is taken over 2000 beamlet intensities.

3.3. Clinical Validation

The parameter optimization results for 10 of the 36 plans described above were clinically evaluated at MSKCC by an expert planner who examined dose volume histograms and isodose contours on three orthogonal slices through the isocenter. As discussed in the Appendix, the dose results based on corrected dose calculation in Section 3.2 differ slightly from the full dose calculation performed in the clinic. Accordingly, the plans from the outer-loop optimization are not guaranteed to be approved by the planner, since certain dose statistics might violate their constraints after full dose calculation. We defined three categories for the clinical validation of the outer-loop plans:

Category I The outer-loop plan was clinically acceptable after the usual re-normalization of the doses. This re-normalization ensures D_{max}^{PTV} to be 109.5-110% and $D_{max}^{rectwall}$ to be 98.5-99%, and requires a trivial amount of time compared to re-optimization.

Category II The outer-loop plan required only minor parameter changes followed by re-optimization to obtain a clinically acceptable plan. Here we define “minor” adjustment to involve less than 5 re-optimizations and 15 minutes of effort using the MSKCC planning system running on a Pentium 4 3.06 GHZ, 2.5 GB RAM PC.

Category III Starting from the outer-loop plan required more than 15 minutes or 5 re-optimizations to obtain an clinically acceptable plan.

Among the 10 clinically validated plans, 4 were considered to be in Category I. Figures 7(a)-(c) show the full dose distributions for one of the immediately acceptable patients in the sagittal, transversal and coronal planes. For the other 6 plans that were not immediately approved, 3 plans were in Category II and 3 plans were in Category III. Hence, in 70% of the cases, the parameters from the reduced-order optimization resulted in a substantial reduction in planning time.

4. Conclusions

We showed that prostate IMRT treatment planning can be made more intuitive and less time-consuming, based on a sensitivity analysis and an automated, reduced-order search over

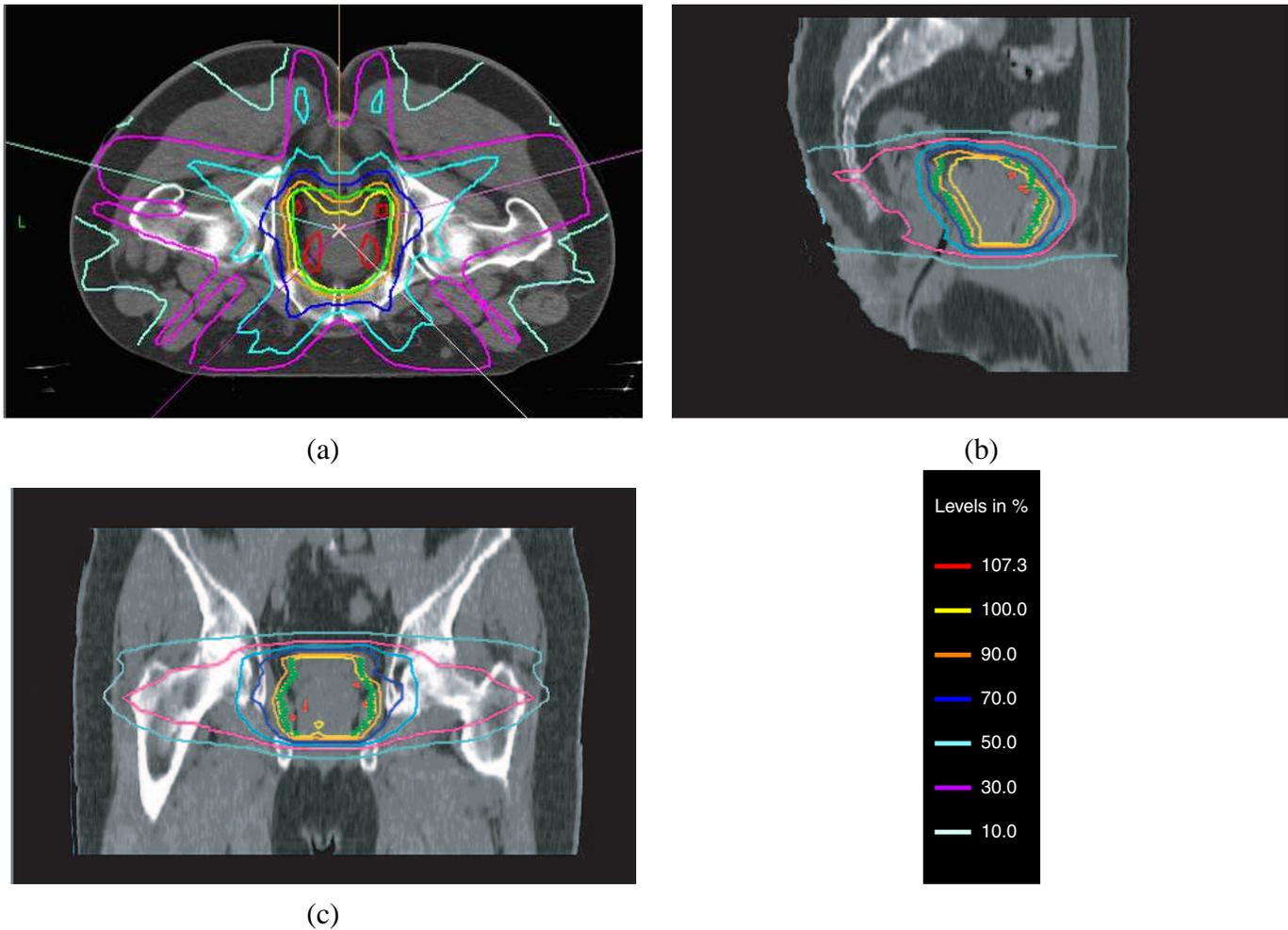


Figure 7: Clinical full dose evaluations for the outer-loop plan for Patient 24. Dose distributions are displayed in the (a) transverse, (b) sagittal, and (c) coronal planes through the isocenter.

the sensitive set. The method is relatively fast and efficient, since we use unconstrained quadratic programming and simplified dose calculations in the inner loop (which takes just seconds and several MB of memory), and stochastic optimization over less than 10 variables in the outer loop. Our methods can be applied to anatomically similar situations such as prostate treatments intended for different total doses, using more treatment beams, or subject to different clinical preferences. We also believe our approach has the potential to improve efficiency for sites where the IMRT planning process is more challenging or tedious, such as head and neck cancer treatment.

As described above, the optimized outer-loop result may not always be clinically acceptable. The optimization may end with a positive score after the decided number of RRS iterations, or the plan may not be approved by planners after full dose calculation and plan evaluation. The first case indicates that a good solution is not available within the sensitive parameter set, and additional parameters should be included in the outer loop. The second case

is caused by the difference between the corrected pencil beam method (in the outer loop) and full dose calculation. However, even in these two cases, we expect the resulting parameters will still provide a better starting point for manual tuning than the default parameters. In the future, we plan to investigate more accurate dose calculation algorithms in the outer loop. We also intend to investigate methods for quickly determining an appropriate sensitive set for a given patient. For example, as mentioned above, the bladder wall DVH parameters had a strong impact on plan quality for some patients, but were not determined to be highly sensitive for the 5 training patients. One possibility might be to add the bladder wall parameters based on a simple rule involving the patient’s organ geometry (e.g., bladder size). We also plan to use our sensitivity analysis to provide a prescriptive basis for further tuning so that given the deficiencies of a candidate plan, planners can more easily determine which parameters should be tuned and to what extent.

A natural idea is to minimize the ranking function (7) directly, using a stochastic algorithm (like RRS) to deal with the non-differentiable nature of the objective function. The major problem with this approach would be dimensionality, since stochastic optimization over thousands of beamlet variables would be extremely time-consuming, compared to stochastic optimization over less than 10 variables in the outer loop of our method.

In discussions with planners, we found that there are some features of the automatically-produced plans that, while not rendering them clinically unacceptable, indicate room for improvement. For example, planners observed that in some cases, they expected that a higher PTV coverage or lower maximum dose in the Mid region could be achievable, and noted that in a “good” plan, the OAR constraints should generally be pushed to their limits. In future work, we will improve our scoring function to further take into account expert planners’ experience (much of which is not explicitly written in the clinical protocol).

While we only considered dose and dose-volume based objective function terms in this work, the same approach should apply to other types of objective function terms as well. For example, a typical dose-response-based objective function to be maximized is given by: $F = \prod_j f_j$, where each term f_j is of the form $f_T = \frac{1}{1 + (\frac{EUD_0}{EUD})^n}$ for tumors or $f_{OAR} = \frac{1}{1 + (\frac{EUD}{EUD_0})^n}$ for normal tissues (Wu et al., 2002). The equivalent uniform dose is defined as $EUD = (\frac{1}{N} \sum_i D_i^a)^{\frac{1}{a}}$, where D_i is the dose to the i^{th} voxel (Niemierko, 1997). In this case, the tunable parameters that could be analyzed include the dose-volume effect parameter a , the desired doses EUD_0 , and the power term n for each structure.

A final goal is to produce a better patient-specific starting point for the outer-loop optimization based on a parametric shape model fit to the organs of interest in CT images, such as the one described in (Jeong and Radke, 2006).

Appendix: Full Dose and Truncated Dose Calculation Algorithms

As described in Section 2.3, while our corrected dose (C-D) calculations enable speed and accuracy in the outer loop optimization, they still differ slightly from the full dose (F-D) calculations performed in the clinic. The assumption of the C-D method is that the full

and truncated doses are locally related by a scaling factor, which, for a given voxel, is approximately constant in optimized plans. As a result of this approximation, the dose outputs from RRS optimization and MSKCC full dose evaluation are not identical. Table 7 compares the dose statistics for 6 patients using both methods. The differences in extreme (i.e., maximum or minimum) dose statistics were generally larger than those in the other DVH statistics, as would be expected, since they only depend on the dose at one sample. In the Mid structure, doses were higher with F-D than with C-D, which makes sense since the Mid structure is thin and is surrounded by target that can scatter into it. For Patient 24, the mean values of the dose differences for all evaluation points in the PTV, rectal wall and bladder wall were -2.14, -1.58 and -0.84, respectively. The standard deviations of the differences were 0.81, 1.41 and 1.83, respectively. Such differences do not necessarily weaken the clinical value of the RRS results, as long as the RRS-generated plans have enough margin beyond the clinical constraints. For Patient 24, the DVH curves from RRS and full dose evaluation were almost identical, as illustrated in Figure 8. After full dose calculation, all the clinical constraints were still satisfied, which suggests that for this case, the approximate dose calculation was accurate enough for parameter optimization.

Patient	V_{95}^{PTV}		D_{min}^{PTV}		$V_{54}^{rectwall}$		$D_{max}^{rectwall}$		D_{max}^{mid}	
	C-D	F-D	C-D	F-D	C-D	F-D	C-D	F-D	C-D	F-D
1	0.94	0.97	82.67	82.93	0.51	0.53	97.36	100.26	99.73	104.81
7	0.99	0.98	84.79	82.51	0.50	0.52	96.34	99.38	105.48	108.05
19	0.95	0.96	75.78	71.29	0.50	0.52	99.45	98.96	103.16	106.56
23	0.94	0.95	71.44	69.23	0.55	0.54	100.47	101.59	102.90	105.59
24	0.91	0.93	77.94	79.61	0.52	0.52	92.97	96.46	102.82	105.86
27	0.98	0.98	83.46	82.49	0.54	0.53	99.60	99.80	105.51	106.56

Table 7: Comparison of the doses from RRS plans using the corrected dose calculation algorithm (C-D), and doses from clinical evaluation using the full dose calculation algorithm (F-D). All doses are expressed as percentages.

Acknowledgments

This work was supported by the National Cancer Institute under grant 5P01CA59017-13, and CenSSIS, the NSF Center for Subsurface Sensing and Imaging Systems, under the award EEC-9986821.

References

- M. Alber, M. Birkner, and F. Nüsslin. Tools for the analysis of dose optimization: II. sensitivity analysis. *Phys Med Biol.*, 47(4):265–270, April 2002.
- M. Alber and F. Nüsslin. Tools for the analysis of dose optimization: I. Effect-volume histogram. *Phys Med Biol.*, 47(14):2451–2458, July 2002.

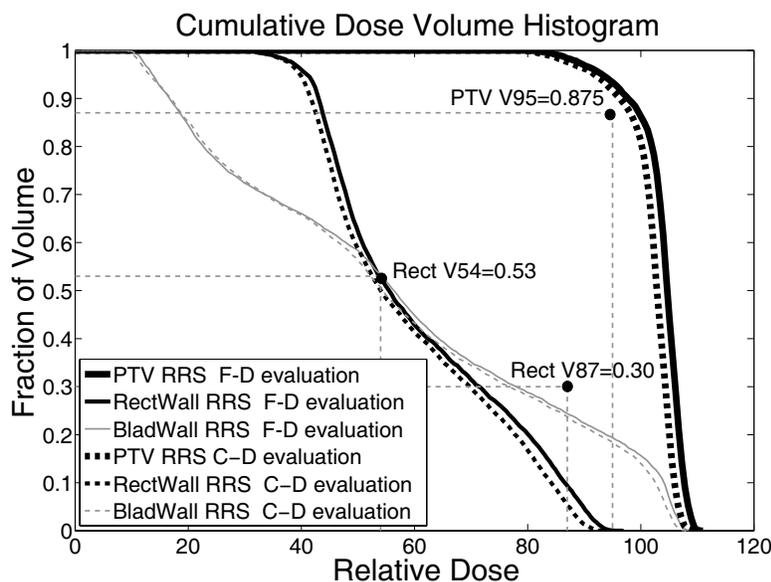


Figure 8: DVHs of the RRS plan using full dose (F-D) calculation (solid line) vs. the corrected-pencil-beam dose (C-D) calculation (dashed line), showing that the differences are slight.

- H. I. Amols and C. C. Ling. EUD but not QED. *International Journal of Radiation Oncology Biology Physics*, 52:1–2, 2002.
- J. Barbiere, M. Chan, J. Mechalakos, D. Cann, K. Schupak, and C. Burman. A parameter optimization algorithm for intensity-modulated radiotherapy prostate treatment planning. *Journal of Applied Clinical Medical Physics*, 3:227–234, 2002.
- J. L. Bedford and S. Webb. Elimination of importance factors for clinically accurate selection of beam orientations, beam weights, and wedge angles in conformal radiation therapy. *Medical Physics*, 30:1788–1804, 2003.
- T. Bortfeld, K.-H. Kufer, M. Monz, A. Trofimov, and A. Niemierko. Problems with current IMRT prescription practices and planning systems [abstract]. *Medical Physics*, 31:1761, 2004.
- C. Cotrutz, M. Lahanas, C. Kappas, and D. Baltas. A multiobjective gradient based dose optimization algorithm for conformal radiotherapy. *Physics in Medicine and Biology*, 46:2161–2175, 2001.
- D. Craft, T. Halabi, and T. Bortfeld. Exploration of tradeoffs in intensity-modulated radiotherapy. *Phys. Med. Biol.*, 50(24):5857–5868, December 2005.
- M. A. Hunt, C.-Y. Hsiung, S. V. Spirou, C.-S. Chui, H. I. Amols, and C. C. Ling. Evaluation of concave dose distributions created using an inverse planning system. *International Journal of Radiation Oncology Biology Physics*, 54:953–962, 2002.
- Y. Jeong and R. J. Radke. Modeling inter- and intra-patient anatomical variation using a bilinear model. *Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 06), in conjunction with CVPR 06, New York, NY, 2006.*
- M. L. Kessler, D. L. Mcshan, M. A. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass. Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans. *Optimization and Engineering*, 6(4):421–448, December 2005.
- M. Kutner, C. Nachtsheim, and J. Neter. *Applied Linear Regression Models*. McGraw-Hill, 2004.
- M. Langer, E. Lee, J. Deasy, R. Rardin, and J. Deye. Operations research applied to radiotherapy, an NCI-

- NSF-sponsored workshop, February 7-9, 2002. *Int J Radiat Oncol Biol Phys.*, 57(3):762–768, November 2003.
- J. Lian and L. Xing. Incorporating model parameter uncertainty into inverse treatment planning. *Med Phys.*, 31(9):2711–2720, September 2004.
- C. C. Ling et al. *A Practical Guide to Intensity-Modulated Radiation Therapy*. Medical Physics Publishing, 2004.
- J. Meyer, M. Phillips, P. Cho, I. Kalet, and J. Doctor. Application of influence diagrams to prostate intensity-modulated radiation therapy plan selection. *Phys Med Biol.*, 49(9):1637–1653, May 2004.
- A. Niemierko. Optimization of inverse treatment planning using a fuzzy weight function. *Med Phys.*, 24(1):103–110, January 1997.
- J. R. Palta et al. *Intensity-Modulated Radiation Therapy: The State of the Art*. Medical Physics Publishing, 2004.
- W. L. Price. Internet traffic engineering - global optimization by controlled random search. *J. Optimiz. Theory App*, 41:333–348, 1983.
- J. Purdy et al. Intensity-modulated radiotherapy: current status and issues of interest. *Int J Radiat Oncol Biol Phys.*, 51(4):880–914, November 2001.
- 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. *Medical Physics*, 48(21):3521–3542, November 2003.
- 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.
- A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, 2004.
- A. Samuelsson and K. Johansson. Intensity modulated radiotherapy treatment planning for dynamic multileaf collimator delivery: influence of different parameters on dose distributions. *Radiother Oncol.*, 66(1):19–28, January 2003.
- D. Shepard, M. Ferris, G. Olivera, and T. Mackie. Optimizing the delivery of radiation to cancer patients. *SIAM Review*, 41:721–744, 1999.
- J. Siebers, M. Lauterbach, S. Tong, Q. Wu, and R. Mohan. Reducing dose calculation time for accurate iterative IMRT planning. *Medical Physics*, 29(2):231–237, February 2002.
- S. Spirou and C. Chui. A gradient inverse planning algorithm with dose-volume constraints. *Medical Physics*, 25(3):321–333, March 1998.
- G. Starkschall, A. Pollack, and C. Stevens. Treatment planning using a dose-volume feasibility search algorithm. *Int J Radiat Oncol Biol Phys.*, 49(5):1419–1427, April 2001.
- S. Webb. *Intensity-Modulated Radiation Therapy*. Taylor and Francis, 2001.
- Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *Int J Radiat Oncol Biol Phys.*, 52(1):224–235, January 2002.
- X. Wu and Y. Zhu. An optimization method for importance factors and beam weights based on genetic algorithms for radiotherapy treatment planning. *Phys Med Biol.*, 46(4):1085–1099, April 2001.
- L. Xing, J. Li, S. Donaldson, Q. Le, and A. Boyer. Optimization of importance factors in inverse planning. *Physics in Medicine and Biology*, 44:2525–2536, 1999a.
- L. Xing, J. Li, A. Pugachev, Q. Le, and A. Boyer. Estimation theory and model parameter selection for therapeutic treatment plan optimization. *Med Phys.*, 26(11):2348–2358, November 1999b.
- T. Ye and S. Kalyanaraman. Internet traffic engineering - a recursive random search algorithm for large-scale network parameter configuration. *ACM SIGMETRICS Perf. Eval. Rev.*, 31:10, 2003.
- Y. Yu. Multiobjective decision theory for computational optimization in radiation therapy. *Medical Physics*, 24:1445–1454, 1997.