Winter 12-15-2014

Applications of Nonlinear Optimization

Yao Xie
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Mathematics

Dissertation Examination Committee:

M Victor Wickerhauser, Chair

Yixin Chen

Joseph O Deasy

Renato Feres

Ed Spitznagel

Blake Thornton

Applications of Nonlinear Optimization

by

Yao Xie

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2014

St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest appreciation to my committee chair Prof. Victor Wickerhauser, who is an excellent advisor and a great mathematician. I learned so much from him on how to think and analyze problems in a mathematical way. Without his guidance and help, this dissertation would not have been possible.

I would like to give my special thanks to Dr. Joseph O Deasy from Memorial Sloan-Kettering Cancer Center in New York, who is very thoughtful, smart and creative. He introduced me to the radiotherapy world, have given me guidance on so many interesting projects, and even helped me a lot on how to write scientific papers.

I would like to thank my co-advisor Prof. Yixin Chen from Computer Science Department at Washington University, who is superb in mathematical optimization. Besides his help on optimization theories, he has given me a lot of guidance on other Computer Science areas such as machine learning.

In addition, I would like to thank my labmate Paras Tiwari, who is a Ph.D. student in Computer Science Department at Washington University. He is a great team-worker. It is always joyful and inspiring to work together with him.

I would like to give my thanks to my former Ph.D. advisors Prof. Jimin Ding, Prof. Nan Lin, and Prof. John McCarthy, to all my course teachers, to all staff at Mathematics Department, and to all my classmates and friends.

Dedicated to My Parents.

ABSTRACT OF THE DISSERTATION

Applications of Nonlinear Optimization

by

Xie, Yao

Doctor of Philosophy in Mathematics,

Washington University in St. Louis, 2014.

Professor M Victor Wickerhauser, Chair

We apply an interior point algorithm to two nonlinear optimization problems and achieve improved results. We also devise an approximate convex functional alternative for use in one of the problems and estimate its accuracy.

The first problem is maximum variance unfolding in machine learning. The traditional method to solve this problem is to convert it to a semi-definite optimization problem by defining a kernel matrix. We obtain better unfolding and higher speeds with the interior point algorithm on the original non-convex problem for data with less than 10^4 points.

The second problem is a multi-objective dose optimization for intensity modulated radiotherapy, whose goals are to achieve high radiation dose on tumors while sparing normal tissues. Due to tumor motions and patient set-up errors, a robust optimization against motion uncertainties is required to deliver a clinically acceptable treatment plan. The traditional method, to irradiate an enlargement of the tumor region, is very conservative and leads to possibly high radiation dose on sensitive structures. We use a new robust optimization model within the framework of goal programming that consists of

multiple optimization steps based on prescription priorities. One metric is defined for each structure of interest. A final robustness optimization step then minimizes the variance of all the goal metrics with respect to the motion probability space, and pushes the mean values of these metrics toward a desired value as well. We show similar high dose coverage on example tumors with reduced dose on sensitive structures.

One clinically important metric for a radiation dose distribution, that describes tumor control probability or normal tissue complication probability, is D_x, the minimum dose value on the hottest $x\%$ of a structure. It is not mathematically well-behaved, which impedes its use in optimization. We approximate D_x with a linear function of two generalized equivalent uniform dose metrics, also known as l_p norms, requiring that the approximation is concave so that its maximization becomes a convex problem. Results with cross validation on a sampling of radiation therapy plans show that the error of this approximation is less than 1 Gy for the most used range 80 to 95 of x values.

1. Introduction to Nonlinear Optimization

A (mathematical) optimization problem is defined in the following way:

Given a function $f : X \to \mathbb{R}$ from some set X to the real numbers, find $x_0 \in X$ so that $f(x_0) \leq f(x), \forall x \in X$ (minimization). Maximization problems can be trivially converted to minimization problems. In short, the problem is denoted by $\min_{x \in X} f(x)$. The argument x_0 that gives minimum function value is denoted as $x_0 = \text{argmin}_{x \in X} f$. The function f is called the *objective function*, and the set X is called the *search space* of the optimization problem.

The search space X can be any set, but the ones we consider are subsets of Euclidean space defined by some constraints. The optimization of interest can thus be described as:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x), \tag{1.1}$$

$$\text{s.t. } h(x) = 0, \tag{1.2}$$

$$g(x) \leq 0, \tag{1.3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$, and $h : \mathbb{R}^n \to \mathbb{R}^m$.

Linear and nonlinear optimization

The above optimization is a *linear optimization* if f, h, g are all linear. It is a *nonlinear optimization* if any of f, h, g is nonlinear. Linear optimization is relatively easy to solve with very fast algorithms available such as Simplex algorithm, while nonlinear optimization is harder to solve but has broader applications [1].

An optimization problem is *convex* if f is a convex function and the set X is a convex set, or equivalently, when both f and g are convex functions and h is linear. A useful

property of convex optimization is that any local optimal solution is also the global optimal solution. In practice, it is preferred to formulate a convex optimization.

Multi-objective optimization

For real application problems, there are usually more than one objective function. A *multi-objective optimization* is an optimization problem with multiple objectives, and is defined as:

$$\min\ (f_1(x), f_2(x), \cdots, f_k(x)) \tag{1.4}$$

$$\text{s.t. } x \in X, \tag{1.5}$$

where $f_i(x) : \mathbb{R}^n \to \mathbb{R}$, and $X \subset \mathbb{R}^n$. Interesting problems are those optimization with conflicting objective functions. A point $x_0 \in X$ is called a *Pareto optimal* solution if there is no $x \in X$ such that $f_i(x) \leq f_i(x_0), \forall i = 1, \cdots, k$ and $f_{i_0}(x) < f_{i_0}(x_0)$, for some i_0. The set of all Pareto optimal solutions is called a *Pareto surface*.

Solving a multi-objective optimization is much harder than solving a single objective optimization. If one problem has several objectives, it is hard (sometimes computationally impossible) to find the whole Pareto surface and then choose the best one that respects the priorities of the conflicting objective functions. If one chooses one solution from part of the Pareto surface, there might be a solution that sacrifices a little on one objective but gains a lot on another objective. Moreover, judging whether one solution is better than the other may require introducing additional equations or even subjective evaluation.

Optimization solver

There are a variety of optimization solvers including both commercial and free solvers. We adopt an solver named IPOPT (Interior Point Optimization) as it is open-sourced and can solve both convex and non-convex problems with functions that are twice continuously differentiable [2].

2. Application in Machine Learning

Our first application is in machine learning.

Machine learning is a relatively new area in Computer Science with origins in the 1950s. As computer hardware advanced rapidly, machine learning has become a large subfield of computer science and artificial intelligence, with applications in computer vision, natural language processing, medical diagnosis, bioinformatics, computational finance, etc. [3]. Tom M. Mitchell made the widely quoted definition of machine learning as: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [4]. Optimization approaches play important roles in this learning process due to their widespread applications and nice theoretical properties [5, 6]. Together with statistics and machine learning, (nonlinear) optimization helps to understand data, explain data, and make value from data.

2.1 Introduction to Maximum Variance Unfolding

Nowadays, practical problems often have high-dimensional data. For example, DNA microarray technology can produce a large number of measurements at once. If a word-frequency vector is used, the number of dimensions of text documents equals the size of the vocabulary.

Practical high-dimensional data are usually self-correlated and there is a need to project the data to a low dimensional space to remove the correlation or to visualize the data. For example, data collected from DNA microarray has spacial correlation, and vocabulary are correlated due to the same letters.

Linear and nonlinear dimensionality reduction (NLDR) methods are developed to project data from high-dimensional spaces to low-dimensional spaces. The well-known linear dimensionality reduction algorithm, principal component analysis (PCA), uses an orthogonal transformation to convert high dimensional data into linearly uncorrelated principal components [7]. As a generalization of PCA, NLDR assumes that the data lies close to an embedded nonlinear manifold within the high dimensional space. NLDR is thus also called manifold learning or graph embedding.

NLDR is widely studied with lots of applications. Many algorithms have been designed [8–14]. There are also various applications of NLDR. A straightforward application is in robot design. A robot analyzes image data and learns a two dimensional manifold (position and orientation). NLDR has applications in the generalization of robots: dynamical systems [15]. Another application is about A* search, which is a computer algorithm that is widely used in path finding and graph traversal in graph theory [16]. NLDR learns the Euclidean embedding of a search space graph as a heuristic for A* search [17]. Better heuristics greatly improve the speed of A* search [18].

Among many NLDR algorithms [8], maximum variance unfolding (MVU) utilizes a property that when an embedded manifold is properly unfolded, the variance of the points is maximized [12]. It also uses a technique for casting the non-convex optimization as a semidefinite programming problem. For this reason, MVU is also called semidefinite embedding.

The major benefit of MVU is that it provides a heuristic that is both admissible (distances are not over estimated) and consistent (distances satisfying the triangular inequality), so that the embedding can be an A* heuristic.

MVU maximizes the variance of the embedding as follows:

$$\underset{x_1,\cdots,x_n\in\mathbb{R}^d}{\text{maximize}} \sum_{i=1}^{n} x_i^2 \tag{2.1}$$

$$\text{s.t. } \|x_i - x_j\| \leq d_{ij}, \qquad (i,j) \in E, \tag{2.2}$$

$$\sum_{i=1}^{n} x_i = 0, \tag{2.3}$$

where $x_i \in \mathbb{R}^d$ is a d dimensional vector. Equation 2.3 normalizes the points to make them invariant under translation.

This problem is non-convex and Weinberger et al. intelligently used one technique to convert it to a semidefinite programming problem [12]. They define kernels $k_{ij} = x_i^T x_j$ and convert the optimization to be:

$$\underset{K}{\text{maximize}}\ trace(K) \tag{2.4}$$

$$\text{s.t. } k_{ii} - 2k_{ij} + k_{ij} \leq d_{ij}^2, \qquad (i,j) \in E, \tag{2.5}$$

$$\sum_{i,j} k_{ij} = 0, \tag{2.6}$$

$$K \succeq 0, \tag{2.7}$$

where K is the matrix consisting of k_{ij}, and $K \succeq 0$ means K is a semidefinite matrix.

Fig. 2.1 is a visualization of maximum variance unfolding. The upper graph represents a data set where some nodes are connected by edges. The output of MVU (lower figure) captures the information about which nodes should be connected and unfolds the graph by utilizing the freedom of unconnected points.

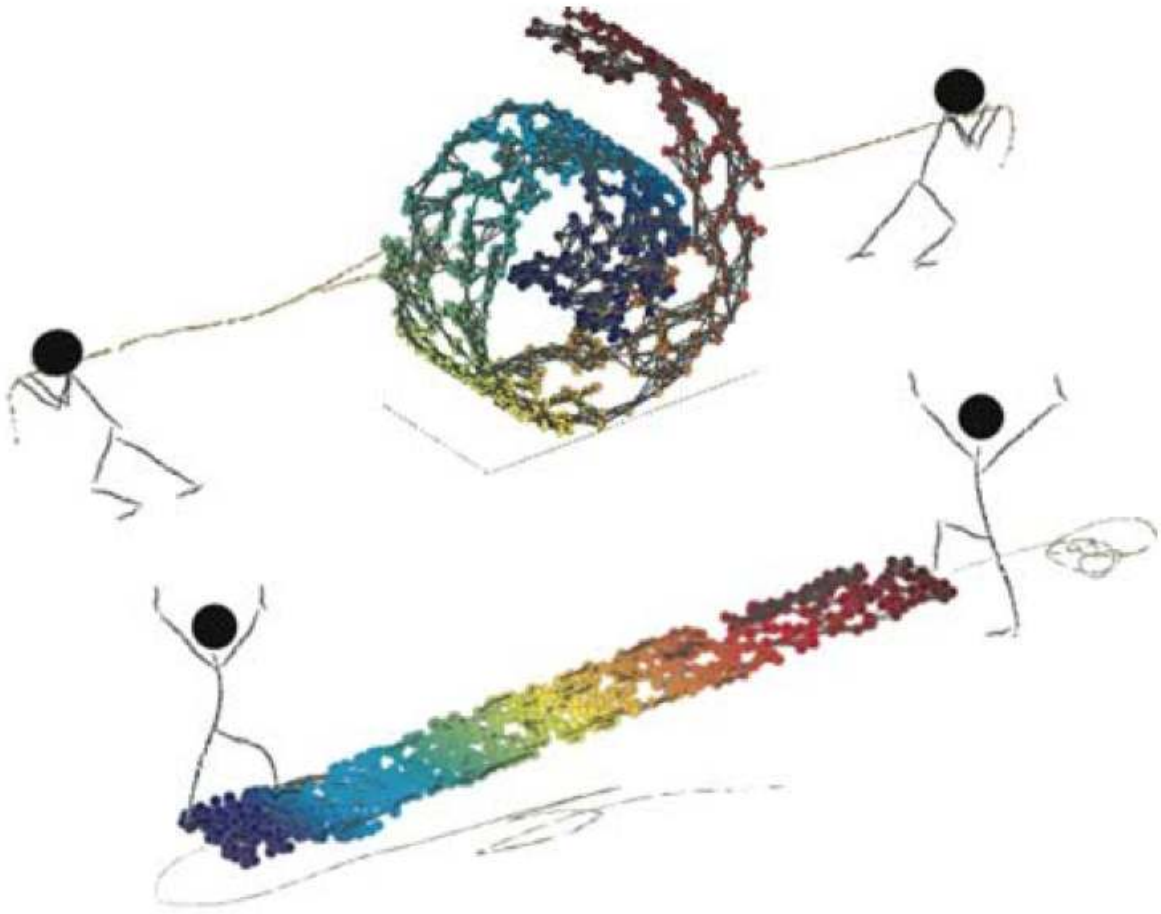

Figure 2.1. Visualization of maximum variance unfolding. Fig. 1 of the paper by Weinberger et al. [12].

Table 2.1 shows that MVU greatly improves A* search. This is extremely helpful in applications that have repeated search tasks (e.g. GPS routing) with low powered devices (e.g. smartphones).

However, MVU itself is very slow as it tries to solve a semidefinite problem. Actually, it is among the least scalable manifold learning algorithms and cannot embed graphs with more than 4000 points. This greatly reduces its usefulness for real applications. There are refinements to MVU. The LandmarkMVU uses landmarks to increase speed with

	game map	6-blocksworld
Differential heuristic	1	1
Eigenmap	0.32	0.88
Isomap	0.50	12.13
MVU	1.12	37.27

Table 2.1
Relative A* search speed up using four heuristics [18].

some cost to accuracy [19]. Maximum Variance Correction (MVC) decomposes MVU into independent sub-problems and scales up MVU embedding greatly [18].

2.2 Improve MVU by IPOPT

Our idea is to use an interior point method solver for non-convex optimization problems to solve MVU directly.

Consider the original problem formulation of MVU:

$$\underset{x_1,\cdots,x_n\in\mathbb{R}^d}{\text{maximize}} \sum_{i=1}^{n} x_i^2 \tag{2.8}$$

$$\text{s.t. } \|x_i - x_j\| \leq d_{ij}, \qquad (i,j) \in E, \tag{2.9}$$

$$\sum_{i=1}^{n} x_i = 0. \tag{2.10}$$

This optimization problem is non-convex and thus hard to solve. As mentioned in Sec. 2.1, the conventional MVU method converts this formulation to a semi-definite programming problem and solves the problem using a semi-definite optimization solver. But it also results in slow computation and inaccurate solution. The solution is inaccurate

because after running the optimization it is needed to derive the actual coordinates from the kernels, i.e., solve $k_{ii} = x_i^T x_i$ for x_i.

The nonlinear optimization solver IPOPT uses interior point methods to solve optimization problems. It works for convex problems and finds very good solutions for non-convex problems as well. Instead of converting the MVU problem to a semi-definite problem, we solve the problem directly with IPOPT.

To quantify the 'unfolding' quality, it is necessary to directly compare the objective function values of solving MVU with IPOPT and other methods.

This method is tested on three well-known benchmark artificial intelligence problems: 'swiss roll' problem [12], N-Puzzle problem [20] and blocks world problem [21].

A 'swiss roll' data set has 8 dimensions, consisting of the three dimensions shown in Fig. 2.2, plus five extra dimensions filled with low variance Gaussian noise.

N-puzzle is a sliding puzzle that consists of N randomly ordered tiles with one tile missing. An action is to slide a tile to the empty space. The task is to find the shortest action sequence from a pre-defined tile order to a goal tile order.

As stated by Jones et al., the definition of block world problem is as follows: "The objects in the problem domain include a finite number of cubical blocks, and a table large enough to hold all of them. Each block is on a single other object (either another block or the table). For each block b, either b is clear or else there is a unique block a sitting on b. There is one kind of action: move a single clear block, either from another block onto the table, or from an object onto another clear block. As a result of moving b from c onto d, b is sitting on d instead of c, c is clear (unless it is the table), and d is not clear (unless it is the table). A problem in this domain is specified by giving two sets of ground atoms, one specifying an initial state of the world, and the other specifying necessary and

sufficient conditions for a state to be a goal state. A solution to this problem is a plan (i.e., a sequence of move actions) capable of transforming the initial state into a state satisfying the goal conditions [21]."

Our algorithm is evaluated for 'swiss rolls' with 500 and 2,000 points, 6-blockworld (6 blocks, 4,051 states/points, 21,300 edges), 7-blockworld (7 blocks, 37,633 states/points, 235,074 edges), 7-puzzle (20,160 states/points, 50,400 edges) and 8-puzzle (181,440 states/points, 483,840 edges).

2.3 Results

We first test on the simple 'swissroll' to visualize the model performance. A 'swiss roll' is simulated and it is assumed that each point is connected to its nearest 3 neighbors. In other words, if we assume the x_j is one of the 3 nearest neighbors of x_i, d_{ij} is set as the distance of x_j and x_i in the 'swiss roll'. The objective is to unfold the 'swiss roll'.

Fig. 2.2 shows the simulated 'swiss roll' with 500 points. Fig. 2.3 presents the solution of coordinates after running the optimization, where the 'swiss roll' is unfolded.

Fig. 2.4 and Fig. 2.5 show the results with more points and thicker 'swiss roll' along y-axis. We can see a similar pattern that the 'swiss roll' has been unfolded.

Table 2.2 compares running MVU with IPOPT and with other methods. Running MVU with IPOPT gives the highest variance value (illustrated by the bold font). When the problem size is not very large (graph with less than $50,000$ edges), the optimization is very fast.

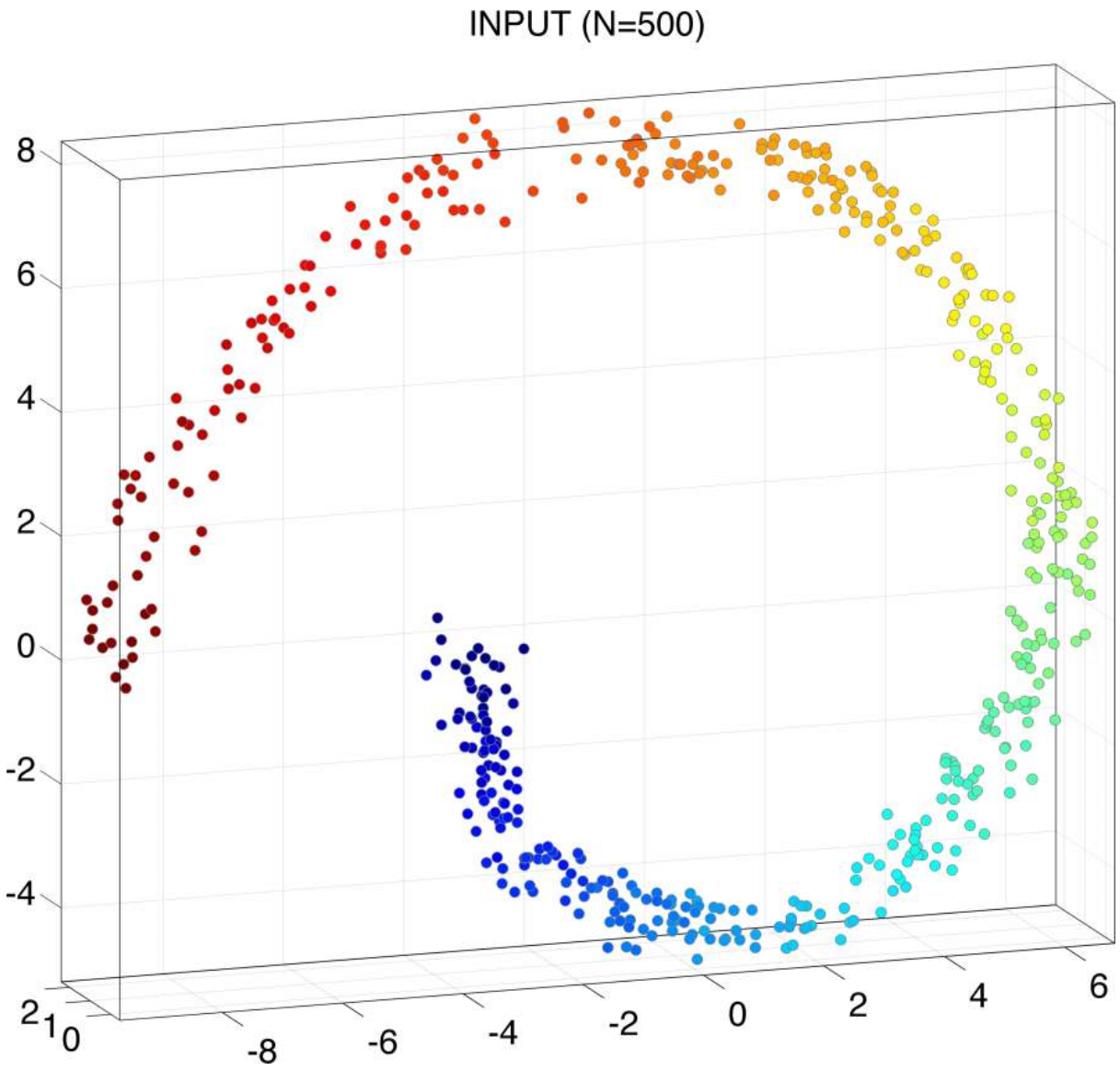

Figure 2.2. Simulated 'swiss roll' with 500 points.

2.4 Discussion and Conclusion

Because MVU problems are non-convex, traditional methods transform MVU problems to semi-definite problems by defining a kernel matrix. This makes the optimization process slow and the solution inaccurate. Instead of transforming the problem, we solve the original problem formulation directly by using an interior point solver IPOPT. The results show that running MVU with IPOPT is more accurate and fast when the graph

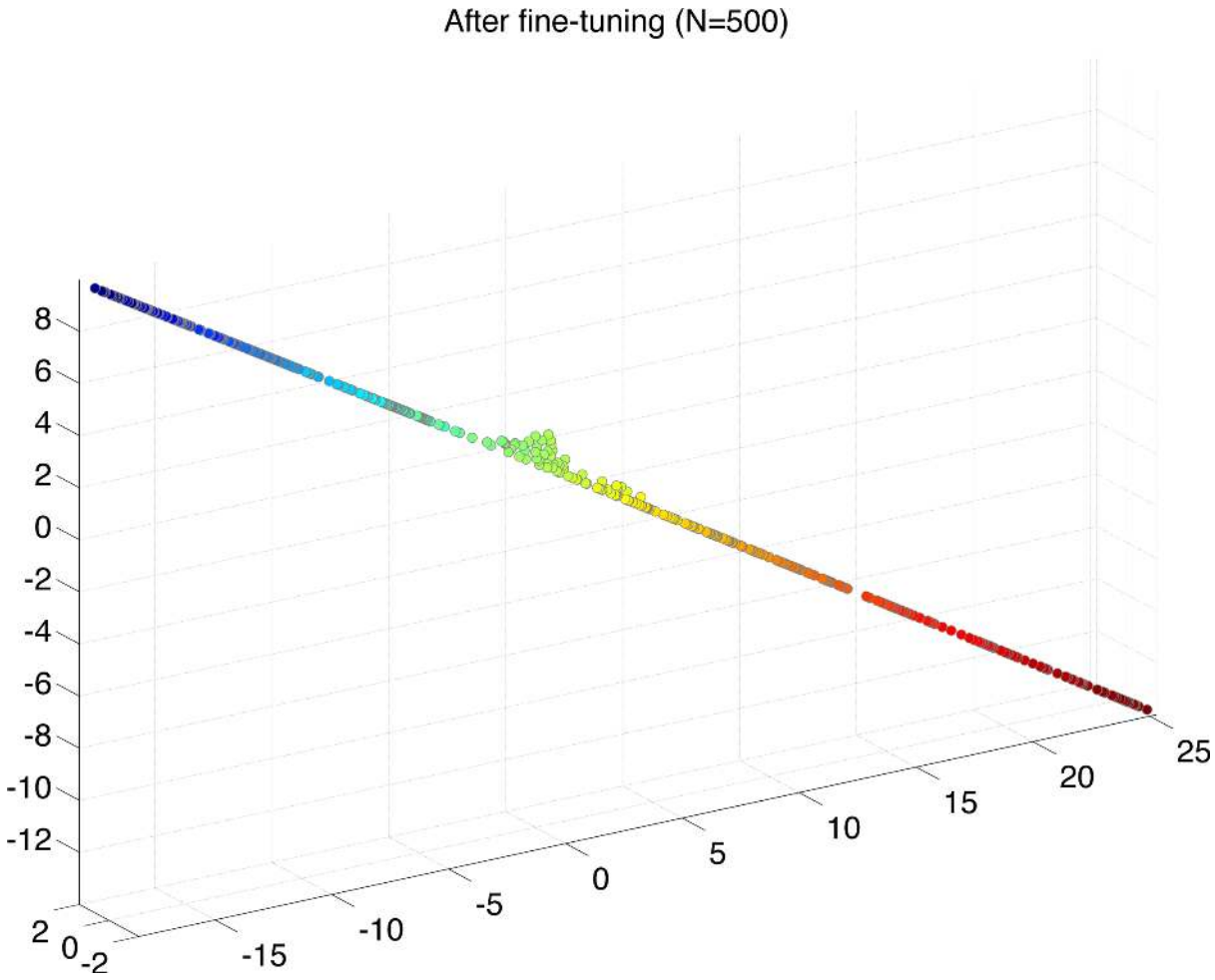

Figure 2.3. Unfolding of the simulated 'swiss roll' with 500 points.

is not too big (at most 50,000 edges). To scale up this method, it is possible to apply MVC to this method. As MVC breaks the original problem into an equivalent system of sub-problems, MVU with IPOPT should be adopted to solve these sub-problems fast and accurately.

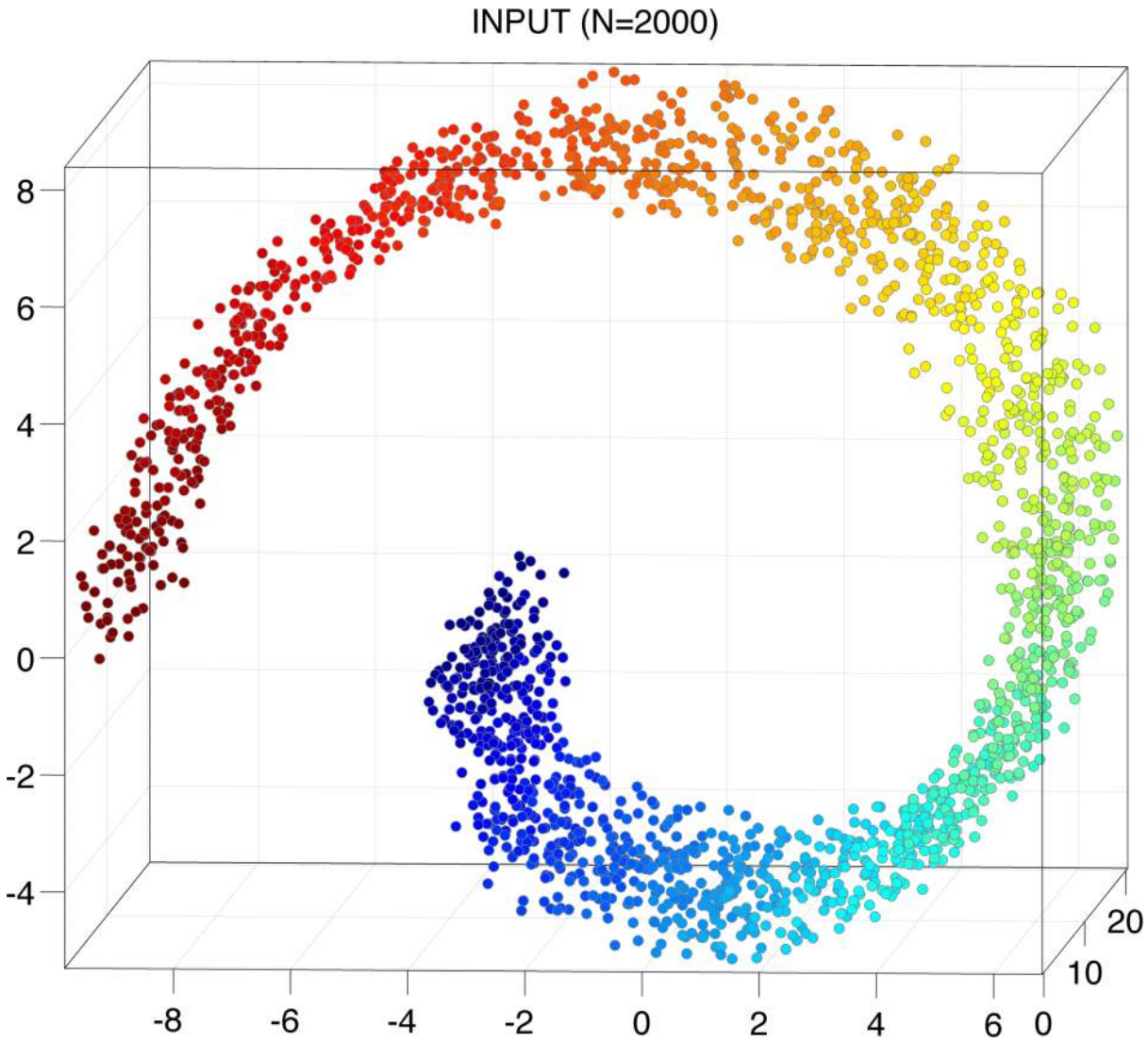

Figure 2.4. Simulated 'swiss roll' with 2000 points.

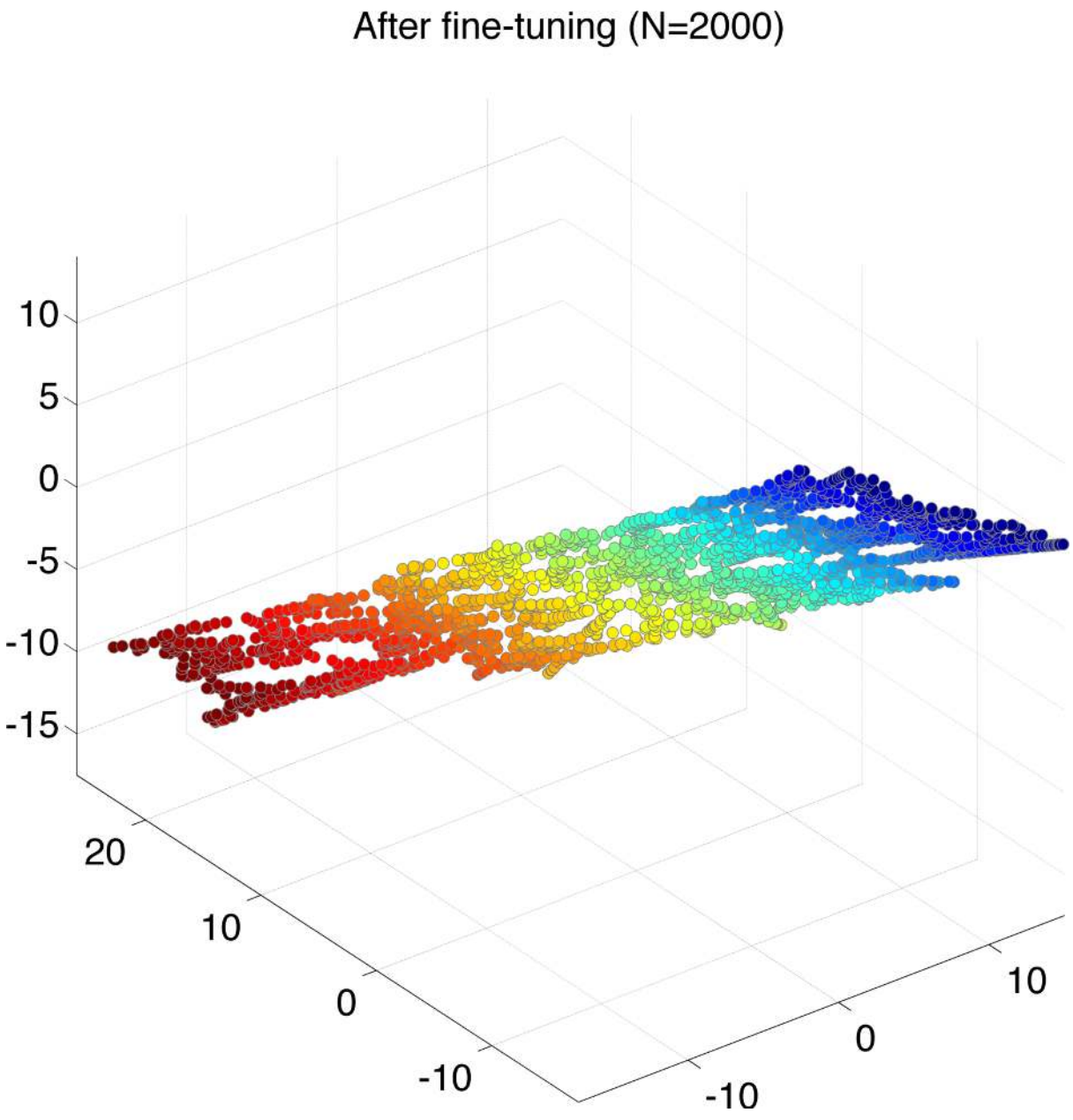

Figure 2.5. Unfolding of the simulated 'swiss roll' with 2000 points.

	6-blockworld		7-puzzle		7-blockworld		8-puzzle	
	Time	Var	Time	Var	Time	Var	Time	Var
gl-MVU	1m	0.138	1m	6.66	8m	1.20	15m	17.79
MVC-10(gl-mvu)	3m	0.27	18m	9.82	14m	2.27	3h	45.75
MVC(gl-mvu)	5m	0.30	26m	9.86	33m	2.95	7h	49.06
MVU(IPOPT)	4m	**0.34**	26m	**10.12**	2d 5h	**3.91**	n/a	n/a

Table 2.2
Comparison of embedding time and variance for different data sets. MVC-10 means running MVC with 10 iterations and MVC means running MVC until convergence. Variance has unit 10^5.

3. Applications in Radiation Therapy

The main contributions of the dissertation are the applications of nonlinear optimization to radiation therapy.

3.1 Introduction to Radiation Therapy

Radiation therapy, or *radiotherapy*, is the therapy using ionizing radiation to kill or control malignant cells, as part of cancer treatment, while sparing healthy tissues. Ionizing radiation works by damaging DNA of cancerous tissue leading to cellular death.

The external beam radiation therapy (EBRT or XRT) is one division of radiation therapy on which this dissertation focuses. By its name, the external beam radiation therapy places radiation sources outside the body. Other divisions include brachytherapy, which places radiation sources inside the body, and systemic radioisotope therapy, which gives systemic radioisotopes to patients by infusion or oral ingestion.

3-dimensional conformal radiation therapy (3DCRT), one type of EBRT occurring after the invention of CT, shapes each radiation beam to fit the profile of the tumor from a beam's eye view using a multi-leaf collimator and uses multiple beams to achieve high dose on tumor. It makes full use of the ability to delineate tumors and normal organs in 3-D using CT and planning software.

Proposed by Brahme and colleagues in the mid 1980s [22], Intensity Modulated Radiation Therapy (IMRT), the next generation of 3DCRT, is an advanced high-precision radiation therapy. IMRT not only shapes the beam, but also modulates dose intensity

distribution of the beam. The beam is divided into small grids, each of which is called a *beamlet*, and IMRT modulates the intensity of each beam. 3DCRT can deliver high dose for a tumor with a convex shape, while IMRT can deal with convex and concave shaped tumors with high precision.

Generally, the IMRT radiation therapy process consists of the following steps: a patient receives a CT (computed tomography) scan, resulting in tomographic images of the patient's body. Given these images, physicians delineate the tumor and organs of interest and prescribe the dose protocol for the tumor and every normal organ. Medical physicists then choose radiation beam directions and run an IMRT optimization to get a beam profile that meets the protocol. Then patients receive radiation treatment for about seven weeks. A standard dose of 2 Gray (Gy, joules of energy deposited per kg) is radiated to the patient per fraction [23–25].

Refer to [26] for the history of radiotherapy.

3.1.1 Intensity Modulated Radiation Therapy Optimization

An important and unsolved problem in radiation therapy is how to simultaneously achieve the difficult needs of radiotherapy treatment planning, that is, to solve a multi-objective nonlinear optimization problem. This 'planning problem' has been the focus of a huge amount of research since IMRT was originally proposed by Brahme and colleagues [22] in the mid-1980s. Investigators realized that they could reformulate the treatment planning problem into the form of a computer optimization problem. This resulted in the 'IMRT revolution' [27]. Today, IMRT is by far the most common type of external beam radiotherapy delivered in the U.S. Typically, the mathematical planning problem would state that a certain function needs to be minimized by controlling the detailed shapes of

the incident radiation beams. Theoretically, if one knew how to predict tumor control probability, along with the probability of potential treatment side-effects, as a function of the shape and intensity of the dose distribution, then the optimization algorithm could simply pick out the 'best' treatment plan [28–30].

Before diving into the mathematical formula of IMRT optimization, it is better to discuss some basic properties of IMRT optimization.

IMRT optimization is a multi-objective optimization problem. There are multiple goals to make a good treatment plan. The major goal is to achieve high uniform dose on tumors to control or kill these cells. The next goal is to spare some normal but sensitive organ such as the brainstem for head and neck cases, or the rectum for prostate cases. Without sparing dose and keeping the functionality of these sensitive structures, it is often useless to control the tumor. Then, there are organs that can withstand relatively higher dose, but they still prefer radiation dose as low as possible. There are yet more clinical objectives and almost all these goals contradict with each other.

IMRT optimization is a nonlinear optimization problem. There are certain functions or metrics that physicians care about, which have also been proven to be highly correlated with organ functionality or tumor control. Most of these functions are not linear. Dose homogeneity is preferred on tumors and the usual best function to describe it is the mean square error. Another function D_x, essentially an order statistic, is the minimum dose of the hottest $x\%$ of one organ. The value of D_{95} for the tumor is widely used to assess one treatment plan. D_x has no analytic form and is approximated by some nonlinear functions in order to be used in optimization.

IMRT is a large scale optimization problem. Each beam is divided into roughly 10^3 beamlets (0.25cm x 0.5cm) and there are about 7 beams. Organs are divided into small

units (called voxels) of size about 0.2 cm $\times$ 0.2 cm $\times$ 0.3 cm (CT thickness). For each treatment plan, there are then voxels of order 10^6 and beamlets of order 10^4. In other words, there are 10^4 variables and 10^6 constraints if all voxels are included. Considering the non-linearity of IMRT, this is a very large scale optimization problem.

The mathematical formulation of IMRT is based on the so-called *influence matrix*, a matrix whose element in row i and column j stores the dose received by voxel i from beamlet j with unit dose. It is a sparse matrix of order $10^6 \times 10^4$. Ahead of IMRT optimization, the influence matrix is computed based on CT information for each patient. Given a beamlet intensity solution (or beam profile), the dose distribution within the patient is just the product of the influence matrix and the beam profile. By setting objective functions and constraints on this dose distribution, IMRT can be easily formulated as a multi-objective nonlinear optimization problem.

The IMRT formulation is as follows: Suppose there are k structures of interest for one cancer case, denoted by $S_1, S_2, \cdots, S_k$. Each structure of interest has its function or metric $f_1, f_2, \cdots, f_k$ that represents its functionality (normal organ) or its tumor control (tumor). Denote the influence matrix by A and the beam profile by w. Then the IMRT formulation is a multi-objective optimization:

$$\min \ (f_1(S_1), f_2(S_2), \cdots, f_k(S_k)) \tag{3.1}$$

$$\text{s.t. } 0 \leq w \leq \overline{w}, \tag{3.2}$$

$$Aw \leq \overline{d}, \tag{3.3}$$

where $\overline{w}$ and $\overline{d}$ are upper bounds of beamlet intensity and voxel dose. There might be more functions for each organ and more constraints besides the two in the above formulation.

It is worthwhile to mention and define some of the popular functions that characterize structures. Some of these are very simple functions, such as minimum dose, maximum dose, or mean dose. Table 3.1 is a typical table of objective values of these functions [31].

Structure	Min Dose	Max Dose	Mean Dose	Median Dose	Std Dev
Rect_wall	582.8 cGy	7352.8 cGy	3935.5 cGy	3544.5 cGy	2062.8 cGy
Blad_wall	100.1 cGy	7536.1 cGy	3046.0 cGy	824.4 cGy	3204.1 cGy
PTV	6084.1 cGy	7570.8 cGy	7342.1 cGy	7355.8 cGy	84.5 cGy
Bladder_I	123.9 cGy	7540.6 cGy	2733.9 cGy	850.7 cGy	2984.0 cGy
OUTER_1	0.0 cGy	7570.8 cGy	418.0 cGy	27.1 cGy	1085.7 cGy

Table 3.1
A table of function values that a physician shows for one treatment case [31].

Besides these simple functions, there is a category of functions that are widely used in clinics, called *dose volume metrics*. Dose volume metrics consist mainly of two functions D_x and V_d. Given a structure and a dose distribution on the structure, D_x is the minimum dose value of the hottest $x\%$ volume of the organ. It is the $x\%$ order statistic of the dose distribution. V_d is the inverse function of D_x, i.e., if $d = D_x$ then $x\% = V_d$. In other words, V_d is the fraction of volume that receives at least dose d.

Dose volume metrics are widely used for plan assessment. As in Table 3.2, $V_{47} \leq 53\%$ on rectum, $V_{47} \leq 53\%$ on bladder, and $D_{50} \leq 50$ on small bowel are required for this prostate case [31]. Although dose volume metrics are important for assessing treatment plans, they are very hard to incorporate into the nonlinear optimization problem as they have no analytic form. In practice, physicians run the optimization multiple times and

improve these dose volume metrics by changing certain optimization parameters based on their training and experience.

Structures	Metrics	Maximum Values
Rectal Wall	V_{47Gy}	53% of total volume
Rectal Wall	Max Point Dose	103.5%
Bladder Wall	V_{47Gy}	53% of total volume
Bladder Wall	Max Point Dose	103.5%
Femoral Heads	Max Point Dose	68 Gy
Large Bowel	Max Point Dose	60 Gy
Small Bowel	D_5	50 Gy
Small Bowel	Max Point Dose	53 Gy

Table 3.2
Normal tissue criteria for one type of prostate case [31].

The plot of the function $x\% = V_d$ with d as the variable is called a *dose volume histogram (DVH)*. Each DVH then represents one dose distribution without ordering voxels. Fig. 3.1 shows a typical DVH. Each line represents a dose distribution of a structure, which starts from the top left point ($V_0 = 100\%$), and then gradually decreases to 0 ($V_{\max} = 0\%$). It is then the goal of IMRT optimization to push target line(s) to top right (uniform dose) and lines of normal structures to the lower left (low dose to protect normal structures).

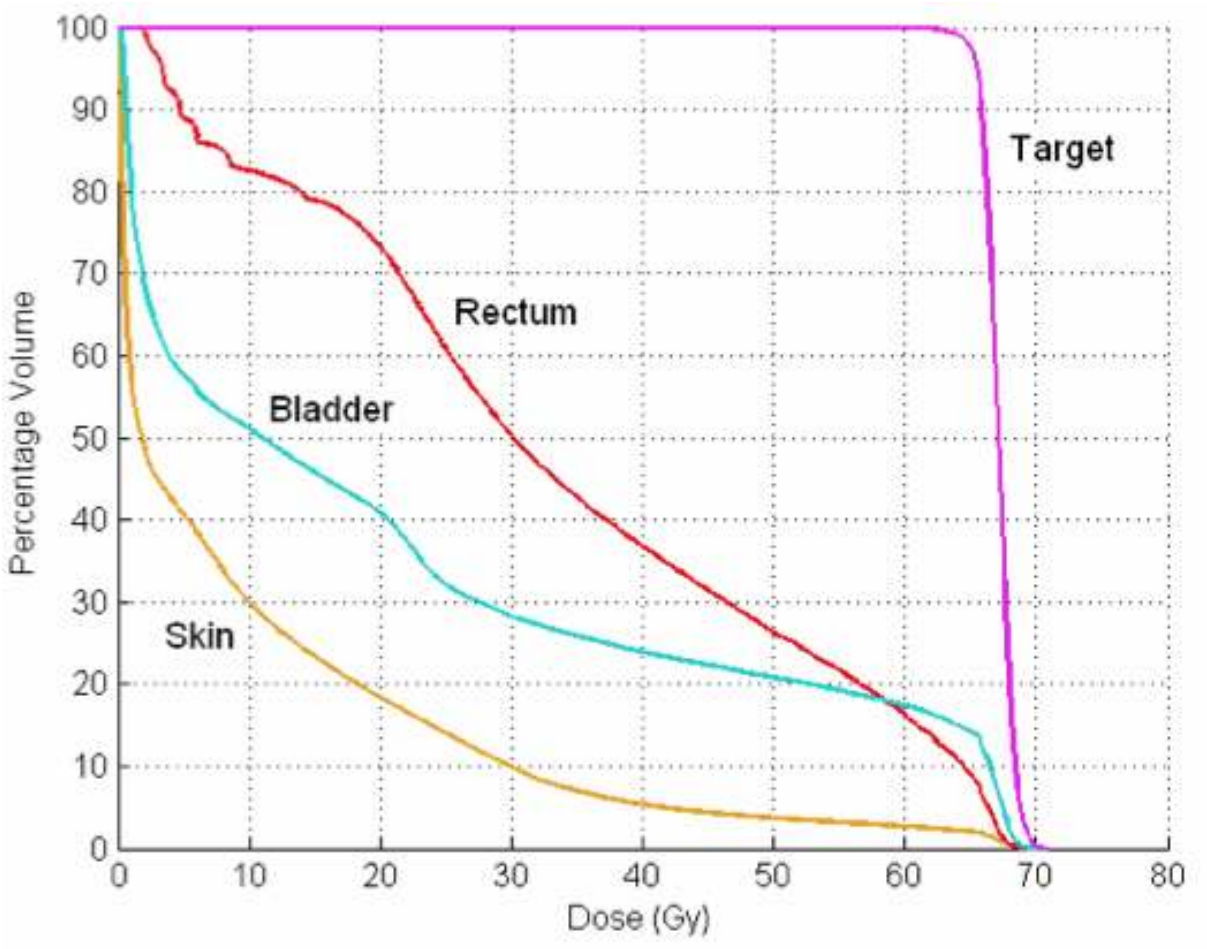

Figure 3.1. The dose volume histogram (DVH) of a prostate case.

3.1.2 Pre-emptive Goal Programming

As IMRT was introduced, the standard method of IMRT treatment planning involved minimizing a function that is the sum of terms representing constraints on normal structures or the tumor [32]. Although alternatives were proposed that involved estimating the probability of treatment outcomes, it was widely recognized that the data needed for building such models of normal tissue complication probability (NTCP) or tumor control probability (TCP) was not available [33]. For this reason, simple dosimetric surrogates are commonly used to drive radiotherapy treatment planning. For example, the surrogate

for tumor target volume dose quality was often just a quadratic sum of the target voxel doses. Doses to the surrounding tissues could be controlled by parameterizing the dose-volume histogram [34], but it was realized that this led to mathematically complicated formulations ('non-convex formulations' [35, 36]). When different tissues are taken into account, this is typically done by variably weighting terms in a linear sum [32]. Other parameters that are required from the user include estimates of DVH goals, such as 'lung volume above 20 Gy less than 20%, [37]' that are typically incorporated as weighted quadratic penalty terms into the objective function if the goal is not met.

Soon after the introduction of IMRT, researchers recognized that the quality of the optimization plans was very sensitive to the required weighting coefficients [38]. Thus, the standard IMRT planning method, which seemingly held the promise of simplified 'inverse planning', is commonly agreed to require even more trial and error user manipulation and planner time than 3-D conformal radiotherapy planning [32]. The typical user experience is that commercially available linear-formulation planning is a massively time-consuming process where the planner never really knows when to quit, and the physician never really knows for sure that more work will not result in a better plan. The current situation with standard planning techniques is summarized well by Peter Voet (Erasmus University) in the introduction to his recent PhD thesis on the topic of goal programming for IMRT (further discussed below):"With this procedure there is no guarantee that, even after many hours of planning workload, an alternative approach in steering the treatment planning system (TPS) software would not have resulted in a significantly improved treatment plan." On top of this, well-known (substantial) variations in the skills, experience, and drive of planners for generation of high quality treatment plans and the quality of the interaction of the treating clinicians with the planners may seriously

impact plan quality for individual patients. Another frequent threat for generation of high quality plans in current clinical practice is lack of planning time [39].

Furthermore, as the science of outcomes modeling matures and predictive models become more accurate, it will become more critical to account for the clinical priority of different toxicity or disease control endpoints, as represented through dose-volume surrogates (e.g., normal tissue complication probability and tumor control probability modeling).

The standard linear formulation fails on the criteria stated by Fraass [40] as "could I at least have confidence that all the easy improvements have been made?" The pre-emptive goal programming method was proposed to address this problem at a jointly sponsored NCI/NSF workshop, in 2002, devoted to "Operations Research Applications in Radiotherapy (ORART)" [41, 42]. Fig. 3.2, from the online proceedings of that meeting, clearly distinguishes the basis of the current proposal compared to the linear formulation.

As described in an article titled "The Physical Basis and Future of Radiotherapy," Bortfeld and Jeraj write that [43] "there is typically more than one normal organ involved and trade-offs between dosing the various normal organs have to be made as well. The standard way to find the most suitable trade-offs for an individual patient is by trial and error. A more scientific approach uses concepts of multi-objective optimization and in particular the concept of Pareto optimality." They reference Jee et al. [44] and Wilkins et al. [45], as well as other work on navigating the Pareto frontier that will be discussed below.

Subsequent to the ORART meeting, several groups developed and published full 3-D goal programming planning methods, all in 2007, including Washington University [45],

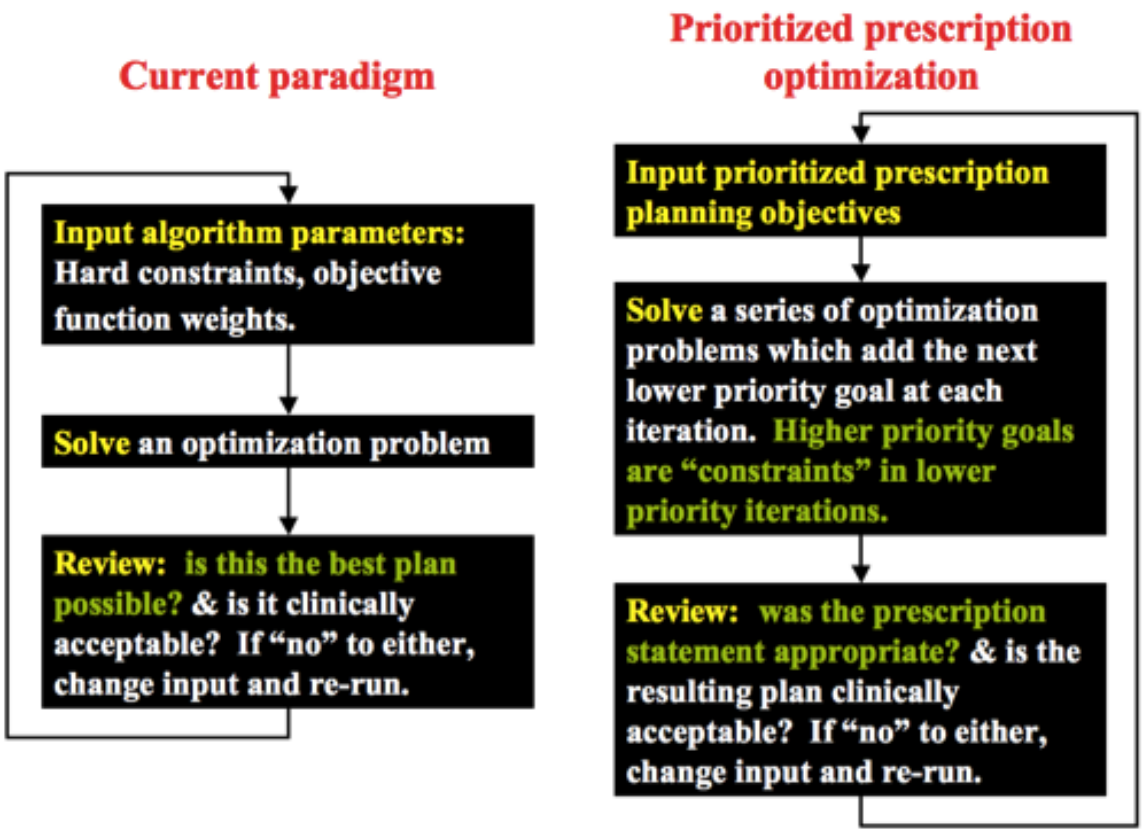

Figure 3.2. Basic description of the prioritized prescription optimization framework, from the online proceedings of the NCI/NSF sponsored ORART workshop [41].

University of Michigan [44], and Erasmus University [46]. These different flavors of goal programming applied to IMRT differ in some important details.

The Erasmus University approach, named 'iCycle,' has been the subject of several publications and improvements [46–49]. 'Goal levels' at each priority level are commonly included. If the optimizer reaches a goal level, then it stops trying to improve on that plan characteristic. However, this can leave metrics short of their potential levels (possibly producing Pareto inefficient solutions). Hence, in a second cycle, the algorithm seeks further improvements in dosimetric goals, past goal levels, in priority order. iCycle incorporates beam selection as well as fluence optimization. A recent publication

showed that, for a given patient cohort, iCycle produced plans in 31 of 32 cases that were preferable by clinicians, compared to those generated with the clinical commercial system (Monaco) [48]. A separate study showed that superior automated treatment plans could be generated for prostate VMAT delivery as well [49].

Actual deliverable plans are generated with iCycle in a two-step approach. iCycle generates an optimal plan based on the priority prescription list. The list of objective DVH parameter results are then used as inputs into the Monaco treatment planning system (inserted by hand or automatically with template scripting), which produces the final, deliverable machine instructions. In all cases, they found that this two-step method did not significantly degrade dose solution quality. The Erasmus team continues to develop iCycle, which was the subject of an ESTRO meeting session in 2014. iCycle is currently used clinically at Erasmus University for head and neck, prostate, cervical, and liver SBRT patients [39].

In contrast, Washington University, Memorial-Sloan Kettering Cancer Center and the University of Michigan have pursued a pre-emptive goal programming approach (also sometimes called lexicographic programming). The general advantage of pre-emptive goal programming is that "It can be the ideal choice when such a priority scheme exists in the mind of the decision maker. It can also be used to avoid unwanted direct comparisons between sensitive criteria [50]." Priority schemes are often implicit, but can be made explicit in radiotherapy, as shown for prostate [51].

The University of Michigan team implemented a pre-emptive goal programming (lexicographic programming) method using large-scale constrained nonlinear programming [44]. Their implementation provides fluence maps that are then converted into deliverable leaf sequences. This was integrated into their in-house planning system, providing a

clinically-usable tool. In follow-up publications, they have shown that they can integrate beam smoothing directly into the algorithm to produce plans with fewer delivery segments without degrading dose characteristics [52]. The incorporation of beam smoothing was also demonstrated by Wilkens et al. [45]. They also showed in detail how a prioritized solution for pancreatic cancer patients, based on using gEUD (generalized equivalent uniform dose) functions as objectives, was an excellent approach for a pancreatic radiotherapy dose escalation trial, consistently showing that significant dose escalation (up to 90 Gy) should be possible [53]. More recently, they also investigated the important problem of sensitivity analysis [54]. As they state: "If a minor sacrifice in high-priority criteria could yield meaningful benefits with respect to lower-priority criteria, the pure lexicographic optimization approach would not recognize or identify this opportunity." Although the method presented was interesting, due to the increased computational time and limitations of the algorithm (it only inter-compares single objectives from nearby priority levels; simplifications to a linear model were necessary to gain enough computational efficiency), there is motivation to approach this issue in a different way, further discussed below, by closely characterizing the solution point.

3.2 Prioritized Prescription Optimization

Given the success of 'iCycle' [48, 49], the similar approach of pre-emptive goal programming is adopted due to its advantages [50]. This section presents the pre-emptive goal proposed by Wilkens et al. and Clark et al. [45, 51], which is also named prioritized prescription optimization (PPO).

As in pre-emptive goal programming, PPO makes use of priorities of a physician's prescription in optimization. Usually targets have the highest priority. Dosimetric (or

radiobiological) indices that are more important regarding the severity or impact of the side-effect are given a higher priority. After determining the objective metrics for different structures and corresponding constraints, a 4-step optimization model is formed to reflect prescribed priorities.

Step I

A sum of objective metrics for all target volumes, denoted by F^I, is minimized. The vector w represents beamlet intensities and is called the *beam profile*; T is the set of target volumes; R^I is the set of sensitive structures that have the highest priority (rectum for the prostate case, for example); D_i^{pre} is the prescribed dose for the i'th target; V_i is the set of all voxels for the i'th target; $|V_i|$ represents the number of voxels in target i; $D_j(w)$ is the dose received by voxel j given w; and N is the number of beamlets. D_i^{max} is the product of *beam weight upper bound ratio* and the dose that, if given uniformly to each beamlet, would ensure that every voxel of the target volume gets at least D_i^{pre} dose.

Step I minimizes the deviation between dose on targets and prescribed dose. The variables t_i's are added to ensure maximize minimum dose values for each target. A voxel based upper bound is set for each sensitive structure involved in this step. The solution of step I is denoted by w^I and similar notations are used for later steps.

The optimization formulation is:

$$\text{minimize } F^I(w) = \sum_{i \in T} G_i(w) + t_i^2, \tag{3.4}$$

$$\text{where } G_i(w) = |V_i|^{-1} \sum_{j \in V_i} [D_j(w) - D_i^{pre}]^2, \qquad \forall i \in T, \tag{3.5}$$

$$\text{s.t. } D_i^{pre} - D_j(w) \leq t_i, \qquad \forall i \in T, j \in V_i, \tag{3.6}$$

$$0.05 D_i^{pre} \leq t_i, \qquad \forall i \in T, \tag{3.7}$$

$$D_j(w) \leq D_i^{max}, \qquad \forall i \in R^I, j \in V_i, \tag{3.8}$$

$$0 \leq w_k \leq w^{max}, \qquad \forall k \in \{1, \ldots, N\}. \tag{3.9}$$

Equation 3.6 requires dose deficit on target versus prescription dose to be less than a parameter t_i. This deficit is minimized (maximizing the minimum dose) by setting t_i in Equation 3.4, but not minimized too hard by giving t_i a lower bound as in Equation 3.7.

Equation 3.8 sets dose upper bounds for all structures in the set R^I, which consists of the most critical structure and other structures that have hard constraints on maximum dose.

Step II

In the second step, the sum of means of the hottest $1-\alpha$ percent (defined as MOHα) on each sensitive structure in a set R^{II} is minimized. For each structure $i \in R^{II}$, multiple α's can be set and they form a set A_i. The variables p_{ji}, y_i, z_{ji} are intermediate variables. Together with related constraints, they help to minimize the MOHα's. In particular, if $y_i^{\alpha_i} = D_{1-\alpha_i}$ where D_x is the minimum dose value of the hottest x percent, and if $p_{ji}^{\alpha_i}$ is replaced by $max(0, D_j(w) - y_i^{\alpha_i})$ to cancel these intermediate variables, then the objective function is exactly MOHα. The variable s is called *quadratic PTV dose slip*, which is used to relax the constraint and enlarge the optimization search space. The objective metric for Step I is treated as a constraint with this relaxation, which is also the basic

idea of pre-emptive goal programming. One more constraint is set to give a voxel based bound both from below and from above for the targets. They are denoted and defined as $D_i^{min} = \min_{j \in V_i} D_j(w^I)$ and $D_i^{max} = \max_{j \in V_i} D_j(w^I)$.

The optimization formulation for Step II is:

$$\text{minimize } F^{II}(w) = \sum_{i \in R^{II}} \sum_{\alpha_i \in A_i} H(w, \alpha_i), \tag{3.10}$$

$$\text{where } H(w, \alpha_i) = y_i^{\alpha_i} + \frac{1}{(1-\alpha_i)|V_i|} \sum_{j \in V_i} p_{ji}^{\alpha_i}, \tag{3.11}$$

$$\text{s.t. } D_j(w) - z_{ji} = 0, \qquad \forall i \in R^{II}, j \in V_i, \tag{3.12}$$

$$p_{ji}^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.13}$$

$$p_{ji}^{\alpha_i} - z_{ji} + y_i^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.14}$$

$$D_j(w) \leq D_i^{max}, \qquad \forall i \in R^I, j \in V_i, \tag{3.15}$$

$$G_i(w) \leq (1+s) G_i(w^I), \qquad \forall i \in T, \tag{3.16}$$

$$D_i^{min} \leq D_j(w) \leq D_i^{max}, \qquad \forall i \in T, j \in V_i, \tag{3.17}$$

$$0 \leq w_k \leq w^{max}, \qquad \forall k \in \{1, \ldots, N\}. \tag{3.18}$$

Equation 3.12, 3.13, 3.14, and 3.10 together achieve the MOHα minimization. All these new added constraints are linear, resulting in a preferable convex optimization. However, this step has greatly increased the number of variables and constraints.

Equation 3.16 uses the slip value to preserve the target dose coverage while minimizing F^{II}.

Step III

In the third step, the sum of mean doses, denoted by $\langle D(w) \rangle_i$ for structure i, for all normal structures in a set R^{III} is minimized. Objective metric for Step I is further

relaxed by a slip value $(1+s)^2$, while the objective metric for step II is bounded above by its value given in the Step II solution.

The optimization formulation for Step III is:

$$\text{minimize } F^{III}(w) = \sum_{i \in R^{III}} \langle D(w) \rangle_i, \tag{3.19}$$

$$\text{where } \langle D(w) \rangle_i = \frac{1}{|V_i|} \sum_{j \in V_i} D_j(w), \qquad \forall i \in R^{III}, \tag{3.20}$$

$$\text{s.t. } D_j(w) \leq D_i^{max}, \qquad \forall i \in R^I, j \in V_i, \tag{3.21}$$

$$G_i(w) \leq (1+s)^2 G_i(w^I), \qquad \forall i \in T, \tag{3.22}$$

$$D_i^{min} \leq D_j(w) \leq D_i^{max}, \qquad \forall i \in T, j \in V_i, \tag{3.23}$$

$$y_i^{\alpha_i} + \frac{1}{(1-\alpha_i)|V_i|} \sum_{j \in V_i} p_{ji}^{\alpha_i} \leq H(w^{II}, \alpha_i), \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.24}$$

$$D_j(w) - z_{ji} = 0, \qquad \forall i \in R^{II}, j \in V_i, \tag{3.25}$$

$$p_{ji}^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.26}$$

$$p_{ji}^{\alpha_i} - z_{ji} + y_i^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.27}$$

$$0 \leq w_k \leq w^{max}, \qquad \forall k \in \{1, \ldots, N\}. \tag{3.28}$$

The square term in Equation 3.22 further increases the slip value for the target and is empirically chosen. This part has to be improved so dose distribution on the tumor is more definitely controlled.

Equation 3.24 requires that Step III keeps the same dose quality on structures in R^{II}. One could also introduce another slip value for these high priority structures but this model does not.

Step IV

In the fourth step the square sum of beamlet intensity is minimized. This is helpful in both smoothing beam profile and avoiding hot spots outside structures. The objective

metric for Step I is further relaxed with slip value $(1+s)^3$. Another slip value s_2 called *minimum PTV dose slip* is adopted to lower the minimum dose values on targets. The mean values on all sensitive structures are bounded by their values given in the Step III solution.

The optimization formulation for step IV is:

$$\text{minimize } F^{IV}(w) = \sum_{k=1}^{N} w_k^2 \tag{3.29}$$

$$\text{s.t. } D_j(w) \leq D_i^{max}, \qquad \forall i \in R^I, j \in V_i, \tag{3.30}$$

$$G_i(w) \leq (1+s)^3 G_i(w^I), \qquad \forall i \in T, \tag{3.31}$$

$$D_i^{min}(1-s_2) \leq D_j(w) \leq D_i^{max}, \qquad \forall i \in T, j \in V_i, \tag{3.32}$$

$$\langle D(w) \rangle_i \leq \langle D(w^{III} \rangle_i, \qquad \forall i \in R^{III}, \tag{3.33}$$

$$y_i^{\alpha_i} + \frac{1}{(1-\alpha_i)|V_i|} \sum_{j \in V_i} p_{ji}^{\alpha_i} \leq H_i(w^{II}), \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.34}$$

$$D_j(w) - z_{ji} = 0, \qquad \forall i \in R^{II}, j \in V_i, \tag{3.35}$$

$$p_{ji}^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.36}$$

$$p_{ji}^{\alpha_i} - z_{ji} + y_i^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.37}$$

$$0 \leq w_k \leq w^{max}, \qquad \forall k \in \{1, \ldots, N\}. \tag{3.38}$$

Equation 3.29 decreases the beam profile variance as it penalizes high beamlet intensity by its square term. For this reason, the beam profile becomes more smooth, as does the dose distribution of the patient body.

Dose distribution on target is further sacrificed by the third power $(1+s)^3$ in Equation 3.31. Again, this is empirically chosen and needs to be more definite.

In sum, at each step an objective function of certain priority is optimized while searching around the solution of the previous step (if any) satisfying some empirically chosen

constraints. Table 3.3 is a summary of structures and objective functions that are included in the PPO model for a prostate case.

	Structures	Objectives
Step I	Targets	Deviation from prescription
Step II	Rectum	MOH_α
Step III	Bladder, Left/Right Femur	Sum of mean dose
Step IV	All structures	Square sum of beam intensities

Table 3.3
Structures and objectives that are used for each step of the prioritized prescription optimization model.

One important remark about prioritized prescription optimization is that the optimization search space is in the domain rather than the range of the objective function. Say f_1 has been optimized with optimal solution x_1, a slip s_1 has been chosen for f_1 and we want to optimize f_2. Then the search space is $\{x : f_1(x) \leq (1 + s_1)f_1(x_1)\}$. This is a convex set containing x_1 but could have irregular shape based on f_1, given that f_1 is a convex function. In prioritized prescription optimization, this means that after optimizing the target, the next step has plenty of room to improve for structures in R^I, and so on.

3.3 Robust Optimization for Treatment Planning

3.3.1 Introduction

Radiotherapy treatment planning for intensity modulated radiotherapy is challenging due to the massive problem size as well as the complexity of the clinical problem, which

is inherently about trade-offs between different dose-volume variables that are known to impact the probability of disease control or morbidity. Current computer approaches are typically time consuming, difficult to drive towards clinical objectives, do not produce optimal results, and do not explicitly account for the expected treatment geometry variability. Importantly, it is difficult or impossible to enforce priorities between dose-volume objectives related to very different endpoints. Trying to balance dose-volume characteristics with current methods is typically time-consuming with highly variable results. These limitations in treatment planning algorithms are also a roadblock to the introduction of adaptive radiotherapy approaches, where the automated, reliable production of high-quality treatment plans is needed to match changes in patient anatomy or the disease response.

Also of high significance is the need to account for underlying geometrical uncertainties, as further discussed below. This should be done in at least two ways: (1) the plan should be such that expected patient motions are unlikely to ruin the underlying dose distribution actually delivered (robust planning), and (2) the plan should have a good average performance across expected geometrical variations (probabilistic planning). These two technology goals represent a great deal of recent research effort. The problem is further complicated by the fact that planning results need to account for geometrical uncertainty, and should retain their high quality under residual motion uncertainties [55–59]. Thus, innovations in radiotherapy course management depend critically on the ability to automatically generate high-quality treatment plans at multiple points during a treatment course [60–62].

All radiotherapy delivery methods have some residual geometric uncertainties due to differences between the simulated image or images (e.g., 4-D breathing computed

tomography scans) used to do treatment planning. If the variations are small shifts in setup between many fractions, then the impact can be reflected primarily as a dosimetric blurring [63].

However, of much more concern are the cases when treatment shifts result in an average treatment position that is significantly different than the planned position. This can be due to a systematic shift between simulation imaging and all treatments, or because that particular patient experienced 'bad' random shifts enough to negatively impact the delivered dose distribution.

It is therefore a key goal of current treatment planning research to produce plans that are robust with respect to non-rare geometry changes. A useful definition of robustness has been introduced by Alber [64], whom we paraphrase: A robust plan is one for which key dosimetric indices of the actual dose delivered are statistically very likely to be within an acceptable range.

A class of geometric changes we do not consider in detail here are those cases where the patient geometry is likely to show a non-random 'drift' over the course of treatment, due, for example, to tumor volume regression, weight loss, or the like. Nonetheless, the method described is general, and could potentially address this class of changes as well.

The current standard way to address geometric uncertainty during delivery is to create a margin around the imaged clinical target volume (CTV), resulting in a planning target volume (PTV). For organs at risk, the relevant motion margin has been called a planning risk volume (PRV) [37, 65]. Doses on PTVs and PRVs are then optimized. The PTV approach has been criticized elsewhere as being a very poor trade-off between disease control and toxicity: in effect, it over-ensures adequate target volume irradiation and ignores the ever-present problem of normal tissue toxicity. A much more attractive

approach is to balance a very small degradation in target volume dosimetric indices for a significant reduction in the risk of toxicity, as represented by normal tissue dosimetric indices [66].

Moreover, when an inhomogeneous dose is planned or prescribed, as is common, for example, in hypofractionated, stereotactic radiotherapy [67, 68], then the PTV enlargement no longer applies. We also note that the dosimetric indices that have been correlated to outcome have, in almost all cases, not included a PRV margin. So the QUANTEC guidelines [66], for example, cannot be applied in the case of PRV planning [69].

We address robustness within the framework of prioritized prescription optimization (also called lexicographic optimization) [70]. The advantages of prioritized optimization have been published in detailed studies previously. A modification of the method has been developed by the Erasmus group [71]. Generally, the method uses sequential large-scale constrained optimization methods to generate high-quality plans that are Pareto optimal while also reflecting clinical priorities in the form of a prioritized list of planning indices to be minimized. These prioritized goals are optimized in turn. After each step, the optimal value is turned into a constraint. A minor but important detail is that some 'slip' or 'relaxation' is added to the problem after each iteration, to increase the feasibility space for lower priority goals. The Erasmus approach also includes a goal for each index: if the goal is reached then optimization is terminated. In addition to the advantages of automation, Pareto optimality, and clinical relevance, the prioritized prescription framework is attractive as a platform for robust planning because it allows the planner to add goals that are optimized without destroying desirable qualities of the previously generated plan: those indices, as computed on the simulation image(s), are included as constraints.

Several approaches to incorporate motion uncertainties into the IMRT optimization planning process have been proposed, that can be broadly classified into three categories:

The first category ('shift robustness') assumes that the position of any CTV or organ at risk (OR) forms a discrete distribution. Each position where the CTV/OR is found is a possible position with certain probability (i.e., a coverage probability). Then the objective metric is averaged over this discrete probability space to form the expectation value of the metric used for planning. Most approaches belong to this category [59, 69, 72–75]. Due to the small number of 'real' positions (usually, shifted CT scans) that are known prior to treatment, the expected metric itself has a significant uncertainty. A technique to extend this by interpolating between patient specific data and population-based knowledge on organ motion by Bayesian inference was proposed [76]. Besides the expected value of the metric, the variance of the metric should also be considered [76, 77].

The second category ('voxel-based robustness') also assumes that the dose value received by each voxel is an independent random variable due to motion uncertainties. The dose distribution on each voxel is controlled within the optimization process, and thus the dose distributions on different structures are controlled [58, 76–78]. A serious limitation of methods in this category is the huge increase in problem size and hence in time (hours) for it is voxel based, although Olafsson et al. proposed a sequential linear programming algorithm to decrease computation time. [57]. A second limitation is that voxel doses are highly correlated, rather than uncorrelated. Both the first and second class of algorithms requires knowledge of motion probability distributions prior to treatment.

The third category ('metric-based robustness') was proposed by Sobotta et al. in 2010 [64], with the assumption that, in the sense of optimization, the robustness of a treatment plan against motion uncertainties is characterized by the robustness of indi-

vidual objective metrics (as we noted above). They showed that controlling the shape of metric distribution outperforms controlling distributions based on voxel values, because this approach allows for an intervoxel dose trade-off whereas the voxel-based robustness approach does not.

Our solution builds on and extends the third category. We propose a new planning algorithm methodology, called ORATOR, short for Outcome-driven, Robust, Automated, Treatment plan Optimizer for intensity modulated Radiotherapy. This method adds a new robustness step to the prioritized prescription optimization. After using prioritized optimization to achieve a high quality treatment plan, a new step is added to minimize a function of the mean and variance of objective metrics by allowing relaxation to the treatment plan for static structures. Besides using different data and models, including more complex objectives and constraints, our work is different from Sobotta et al.'s work in several fundamental ways:

1) It is embedded as a last step in a prioritized prescription/lexicographic framework, ensuring minimal impact on other important plan characteristics.

2) Although their model works for any metric, and the resulting Chebyshev inequality was used to give an error bound, and the metric becomes normally behaved when the number of fractions increases, their optimization model still assumes that the metric has a normal distribution. It seems likely that this assumption is commonly violated because, for example, anterior-posterior shifts may make much more of an impact on rectal toxicity than shifts in the opposite direction. This assymmetry is exacerbated because the operating point of normal tissue complication probabilities is usually in the low, relatively quadratic, part of the corresponding response curve. We make no assumption on metric distribution but only use two computed values: the mean and variance of the metric

averaged over the simulated treatment courses. A limitation of our method is that we assume uncertainty distribution parameters.

3) Sobotta et al. maximize the probability of a metric in a fixed interval, which essentially controls the mean value of the metric within a range, without a preference for the direction of improvement. In contrast, we directly try to improve metric mean values during our robustness step.

4) Also, we include the metric variance explicitly.

3.3.2 Methods and Materials

Robustness in Optimization

In the sense of optimization, robustness means robustness of objective metrics [64] for they are the functions that are being optimized.

For each structure, one objective metric is usually used in IMRT optimization. It can be either a physical or a biological metric. When motion uncertainties are involved, this metric becomes a random variable with a probability distribution with respect to the motion uncertainties. When multiple metrics are needed, they form a joint probability distribution.

We define the *goal value* for an objective metric to be the desired metric value that the perfect treatment plan can achieve or a physician is satisfied with. For example, if the metric for structure Brain is the mean dose, then the goal value can be 0 Gy or any dose satisfying a physician. The lower the dose, the fewer possible complications happen to that structure. Goal values are usually specified (e.g. mean, D_x, MOHx, etc), but if not, they can be set as the metric value from the solution of the prior step of optimization.

There are two parts to an objective metric being robust. First, robustness means stable outcomes. So the variance of the metric distribution under motion uncertainty needs be small. Second, the metric distribution needs to be close to the goal value. Controlling only the variance cannot guarantee robustness around a desired metric value. Robust solutions should satisfy both conditions.

Objective Metric for Robustness

Say Y is an objective metric for a structure that is used in an IMRT optimization and Y^{goal} is its goal value. Y becomes a random variable due to motion uncertainties. Then both the absolute different $|E(Y) - Y^{goal}|$ and the variance $Var(Y)$ need be small, where $E(Y)$ and $Var(Y)$ are the expected value and variance of Y, respectively.

If there are in total I structures, each of which has objective metric Y_i, then the objective function for the robustness optimization step is

$$\sum_{i=1}^{I} \lambda_i \left[(E(Y_i) - Y_i^{goal})^2 + Var(Y_i) \right], \tag{3.39}$$

where λ_i's are weights for different structures.

There are a few comments on Equation (3.39):

i) The first term has a square so as to have the same units as the second term.

ii) Equation (3.39) can be simplified easily as $\sum_{i=1}^{I} \lambda_i \left[E(Y_i^2) - 2 \cdot E(Y_i) \cdot Y_i^{goal} \right]$ after expansion and dropping an unnecessary constant.

iii) When implementing the above simplified objective metric, motion sampling is needed to estimate both $E(Y_i)$ and $E(Y_i^2)$, before the robustness optimization step. We choose unbiased estimates $\frac{1}{M}\sum_{m=1}^{M}(Y_i^m)$ and $\frac{1}{M}\sum_{m=1}^{M}(Y_i^m)^2$ for M sampled motions.

iv) Weights can be added to emphasize mean or variance.

Robustness slips

Our goal is to fine-tune the solution obtained by an existing prioritized model in the final robust step. There is a natural trade-off between preserving the high solution quality of the existing prioritized model, and pursuing robustness in the final robust step. When structures have relatively stable positions, we tend to be more conservative on preserving the existing solution. When structures have large position variations, we then want more robustness. We define a *robustness slip*, denoted by s_r, in the final robust step, to characterize this natural trade-off.

Specifically, robustness slip is used in inequalities related to Equations (3.29, 3.31, 3.33, 3.34), as follows:

$$G_i(w) \leq (1+s_r)(1+s)^4 G_i(w^I), \tag{3.40}$$

$$\langle D(w)\rangle_i \leq (1+s_r)\langle D(w^{III}\rangle_i, \tag{3.41}$$

$$y_i^{\alpha_i} + \frac{1}{(1-\alpha_i)|V_i|}\sum_{j\in V_i} p_{ji}^{\alpha_i} \leq (1+s_r)H_i(w^{II}), \tag{3.42}$$

$$F^{IV}(w) \leq (1+s_r)\cdot F^{IV}(w^{IV}). \tag{3.43}$$

Step I-IV optimize targets, sensitive structures with high priority, other sensitive structures, and beam profile smoothness respectively. $G_i(w^I)$, $H_i(w^{II})$, $\langle D(w^{III})\rangle$, and $F^{IV}(w^{IV})$ are the best objective function values for these 4 steps. A multiplier $(1+s_r)$ is added to characterize the deviation from these function values in the function range space. When s_r is close to 0, conservative solutions close to the existing solution are pursued, but when s_r is large, solutions more robust to motion uncertainties are pursued.

Step V (robustness step)

Define V_i^m as the voxel index set for the m'th motion of structure i, and define

$$G_i^m(w) = \frac{1}{|V_i^m|} \sum_{j \in V_i^m} [A_j^m w_j - D_i^{pre}]^2,$$

$$H^m(w, \alpha_i) = y_i^{\alpha_i} + \frac{1}{1-\alpha_i} \sum_{j \in V_i^m} p_{j_i}^{\alpha_i},$$

$$\langle D(w) \rangle_i^m = \frac{1}{|V_i^m|} \sum_{j \in V_i^m} D_j(w),$$

which are just intermediate variables for rigorous notation.

Then the optimization formulation for Step V is

$$\text{minimize } F^V(w) = \sum_{i=1}^{I} \lambda_i E(Y^2) - 2E(Y)Y_i^{goal}, \tag{3.44}$$

$$\text{where } E(Y^2) = \frac{1}{M}\sum_{m=1}^{M}(Y_i^m)^2, \tag{3.45}$$

$$\text{and } E(Y) = \frac{1}{M}\sum_{m=1}^{M} Y_i^m, \tag{3.46}$$

$$\text{s.t. } Y_i^m = G_i^m(w) \qquad \forall i \in T, m = 1, \ldots, M, \tag{3.47}$$

$$Y_i^m = \sum_{\alpha_i \in A_i} H^m(w, \alpha_i), \qquad \forall i \in R^{II}, m = 1, \ldots, M, \tag{3.48}$$

$$Y_i^m = \langle D(w) \rangle_i^m, \qquad \forall i \in R^{III}, m = 1, \ldots, M, \tag{3.49}$$

$$D_j(w) \leq D_i^{max}, \qquad \forall i \in R^I, j \in V_i, \tag{3.50}$$

$$G_i(w) \leq (1 + s_r)(1 + s)^3 G_i(w^I), \qquad \forall i \in T, \tag{3.51}$$

$$D_i^{min}(1 - s_2) \leq D_j(w) \leq D_i^{max}, \qquad \forall i \in T, j \in V_i, \tag{3.52}$$

$$H(w, \alpha_i) \leq (1 + s_r) H(w^{II}, \alpha_i), \qquad \forall i \in R^{II}, \alpha_i \in A_i, \tag{3.53}$$

$$D_j(w) - z_{ji} = 0, \qquad \forall i \in R^{II}, j \in V_i, \tag{3.54}$$

$$p_{ji}^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.55}$$

$$p_{ji}^{\alpha_i} - z_{ji} + y_i^{\alpha_i} \geq 0, \qquad \forall i \in R^{II}, \alpha_i \in A_i, j \in V_i, \tag{3.56}$$

$$\langle D(w) \rangle_i \leq (1 + s_r) \cdot \langle D(w^{III} \rangle_i, \qquad \forall i \in R^{III}, \tag{3.57}$$

$$F^{IV}(w) \leq (1 + s_r) \cdot F^{IV}(w^{IV}), \tag{3.58}$$

$$0 \leq w_k \leq w^{max}, \qquad \forall k = 1, \ldots, N. \tag{3.59}$$

Table 3.4 is a summary of the new 5-step model.

Motion Simulation

Motion simulation is needed to estimate Equation (3.39), so as to incorporate it into the optimization formulation.

	Structures	Objectives
Step I	Targets	Deviation from prescription
Step II	Rectum	$MOH\alpha$
Step III	Bladder, Left/Right Femur	Sum of mean dose
Step IV	All structures	Square sum of beam intensities
Step V	All stuctures	Expected squared deviation from ideal objectives plus variance

Table 3.4
Structures and objectives that are used for 5-step ORATOR model.

We make a few assumptions on the motion simulation. 1) Only one CT scan position is available which is the average positions of structures. When more CT scans are available, other assumptions can be changed accordingly. 2) The positions due to systematic and random errors are normally distributed. Σ is the standard deviation of systematic uncertainties, and σ_i's are the standard deviations of random uncertainties. 3) All structures move rigidly and move together (for one patient) when restricting to systematic uncertainties but independently when restricting to random uncertainties. 4) PTV is the 1-cm enlargement of CTV. PRV is the 1-cm enlargement of OR.

Fig. 3.3 illustrates the CTV, CTV motions and PTV given the above assumptions. Similar figure exists for OR, OR motions and PRV.

For beam profile optimization and dose distribution verification, different motion simulation methods can be used depending on the number of fractions. Due to the limit of computer memory size, the product of the number of systematic motions and random motions needs be less than 300-600 for a computer with 16 GB memory, depending on

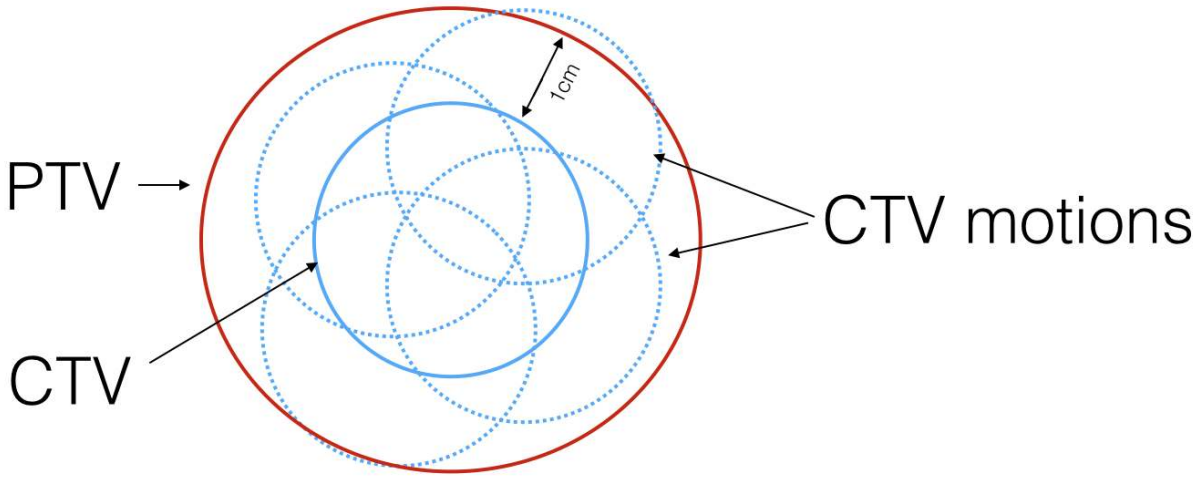

Figure 3.3. Illustration for motion simulation.

the tumor size and voxel size. When there is a large number of fractions (around 30), the convolution method [79] is used to smear the beamlet intensities and hence the influence matrix, by convolving the static dose calculation with a Gaussian distribution. Then more than 100 systematic motions may be sampled. On the other hand, when the number of fractions is relatively small (e.g. around 5) for come treatment cases, the same number of random motions are sampled. Fewer systematic motions (30 or 50) may be simulated so as to be computationally tractable. There are *number of systematic motions* $\times$ *number of random motions* motions in total.

As for the dose verification, we use 100 systematic motions each with 30 fractions. Then 30 dose metrics are computed for each systematic motion and are summed together as one dose metric. These 100 metrics are illustrated later in Fig. 3.4 and Fig. 3.5.

Experimental Details

Data and software:

The data set we have used consists of 10 anonymized prostate treatment cases. Each dataset has 256×256 voxels with size 0.2344×0.2344 cm or 0.1954×0.1954 cm and 0.3 cm thickness in the transverse plane.

Data preparation and motion simulation are realized by the software CERR [80]. The solver for our convex nonlinear optimization is Interior Point OPTimizer (IPOPT) [2] with HSL Mathematical Software Library [81], a powerful solver that is able to efficiently deal with general differentiable objective functions or constraints and can even solve non-convex problems very fast. Since we have a large size problem, MA86 of HSL was used as the linear solver and was compiled in a way that it can run in parallel.

The connection between MATLAB and IPOPT we have used is IPOPT's MATLAB interface, which requires that we write functions for objectives, gradients, Jacobians, Jacobian structures (logical matrix to represent the nonzero elements of Jacobians), Hessians, and Hessian structures (defined similar to Jacobian structures). These function evaluations are also implemented in a way that they run in parallel.

Parameter selection:

In Table 3.5, we list all parameters that were used for implementing the ORATOR model. All parameters for the first 4 steps are the same as the the original paper [51]. Notice that some of the variables have multiple values for the purpose of comparison.

Techniques for speed-up:

A serious problem along with robust IMRT optimization is the long running time. As the number of motions increases, the time that is needed increases rapidly for nonlinear optimization if using a poor implementation. Our initial implementation spent hours on

Mean of the hottest $\alpha\%$	84%
Beam weight upper bound radio	1.5
Quadratic PTV dose slip s	0.5
Minimum PTV dose slip s_2	1.5%
Robustness slip s_r	0-1.2
Number of motions	10, 20, 30, 60, 100, 300, 600
Number of fractions	1 (convolution), 10, 30, 60
Standard deviation for systematic errors Σ	3 mm
Standard deviation for random errors σ	3 mm

Table 3.5
Parameters for the implementation of ORATOR.

one run. We found a few useful techniques that greatly helped us speed up the optimization process. After utilizing these techniques, ORATOR takes less than 6 minutes. Detailed time comparison may be seen in Section 3.3.3.

1) Choosing a good linear parallel solver and implementing function evaluations in parallel.

When the optimization formulation has a large number of constraints, the choice of linear solver mostly determines the speed of IPOPT, for the most time-consuming step is the linear factorization. We choose MA86 of HSL as our linear solver for two reasons. First, it is designed for large problem size and our problem is nonlinear with 10^6 constraints. Second, it is designed for multi-core, which utilizes the power of parallel computation. Another time-consuming part for IPOPT is the function evaluation. We parallelized our

function evaluation computation. For comparison, running IPOPT+MA86 with parallel implementation for function evaluation is about *20 times* faster than using MUMPS, MA27, or other similar linear solvers.

2) Pre-store influence matrices.

Due to the large size of the influence matrix, it is slow to repeatedly read the influence matrix when computing dose distributions. This is especially slow if there are multiple motions as each motion requires a visit to the influence matrix. Instead, we pre-store influence matrices for all involved structures and their motions. This increases memory usage but makes calculations related to dose distribution much faster.

3) Solution of one step is set as the initial value of the next step.

Usually, for a nonlinear optimization problem, it is very hard for the solver to find a feasible initial value. In our experiments, it takes IPOPT a lot of iterations (around 40 for each step in our model) just to find a feasible solution. Sometimes, IPOPT even failed to find any feasible solution after 100 iterations.

Multi-step optimization has an advantage to avoid this trouble, that is, the solution of one step is a perfect initial value for the next step. This is because the search space of the next step is always a set around the solution of current step in the domain. For example, the quadratic PTV slip term makes the solution of Step I trivially satisfy the new added constraint in Step II.

A special case is Step I, which has no prior step. But it usually converges around 40 steps because of the simple formulation.

This technique saved us more than half the iterations and thus gave 2 times speed up. Furthermore, after using this technique, each step converges in less than 50 iterations.

4) Utilize the special properties of linear and quadratic functions to quickly evaluate Hessians.

IPOPT asks users to evaluate objective functions, gradients, Jacobians, and Hessians, among which computing Hessians could be the most time-consuming one when there are nonlinear and non-quadratic objective functions or constraints.

In Step I, the Hessian is simply a constant matrix. So we need only compute it once.

From Step II to Step IV, the objective functions are linear (Step II and III) or quadratic (Step IV) and one of the constraints is quadratic for each step. So the Hessian is a linear function of Lagrange multipliers with constant coefficient matrices. These constant coefficient matrices were computed only once and were stored and called at each iteration.

In Step V, if Y_i^m's are not set as individual variables, then the Hessian matrix is neither a constant nor a linear function of multipliers with constant coefficient matrices. However, after setting them as variables, the Hessian matrix of the formulation is a linear function of multipliers with constant coefficient matrices that we could pre-store and call at each iteration.

This technique gives about 1.1 times speed-up for Hessian computation for each step from Step I to Step IV, and *100 times* speed-up for Step V when 100×1 motion simulation is used.

Notice that this technique only works for linear and quadratic functions. If biological metrics such as EUD or TCP are chosen as metrics of structures, then the Hessian

matrices need to be re-calculated every time step and some other technique needs be used.

3.3.3 Results

The base model is the PPO model with CTV and ORs (no enlargement). We denote this base model simply by *CTV*. The proposed model adds a robustness step onto *CTV*. We denote this model by *Robust*. The PPO model with PTV and PRVs (enlargement of CTV and ORs) is another model we use for comparison, denoted by *PTV*.

Motion simulation is an important factor of the proposed model and may have an effect on the beam profile solution. We thus compare different motion simulations first. As in Fig. 3.4 and Fig. 3.5, solutions with different numbers of systematic motions and numbers of random motions are compared. As the number of systematic motions increases from 10 to 600 while fixing the product of it and the number of random motions as 600, final beam profiles are similar without significant difference. Here 600x1 means systematic motions + convolution for random errors. Using convolution, as the number of systematic motions decreases from 600 to 100, the beam profiles are also not significant different. Therefore, we claim that different motion simulations are not significantly different for this case. We will use 100x1, that is, 100 systematic motions + convolution, for its relatively small size from which fast speed results.

As one could tell, the structure weights λ_i's in Equation (3.39) have some impact, though modest, on resulting plans. A higher weight on one structure effectively increases the relative robustness of that structure, when there are conflicts with other structures.

Box-plot and Scan Comparison

We first assume that the robustness weight is 100 for CTV and 1 for other structures.

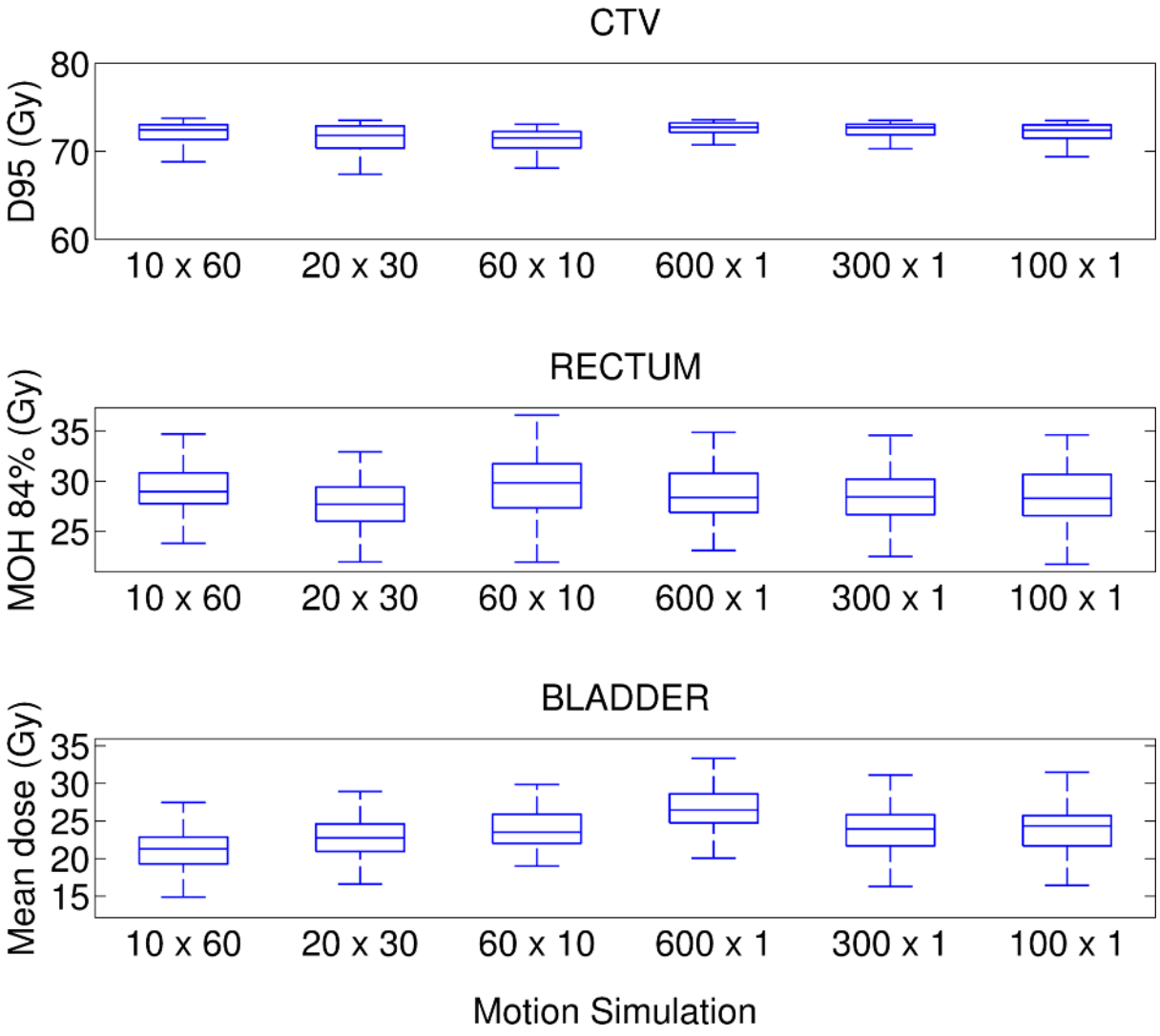

Figure 3.4. Comparison of different motion simulations evaluating with 30 fractions. The robustness slip is 0.2. In the x-axis, $m \times n$ means m systematic motions $\times$ n random motions. One exception is $n = 1$ means dose convolution for random errors. The solution generated given different motions simulations is evaluated with 100 systematic motions each with 30 fractions. The testing case is case 8.

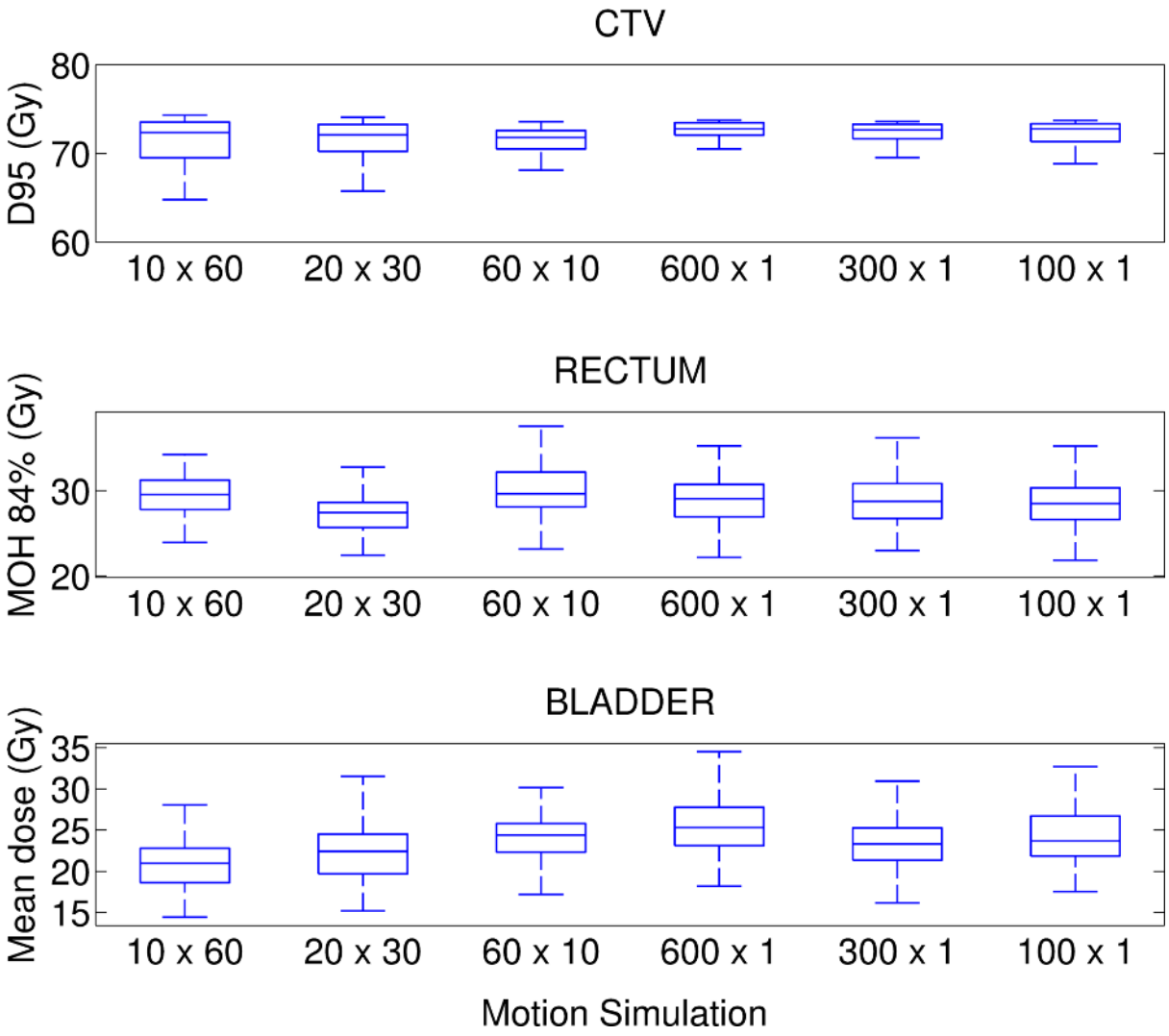

Figure 3.5. Comparison of different motion simulations evaluating with hypo-fractions. The robustness slip is 0.2. In the x-axis, $m \times n$ means m systematic motions $\times$ n random motions. The solutions generated given different motions simulations are evaluated with 100 systematic motions each with 5 fractions, different with 30 fractions for Fig. 3.4. The testing case is also case 8.

Robust step is designed to be a fine-tune step to keep the high solution quality of *CTV* model while improving its robustness against motion uncertainties in the added final fifth optimization step. As in Fig. 3.7, the high quality solution of CTV model becomes worse (2 Gy downgrade for D_{95} on target) when motion uncertainties are incorporated. After adding the fifth robustness step, the solution quality improves, depending on how large the robustness slip is. As the robustness slip becomes larger, the metric on target becomes larger and its variance becomes smaller. Meanwhile, the metrics on sensitive structures that have overlaps with the target (rectum and bladder for this case) become worse (see Fig. 3.8 for structure positions), as expected.

Fig. 3.6 tells us this robustness weight combination weights too much on CTV as the most sensitive structure rectum receives very high dose. Rectum is another very important organ which needs to have robust dose. In Fig. 3.7, we assume that the robustness weight is 100 for CTV, 50 for rectum and 1 for other structures. As the robustness slip increases, CTV metric is not significantly different with that in Fig. 3.6. Meanwhile, the rectum metric decreases significantly. As the robustness slip increases, MOH84 decreases.

Fig. 3.8 illustrates the exact dose distributions for one CT slice. The contoured structures are in their nominal positions. The center is CTV, with bladder and rectum above and below it. *Robust* model achieves higher dose around the target than *CTV* model. Also, *Robust* model yields less dose on bladder and rectum than *PVT*, whose figure shows a large area of high dose on bladder and rectum because PTV overlaps a lot with them.

DVH Comparison

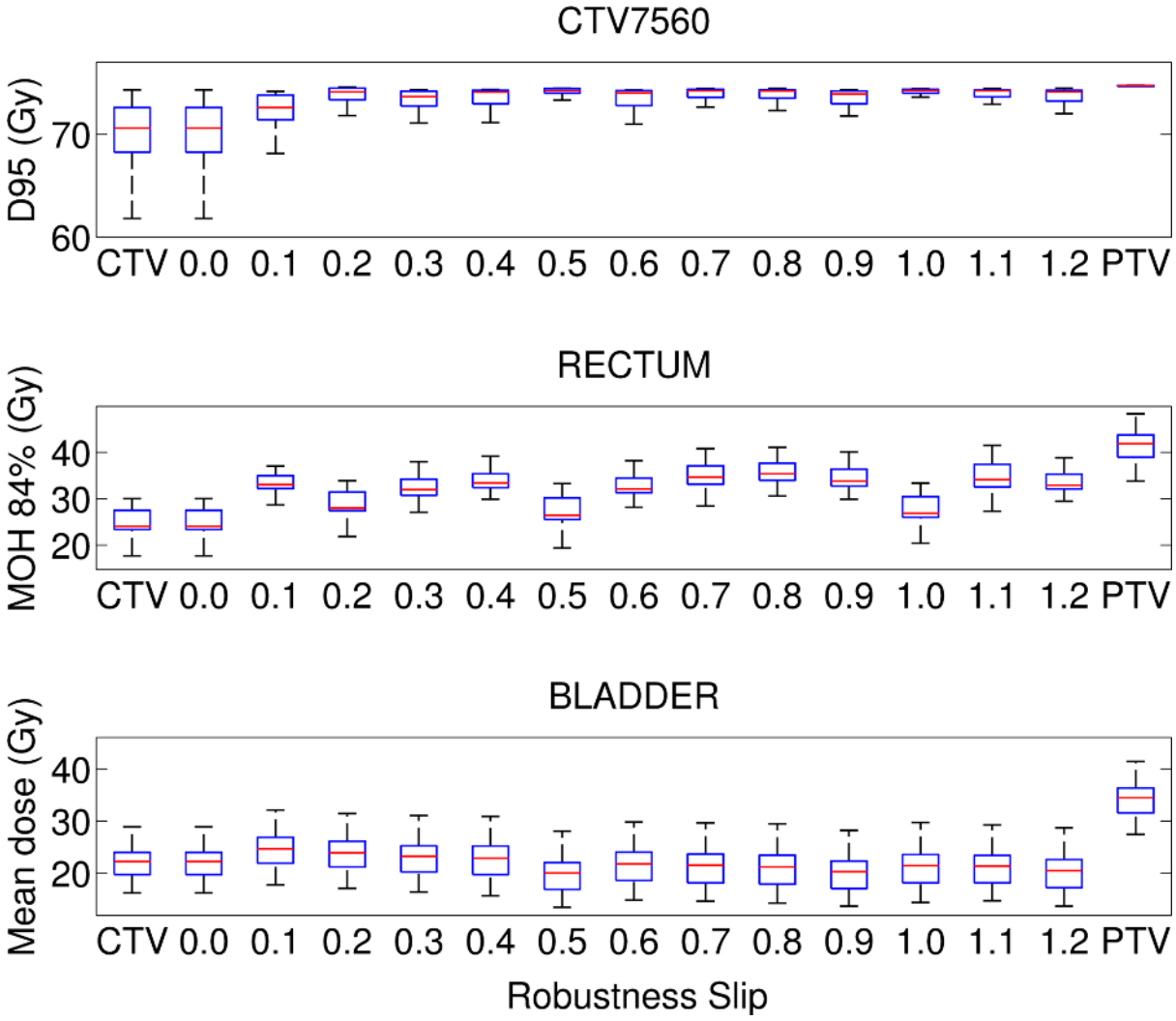

Figure 3.6. Box-plot metric comparison for *CTV*, *PTV*, and *Robust* with different robustness slip values. Every box-plot is plotted with metrics from 100 motion samples (systematic motion) each of which is an average of 30 fractions/random motions. The robustness weight is 100 for CTV, and 1 for else. The data set being used is case 8.

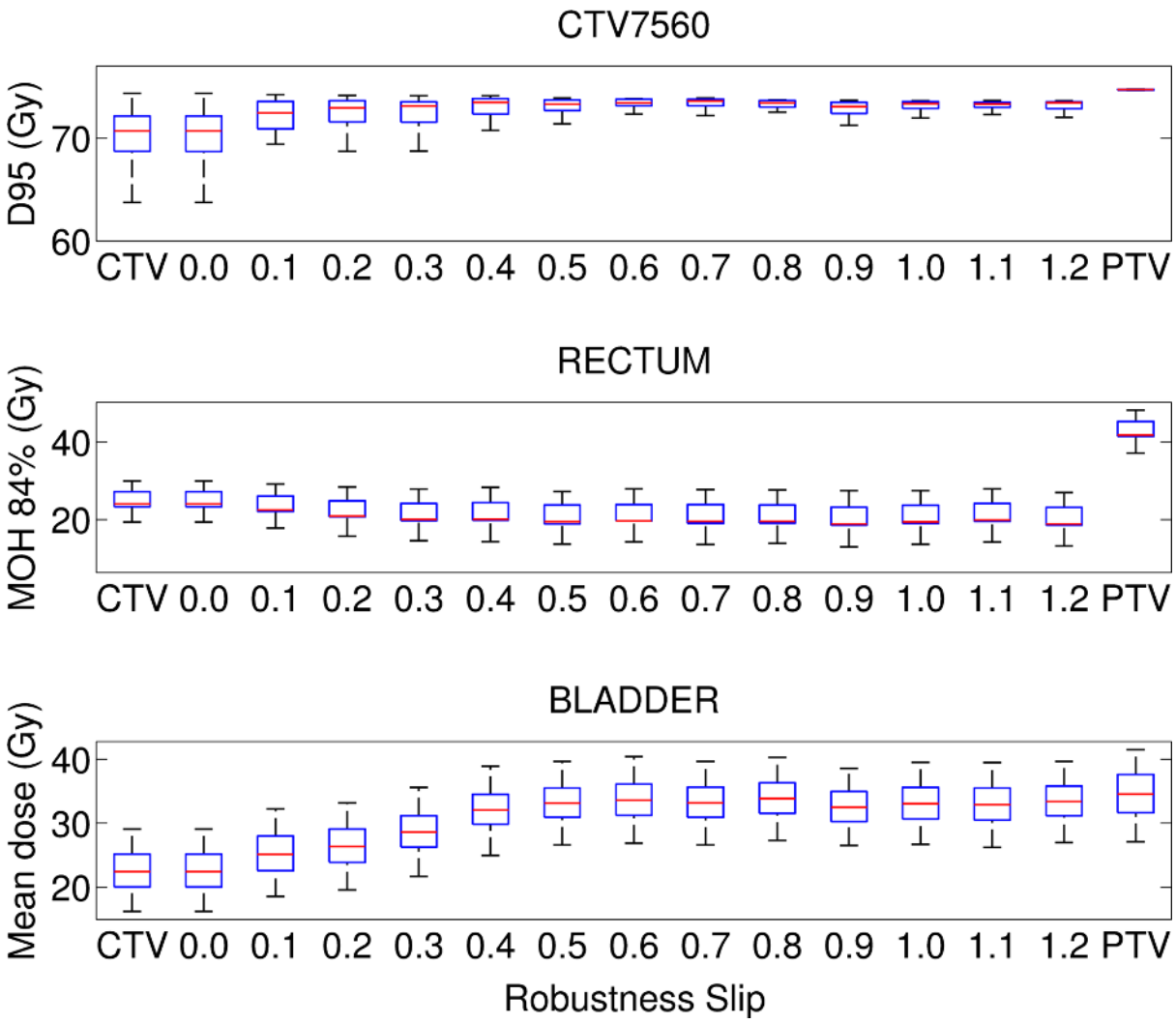

Figure 3.7. Box-plot metric comparison for *CTV*, *PTV*, and *Robust* with different robustness slip values. Every box-plot is plotted with metrics from 100 motion samples (systematic motion) each of which is an average of 30 fractions/random motions. The robustness weight is 100 for CTV, 50 for rectum and bladder, and 1 for else. The data set being used is case 8.

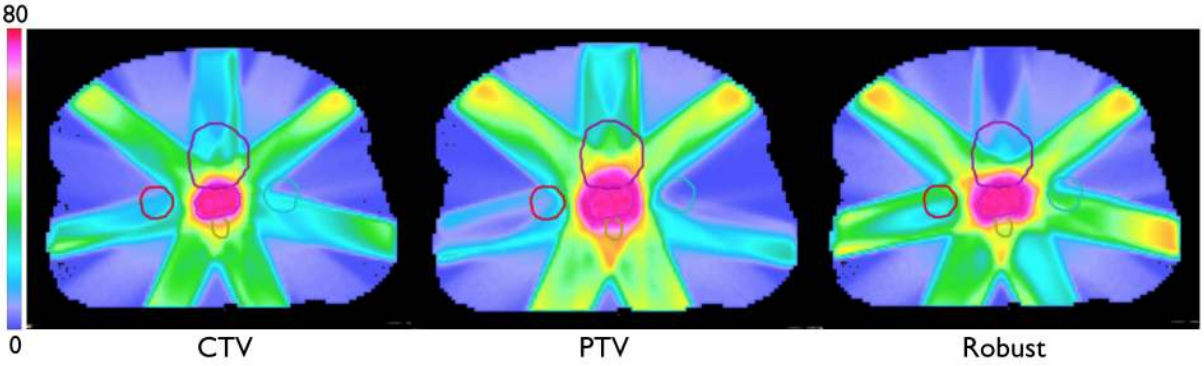

Figure 3.8. Dose scan comparison of one slice for *CTV*, *PTV* and *Robust* optimization for case 8 with robustness slip 1.

To get a whole picture of the dose distributions, we compare these models using DVHs. As in Fig. 3.9, compared to the *CTV* model, the *Robust* model generates higher and more uniform dose distribution on the target but only slightly increases the doses on sensitive structures. Compared to the *PTV* model, the *Robust* then generates very similar dose distribution on the target, but with much less dose on sensitive structures.

Results for 10 prostate cases

To check how this model behaves for different cases, we list the metric comparison of all 10 cases we tested in Fig. 3.10. Each metric is a random variable with respect to the motion probability space. Its mean value and standard deviation value are thus considered for comparison. Similar to the case above, *Robust* model achieves about 2% higher target dose and about 50% less variance than *CTV* model, without increasing the dose too much on sensitive structures. Although *PTV* has almost 0 variance as expected, it sacrifices a lot on sensitive structures (e.g.: above 10 Gy higher for MOH84 on rectum for some cases). Notice that the *PTV* model still has large variation for metrics of sensitive structures because these metrics do not require uniform dose distribution like

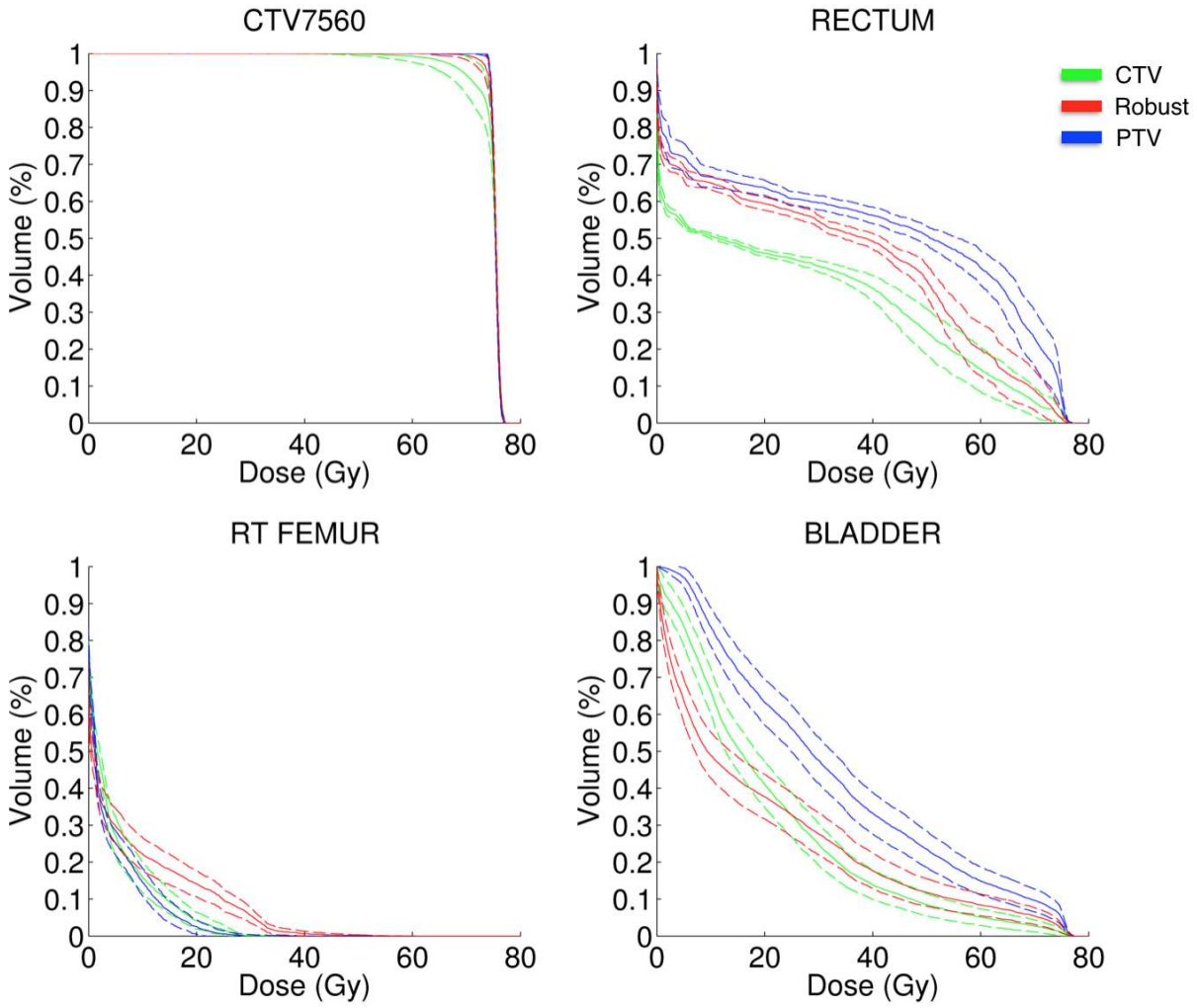

Figure 3.9. DVH comparison of *CTV*, *PTV* and *Robust* optimization with robustness slip as 1 for case 8. Solid lines are mean DVHs of 100 motion samples each with 30 fractions and dashed lines are +/- one standard deviations.

the target metric, and the dose distribution has variation not matter how large the PRV is.

Finally, we compare the time needed for each step. We mentioned that our initial implementation took hours for one run of ORATOR. After utilizing the speed up techniques introduced before, the whole optimization can be finished in less than 6 minutes, as in Fig. 3.11. The final robustness step takes about 3 minutes. In Clark et al.'s work

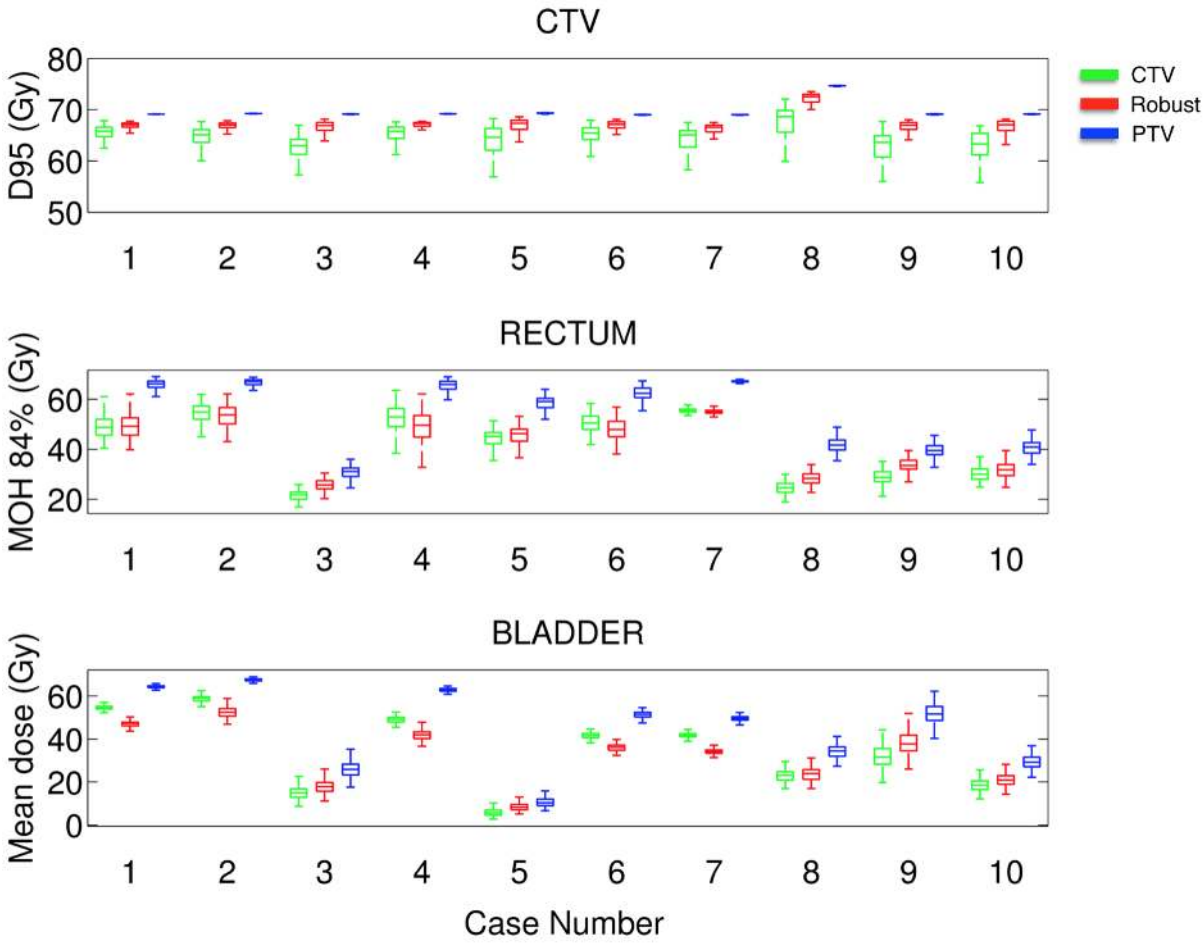

Figure 3.10. Metric comparison of CTV, Robust, and PTV optimization for 10 prostate cases. Robustness slip is $s_r = 0.2$. 100 systematic motions are used, as well as convolution method for random errors. Dose verification uses 100 systematic motions each with 30 fractions.

[51], each step takes 5 minutes. As a comparison, we have improved the speed by about 8 times.

3.3.4 Discussion

Robustness to motion is created by the introduction of a final step in the pre-emptive optimization list, using the objective function shown in Equation 3.39. This function is averaged over a set of shifts. It pushes objective values towards their ideal goals (e.g., zero

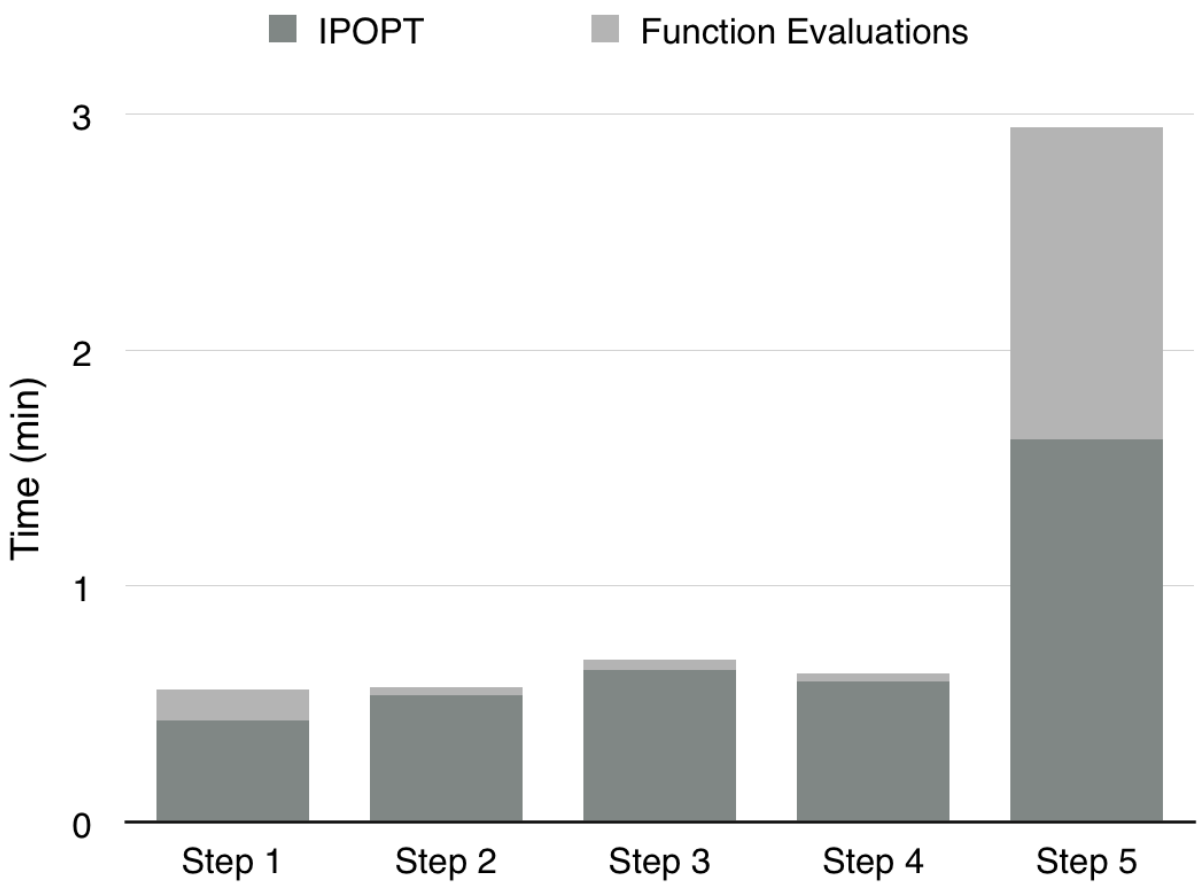

Figure 3.11. Average time comparison for 10 prostate cases that have 20853 voxels in average.

mean dose for sensitive organs), and at the same time reduces the variance of objective values. This is different from the Sobotta et al. approach in that we directly use variance, which we believe is a simpler and more reliable approach compared to their approach of trying to limit the tails of objective function probability distributions [64].

Conventionally, a physician contours a PTV for a CTV and a PRV for an OR. Fig. 3.7 and Fig. 3.8 showed that this method can be too conservative toward the target and hence yields high dose on some sensitive structures. This also confirms the intuition that there is a trade-off between pursuing robustness and sparing dose on sensitive structures. Notice that even though both PTV and PRV are used, PTV has a higher priority so the

solution leans to higher quality on PTV and hence worse quality on PRV. Another way to describe this trade-off is to view the structure metric as a random variable and to analyze its variance. PTV and PRV reduce the variance of the metric to almost 0 since PTV is designed to cover almost all the possible positions of CTV. But actually, sacrificing a little on the variance could reduce the dose a lot on sensitive structures, as in Fig. 3.10.

Fig. 3.6 indicates how the proposed model handled this trade-off and quantified it with the robustness slip. Besides quantifying the trade-off between achieving high dose on target and low dose on sensitive structures, robustness slip also describes how robust the existing model is. If the current model is already robust, then this robustness slip could be set as a tiny positive number or even 0. If the current model is not robust, then it can be set as a large number, like 1. This observation also suggests that it is worthwhile to test the proposed method on some more robust model that uses a target contour between CTV and PTV or even use PTV. But *Robust* keeps the target size as small as possible compared to PTV, which might overlap with sensitive structures.

The results shown in Fig. 3.6 and Fig. 3.7 demonstrate that different robustness weights resulted in different robustness priorities and hence different treatment plans. These results show that the proposed *Robust* model yields more robust treatment plans for a large family of robustness slip combinations. Different choices of robustness weights allow physicians to get robust dose control on structures they think more important. Automatically choosing robustness weights may be possible and will be the subject of future research.

The results showed that a robustness slip parameter can balance preserving the existing high quality prioritized optimization solution against the pursuit of plan robustness. Although the new added robust step enlarged the problem size and increased computa-

tion time with the number of sampled shifts, several useful techniques were introduced to speed-up the algorithm so that the average running time was less than 6 min., for all prostate cases, using a commodity 4-core computer.

Fig. 3.11 compared the time usage of *PTV* and *Robust*. These two methods share the same first 4 steps (except the CTV, PTV difference) but the latter has one more step. For IPOPT, there are two main factors in time usage: IPOPT itself, and function evaluation including evaluating objective, Jacobian and Hessian. The latter part is relatively fast for the first 4 steps because they have very few nonlinear constraints. The fifth step spends much more time on function evaluations because each motion of the target corresponds to a nonlinear constraint and to sum up many large Hessian matrices is time-consuming. IPOPT spends about twice as much time in the fifth step because of the increased problem size. Although the fifth step takes almost as much time as the first 4 steps, it is still less than 3 minutes and the whole optimization takes less than 6 minutes, despite the complicated formulation. Furthermore, reported times are for a 4-CPU-core machine. Since ORATOR was implemented in parallel, a machine with more cores may further decreases run-time.

3.3.5 Conclusion

Preliminary results show that ORATOR is likely to prove innovative in several important respects; in particular, we expect that it will simultaneously: (1) produce high-quality treatment plans in an automated fashion, not requiring planner input, (2) produce plans that follow clinical prescription guidelines regarding dose-volume characteristics (e.g., higher priority goals, such as a high minimum dose to the target volume, are protected when considering lower priority goals, such as lowering mean dose to sensitive

structures), and (3) produce robust plans that do not degrade under expected variations in tissue localization. All of this is done while providing Pareto optimal solutions, an important assurance that, for our hierarchical formulation, there are no other plans possible that could improve a dosimetric/planning goal without requiring a trade-off in a higher priority goal. In other words, with respect to the stated criteria, there are no better plans.

Although promising, several important steps would be required to validate the clinical use of ORATOR: (1) other sites would need to be tested, including sites with many more objectives (e.g., head-and-neck planning), (2) direct comparisons with expert clinical plans are needed, with comparable dosimetric calculation models.

We conclude that the proposed robustness formulation can be used to generate robust treatment plans against geometrical uncertainties. Although further testing is needed, the model has the potential to be clinically useful.

3.4 Dose Volume Metric Approximation

3.4.1 Introduction

Statistics or metrics that summarize the volume of the radiation dose to the target tumor or normal tissue are of essential importance in IMRT treatment planning and have been used extensively since the beginning of the 3D treatment planning era [82–84]. In addition to being used to judge treatment plan quality [85, 86], metrics are also used to analyze outcomes [87]. For target volumes, the most commonly used dose-volume metrics are of the form D_x, which is defined as the minimum dose to $x\%$ of the volume receiving the highest doses (i.e., the minimum dose to the 'hottest' $x\%$). In particular, D_{95} and D_{98} are in widespread use, but it is likely that other D_x values are being routinely used

as well. With the usual goal of exposing most of the target to high doses of radiation, higher D_x values are more desirable. Nonetheless, the physical characteristics of radiation delivery itself (involving Gaussian dose falloff due to lateral scattering and imperfect beam penumbra), as well as the desire to avoid any nearby radiation-sensitive normal tissue structures, means that some dose-falloff within the planning target volume (PTV) is unavoidable and therefore, should be quantified.

Although D_x values are easy to understand and determine from dose-volume histograms (DVHs), D_x values are difficult to manipulate in optimization problems for two reasons: (1) D_x is essentially an order statistic [88] that requires a sorting operation (performed when DVHs are formed, for example) and (2) D_x is not an analytic function, as is the case in general for order statistics. Of course, this implies that D_x is not differentiable (for example, with respect to fluence weights). Both of these points are obstacles to the development of efficient optimization algorithms that usually depend on convexity properties and differentiability as requirements for computational efficiency. IPOPT, for example, requires twice continuous differentiability.

A more attractive metric for treatment planning is the generalized equivalent uniform dose (gEUD) metric. Well-known in mathematics, the gEUD function was applied to IMRT as early as 1992 [84], but the name was coined by Niemierko [89] in 1999. The metric is defined as:

$$\text{gEUD}[a] = (\frac{1}{N}\sum_{i=1}^{N} d_i^a)^{\frac{1}{a}},$$

where N is the number of voxels in a given region of interest, d_i is the dose given to voxel i, and a is a selected or fitted parameter. The mathematical form of the gEUD function

has many uses outside radiotherapy and is known as the Hausdorff measure [90], the l_p-norm, and the generalized (or 'weighted') mean, by Abramowitz and Stegun [91].

Just as for D_x, gEUD is also a number between the minimum and maximum dose to a given structure. When the parameter a goes to ∞, gEUD goes to the maximum dose; or when a goes to $-\infty$, gEUD goes to the minimum dose value. Unlike D_x, gEUD is infinitely differentiable and hence, more amenable to be incorporated into IMRT treatment planning, either as an objective function or as a constraint. Moreover, gEUD itself has desirable convexity properties [92] that are not altered by including gEUD in linear equations.

The hypothesis of this work is that we can substitute a linear combination of gEUD terms as a reasonably accurate approximation for a given D_x value, thus making IMRT plan optimization more computationally tractable.

3.4.2 Methods and Materials

Input dose-volume histogram

Our analyzed dataset includes 498 de-identified head-and-neck treatment plans taken from a single institution. A sample of the DVHs for clinical target volumes (CTVs) is given in Fig. 3.12. As shown, treatment to different CTVs was not uniform. However, our main goal is to estimate D_x reasonably well with a concave function to use in IMRT optimization. A few DVHs had unusually low prescribed doses, around 40 Gy, and these were removed from the analysis. All other DVHs have prescribed doses between 50 Gy to 80 Gy. In addition, in order to focus on reasonably good-quality treatments, we filtered results to only include values of D_{95} that were greater than 95% of the prescribed dose, resulting in 432 DVHs. As noted below, this means that the formula may be less

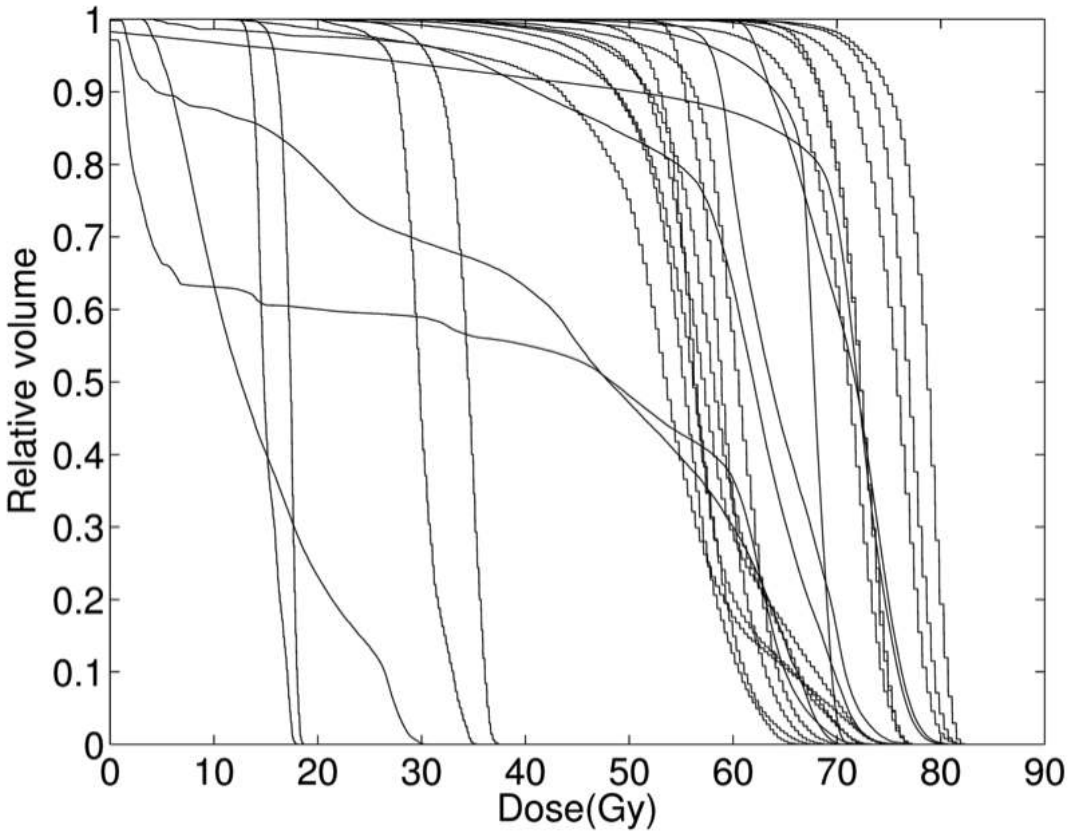

Figure 3.12. Sample of head-and-neck DVHs used to derive the relationship between gEUD and D_x. The full dataset includes 498 DVHs, but here we only show a sample, taking, at most, one DVH randomly from each 2-Gy bin in D_{95} dose.

accurate for poor-quality dose distributions but in such cases, high accuracy is less crucial. A sample of the resulting DVHs used for fitting is given in Fig. 3.13.

Fitting functions

We fitted the concave gEUD function to D_x values using either one- or two-gEUD functions, as follows:

Single-gEUD model:

$$D_x = c_1 \times \text{gEUD}[a] + c_0.$$

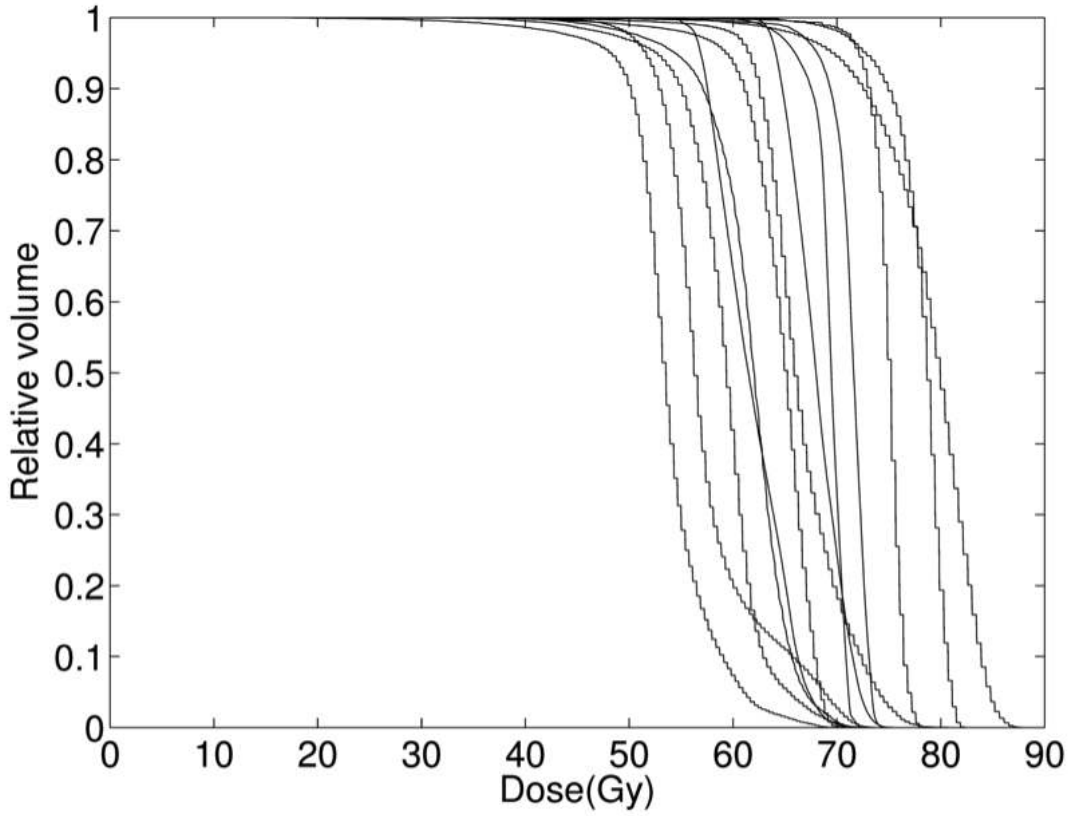

Figure 3.13. Sample of DVHs from 432 of the original 498 head-and-neck treatment plans with prescribed doses between 50 Gy and 80 Gy and D_{95} values greater than 90% of the prescribed dose. Again, here we show, at most, one DVH randomly from each 2-Gy bin in D_{95} dose.

Using this equation, we fitted the parameters a, c_0, and c_1, to minimize the root-mean-square error (RMSE) in the predicted D_x value. Two-gEUD model:

$$D_x = c_2 \times \text{gEUD}[a_2] + c_1 \times \text{gEUD}[a_1] + c_0.$$

In the two-gEUD model, we fitted a_1 and a_2, as well as c_0,c_1,c_2.

For both models, the MATLAB routine LSQLIN was used as the solver [93]. For each fit, we randomly divided the data between a training set (70%) and a testing set (30%). Values of x from 80 to 99 were considered. In order to avoid some non-concave solutions, we found it useful to limit the absolute magnitude of coefficients to less than 20.

3.4.3 Results

The single-gEUD model

Results for the single-gEUD model are given in Fig. 3.14.

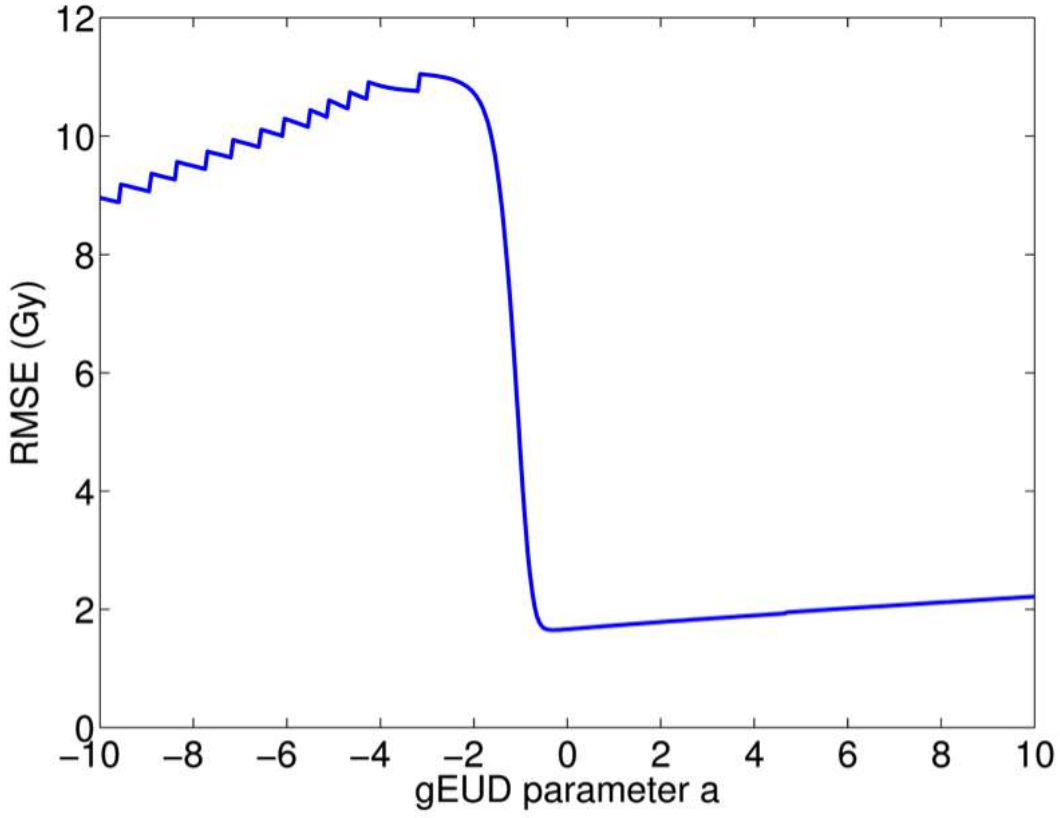

Figure 3.14. Fitting error for D_{95}, estimated using the single gEUD model.

The best fit occurs near a gEUD parameter $a = -0.35$. The error plot suggests that negative gEUD 'a' values with large magnitudes do not produce good fits, while positive or small-magnitude negative 'a' values generate good fits with average errors that are less than 2 Gy. The parameters for best fit are given in Table 3.6.

The two-gEUD model

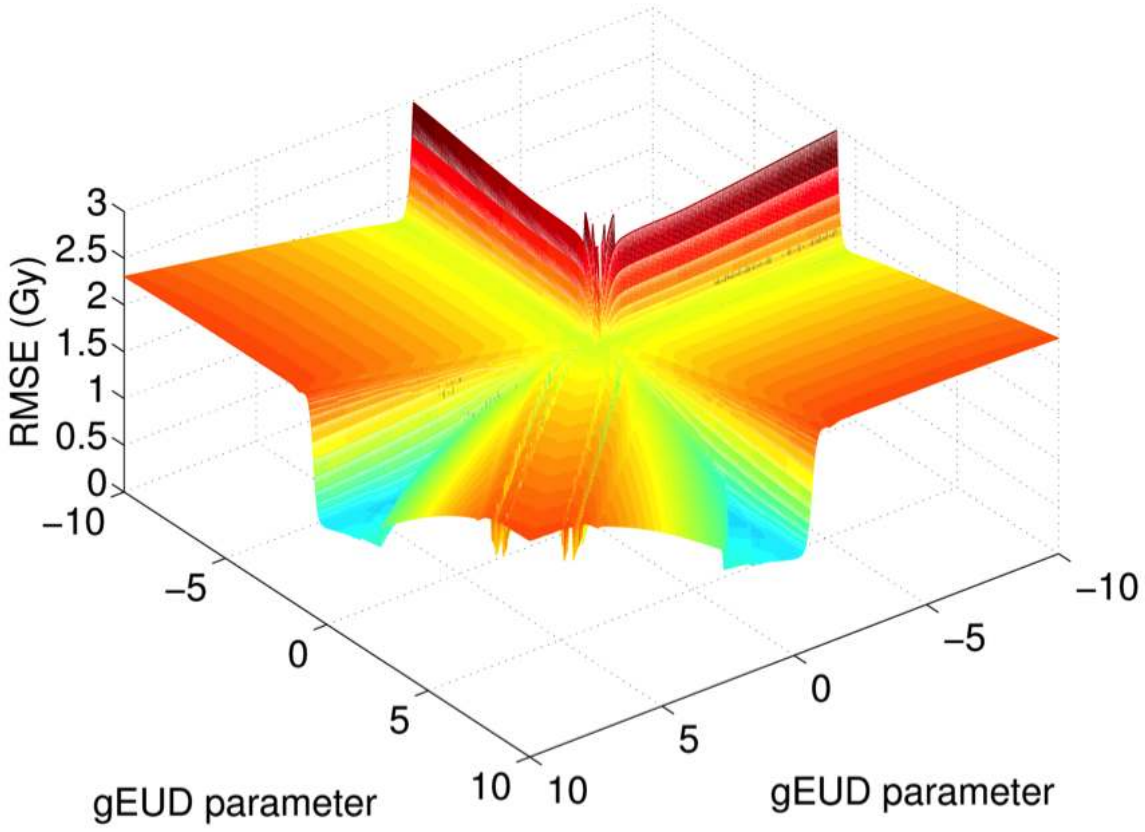

Figure 3.15. Best-fit errors for the two-gEUD model. Some points with error greater than 3 Gy are truncated.

For the fit with D_{95} alone, relatively high errors (>9 Gy) were produced when the best fit consisted of a combination of two negative gEUD 'a' parameters. The best results, with errors less than 1 Gy, resulted from combinations of two positive 'a' values. This is consistent with the finding that a single-gEUD model with either positive or slightly negative 'a' values resulted in better fits than large negative values (see Table 3.6 for the best fit results). The best two-gEUD fits are summarized in Fig. 3.15.

Similar patterns for fits obtained with the single- and two-gEUD models were observed when we cross-validated the fitting procedure using randomly chosen training datasets as

discussed below in this section.

	Best-fitted equation	Training error (Gy)	Testing error (Gy)
Single-gEUD model	$D_{95} = 0.97 \cdot \text{gEUD}[-0.3] - 3.49$	1.65	1.88
Two-gEUD model	$D_{95} = 16.25 \cdot \text{gEUD}[0.45] - 15.3 \cdot \text{gEUD}[1.75] - 0.69$	0.91	0.97

Table 3.6
Cross-validation of the best-fit equation, root mean square (RMS) training error, and testing error of one random run for both models.

Prediction error analysis

To test the robustness of the prediction method, we repeated the entire fitting process for data randomly divided into a 70% training set and a 30% testing set. This was repeated 5 times, and the average prediction performance was tabulated in Table 3.7. Scatter plots (Fig. 3.16 and Fig. 3.17) show the fitted distributions for both the training and testing set.

When $D_x = D_{95}$, both single- and two-gEUD models can generate very good fits with a mean-square error that is less than 2 Gy. By comparison, the two-gEUD model adds only one more gEUD parameter and one more coefficient but generates a better fit, with significantly less RMSE (0.93 Gy), as shown by the Fig. 3.16.

Fit for variable dose cutoffs in D_x

For optimization, it may be useful to have a single equation valid for various dose cutoffs in D_x, even if some price in accuracy is paid. In this fit, we let x in D_x vary from

	Mean (Gy)	Std (Gy)
Single-gEUD model training	1.722	0.0327
Single-gEUD model testing	1.726	0.0764
Two-gEUD model training	0.936	0.0241
Two-gEUD model testing	0.932	0.0415

Table 3.7
RMS training and testing errors for cross-validation runs. The mean and standard deviations of the RMSE are given for 5 runs.

80 to 99, where the number 80 is chosen as likely lower limit of what is clinically useful. We first compare how the fitting methods above behave for this range of x.

As shown by Fig. 3.18, when x increases from 80% to around 96%, the error goes up slightly, but when x becomes greater than 96%, the error increases rapidly. Moreover, when x is less than approximately 96%, the single-gEUD model generates an error less than 2 Gy, and the two-gEUD model generates an error less than 1 Gy. Comparing the two models, the two-gEUD model generates half the error that is generated by the single-gEUD model, when x is less than 96%.

Lastly, we investigate whether we can produce a very simplified equation that still yields a useful approximation across a wide range of x values.

First, we modify the two-gEUD model so that the best c_2, c_1, c_0, a_2, and a_1 are constants. The best-fit equation was found to be:

$$D_x = 3.18 \times \text{gEUD}[0.2] - 2.23 \times \text{gEUD}[6.2] + 0.3328. \tag{3.60}$$

The RMSEs were much higher in Equation 3.60 than in the original two-gEUD model, as shown in Fig. 3.19.

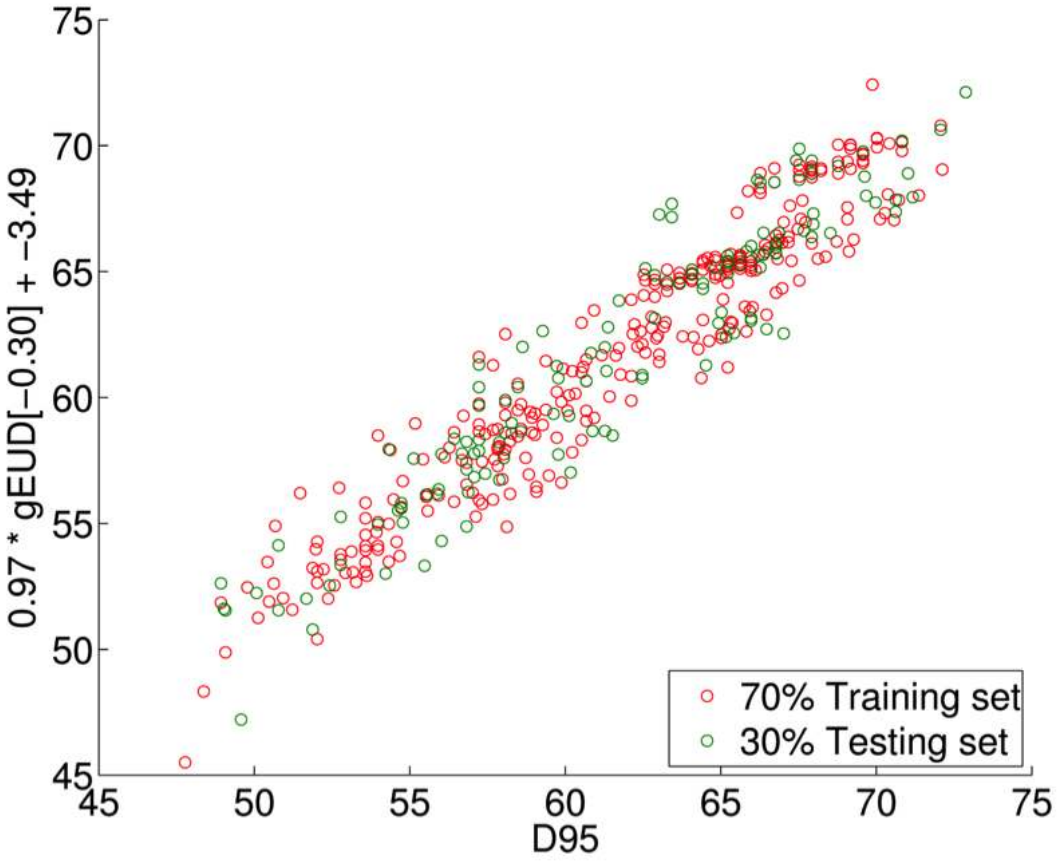

Figure 3.16. The scatter plot of D_{95} vs. the best-fitted single-gEUD model. Red circles are training points, whereas green circles are testing points.

We then modify the two-gEUD model so that the best c_2, c_1 , c_0, a_2, and a_1 for all x are each a function of x. It was found that linear functions of x were adequate, so linear regression is used to estimate these parameters. The best-fit equation was found to be:

$$D_x = (0.15x{-}9.66){\times}\text{gEUD}[0.45]{+}(-0.15x{+}11.08){\times}\text{gEUD}[-0.03x{+}10.57]{+}0.13x{-}10.75. \tag{3.61}$$

The RMSE curve for Equation 3.61 (Fig. 3.19) shows a performance that is much closer to that of the two-gEUD model than that of the model with constant coefficients and gEUD parameters in Equation 3.60.

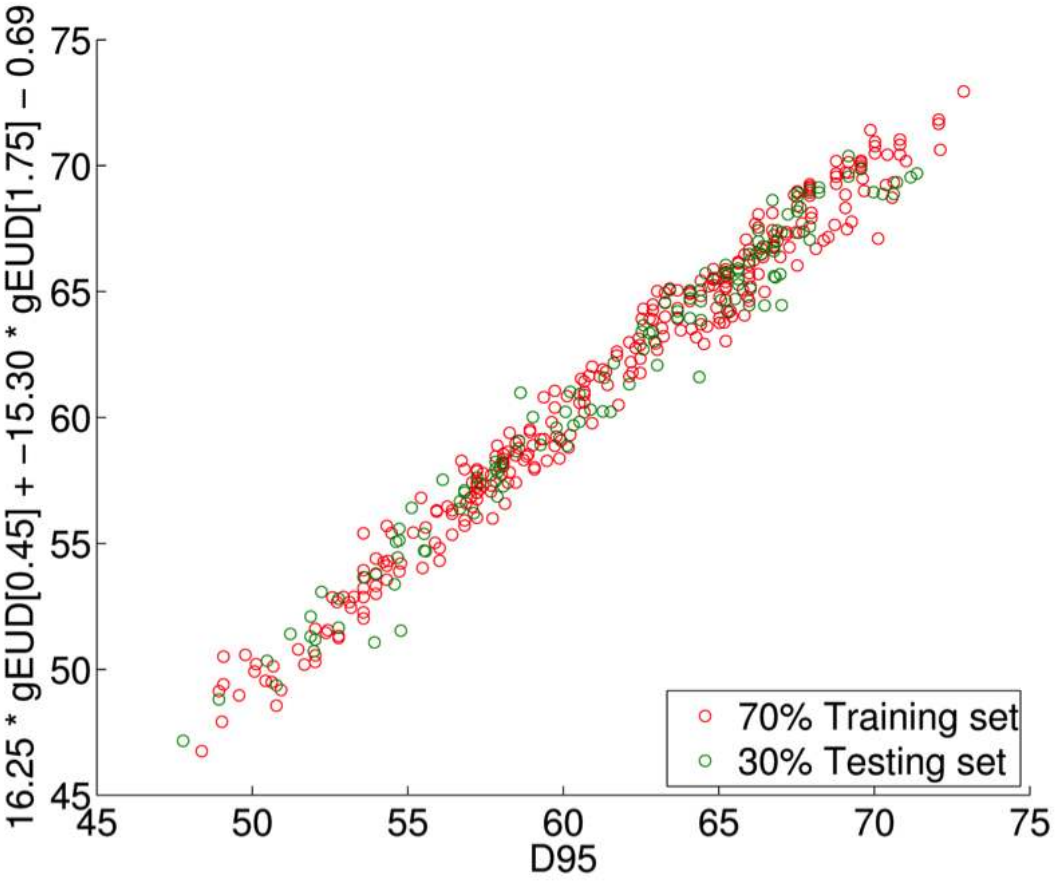

Figure 3.17. The scatter plot of D_{95} vs. the best-fitted equation of the two-gEUD model. Red circles are training points, whereas green circles are testing points. Note the reduced scatter, compared to the single-gEUD model (Fig. 3.16).

3.4.4 Conclusion

These concave metrics fulfill an important need for computationally efficient IMRT treatment planning objectives. Previously, the D_x objectives that were commonly used introduced numerical difficulties due to a lack of convexity. As an important dose-volume metric, D_x has another drawback: it is not analytic or differentiable and hence is difficult to implement for IMRT optimization in a computationally efficient manner. We have proposed to use a linear combination of gEUDs to estimate D_x. In particular, two models

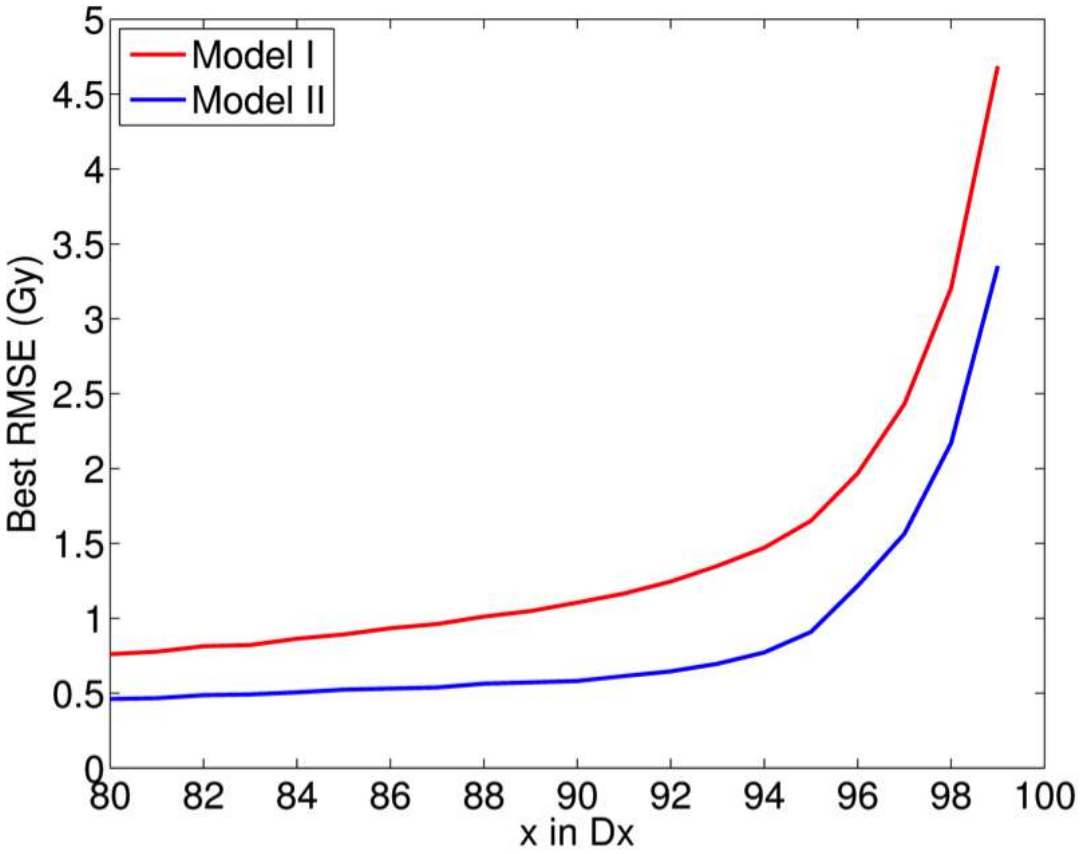

Figure 3.18. The plot of x in D_x vs. best RMSE for the single-gEUD model (red line) and two-gEUD model (blue line). The models are re-fitted for variable x values.

were proposed to represent D_{95}, one involved a single gEUD and the other involved two gEUDs. A more general equation that represented D_x accurately for a wide range of x values (80% to about 96%) was also derived. Cross-validation showed that the RMSE for the best fits was less than 2 Gy and typically less than 1 Gy.

Although these results establish that this type of approximation can be useful for describing a cohort of DVHs, it is important to note that these fits are probably less accurate for a different clinic using a different planning system and different directives on a different patient cohort. It would be fairly straightforward to test the equations

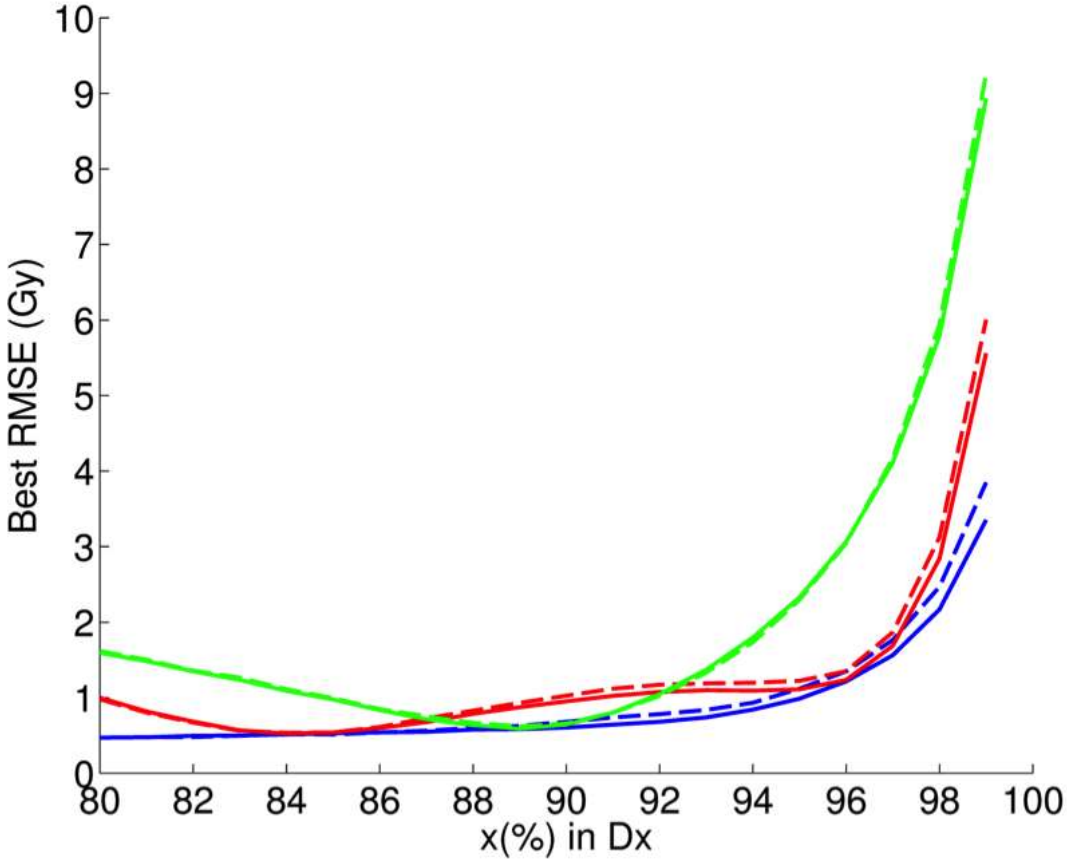

Figure 3.19. A plot of x in D_x vs. RMSE for the two-gEUD model (red lines) and two-gEUD model (blue lines), with variable parameters, i.e., finding the best combination of gEUD parameters when varying x. The green lines are for Equation 3.60, and the red lines are for the Equation 3.61. Solid lines are training/cross-validation errors, and dashed lines are testing cohort errors. Notice the training and testing errors for Equation 3.60 coincide with each other almost everywhere. The figure shows that a good approximation to D_x can be obtained with two gEUD terms between 80% and 96%.

for new DVHs, or even to re-fit the equations to new datasets. We have focused on D_x values, rather than V_x values, because the statistics of V_x values tend to be noisier. Future research will test fits for V_x values, in particular, for normal tissues. A second caveat is that these results were derived for 'good DVHs'. This means that the approximation is likely to be worse for poor-quality DVHs. However, we do not believe this is a significant

problem because the planning objective surrogate only needs to be highly accurate when the plan approaches acceptability.

A given IMRT optimization problem might seek to control more than one point on the DVH curve and therefore use the variable-x equation applied to more than one x value. We have shown that non-convex D_x values can be well-approximated for a range of x values using a two-gEUD model for a given cohort of DVHs. The resulting equation can be used for IMRT optimization and has the advantages of simplicity, differentiability, convexity, and generality for multiple x values.

4. Conclusion

In this dissertation, nonlinear optimization is introduced and some of its applications in machine learning and medical physics are described.

In Chapter 2, a problem called MVU (maximum variance unfolding) in machine learning is introduced as a non-convex nonlinear optimization, that is essentially to learn a low dimensional manifold from a large dimensional space. Instead of converting the non-convex nonlinear optimization to a semi-definite optimization, we directly solve it with an interior point method and got more accurate results.

In Chapter 3, radiation therapy, IMRT in particular, is introduced as a multi-objective nonlinear optimization as physicians make multiple contradicting goals for tumors and other structures. There are mainly two problems that have been discussed. First, physicians describe treatment plan quality by favoring certain metrics/functions, but not all of these functions are nicely behaved. One of these metrics D_x, an order statistic, is approximated by a nicely behaved (infinitely differentiable) function, which could potentially be incorporated into IMRT optimization as a surrogate for D_x. Second, good treatment plans have to be robust against motion uncertainties, which are either due to device set-up error or patient movement. Based on the prioritized prescription optimization algorithm, a new optimization model has been studied which includes a robustness optimization step that improves the expected performance of existing treatment plans.

REFERENCES

[1] Dimitri P Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.

[2] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[3] Dan Jurafsky and James H Martin. *Speech & language processing*. Pearson Education India, 2000.

[4] Tom M Mitchell. *Machine learning*. McGraw Hill, 1997.

[5] Suvrit Sra, Sebastian Nowozin, and Stephen J Wright. *Optimization for machine learning*. Mit Press, 2012.

[6] Kristin P Bennett and Emilio Parrado-Hernández. The interplay of optimization and machine learning research. *The Journal of Machine Learning Research*, 7:1265–1281, 2006.

[7] Karl Pearson. LIII. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[8] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer, 2007.

[9] Lawrence K Saul, Kilian Q Weinberger, Jihun H Ham, Fei Sha, and Daniel D Lee. Spectral methods for dimensionality reduction. *Semisupervised Learning*, pages 293–308, 2006.

[10] John C Platt. Fast embedding of sparse music similarity graphs. *Advances in Neural Information Processing Systems*, 16:571578, 2004.

[11] Vin De Silva and Joshua B Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. In *Advances in Neural Information Processing Systems 15*, pages 705–712. MIT Press, 2003.

[12] Kilian Q Weinberger, Fei Sha, Qihui Zhu, and Lawrence K Saul. Graph Laplacian regularization for large-scale semidefinite programming. In *Advances in neural information processing systems*, pages 1489–1496, 2006.

[13] David Leigh Donoho and Carrie Grimes. *When does ISOMAP recover the natural parameterization of families of articulated images?* Department of Statistics, Stanford University, 2002.

[14] Alexander Paprotny and Jochen Garcke. On a connection between maximum variance unfolding, shortest path problems and isomap. In *International Conference on Artificial Intelligence and Statistics*, pages 859–867, 2012.

[15] Michael Gashler and Tony Martinez. Temporal nonlinear dimensionality reduction. In *Neural Networks (IJCNN), The 2011 International Joint Conference*, pages 1959–1966. IEEE, 2011.

[16] Daniel Delling, Peter Sanders, Dominik Schultes, and Dorothea Wagner. Engineering route planning algorithms. In *Algorithmics of large and complex networks*, pages 117–139. Springer, 2009.

[17] D Chris Rayner, Michael H Bowling, and Nathan R Sturtevant. Euclidean heuristic optimization. In *AAAI*, 2011.

[18] Wenlin Chen, Yixin Chen, and Kilian Q Weinberger. Maximum variance correction with application to A^* search. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 28(1):302–310, 2013.

[19] Michael Gashler, Dan Ventura, and Tony Martinez. Manifold learning by graduated optimization. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(6):1458–1470, 2011.

[20] M Tim Jones. *Artificial Intelligence A System Approach*. Laxmi Publications Pvt Limited, 2008.

[21] Naresh Gupta and Dana S Nau. On the complexity of blocks-world planning. *Artificial Intelligence*, 56(2):223–254, 1992.

[22] Anders Brahme. Optimization of stationary and moving beam radiation therapy techniques. *Radiotherapy and Oncology*, 12(2):129–140, 1988.

[23] Clifton D Fuller, Mehee Choi, Britta Forthuber, Samuel J Wang, Nancy Rajagiriyil, Bill J Salter, and Martin Fuss. Standard fractionation intensity modulated radiation therapy (IMRT) of primary and recurrent glioblastoma multiforme. *Radiation Oncology*, 2(26):412–418, 2007.

[24] Jack F Fowler and Mark A Ritter. A rationale for fractionation for slowly proliferating tumors such as prostatic adenocarcinoma. *International Journal of Radiation Oncology* Biology* Physics*, 32(2):521–529, 1995.

[25] Angela van Baardwijk, Geert Bosmans, Søren M Bentzen, Liesbeth Boersma, André Dekker, Rinus Wanders, Bradly G Wouters, Philippe Lambin, and Dirk De Ruysscher. Radiation dose prescription for non–small-cell lung cancer according to normal tissue dose constraints: An *in silico* clinical trial. *International Journal of Radiation Oncology* Biology* Physics*, 71(4):1103–1110, 2008.

[26] M Kara Bucci, Alison Bevan, and Mack Roach. Advances in radiation therapy: conventional to 3D, to IMRT, to 4D, and beyond. *CA: A Cancer Journal for Clinicians*, 55(2):117–134, 2005.

[27] Steve Webb. Optimizing the planning of intensity-modulated radiotherapy. *Physics in Medicine and Biology*, 39(12):2229, 1994.

[28] Anders Brahme and Anna K Årgren. Optimal dose distribution for eradication of heterogeneous tumors. *Acta Oncologica*, 26(5):377–385, 1987.

[29] Anders Brahme. Biological and physical dose optimization in radiation therapy. *Accomplishments in Cancer Research*, pages 265–298, 1991.

[30] Patrick Källman, Anna K Ågren, and Anders Brahme. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *International Journal of Radiation Biology*, 62(2):249–262, 1992.

[31] Personal communication with Margie A Hunt at Memorial Sloan-Kettering Cancer Center, 2014.

[32] Joseph O Deasy, James R Alaly, and Konstantin Zakaryan. Obstacles and advances in intensity-modulated radiation therapy treatment planning. *Frontiers of Radiation Therapy and Oncology*, 2007.

[33] Joseph O Deasy, Andrzej Niemierko, Donald Herbert, Di Yan, Andrew Jackson, Randall K Ten Haken, Mark Langer, and Steve Sapareto. Methodological issues in radiation dose–volume outcome analyses: summary of a joint AAPM/NIH workshop. *Medical Physics*, 29(9):2109–2127, 2002.

[34] Steve Webb. *Intensity-modulated radiation therapy*. CRC Press, 2001.

[35] Joseph O Deasy. Multiple local minima in radiotherapy optimization problems with dose–volume constraints. *Medical Physics*, 24(7):1157–1161, 1997.

[36] Beong Choi and Joseph O Deasy. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Physics in Medicine and Biology*, 47(20):3579, 2002.

[37] ICRU. Report 62. *Prescribing, Recording and Reporting Photon Beam Therapy (Supplement to ICRU Report 50)*, 1999.

[38] Joseph O Deasy, Renato De Leone, and T W Holmes. Beam weight optimization using the MINOS code. In *The Use of Computers in Radiation Therapy: Proceedings of the XI the International Conference on the Use of Computers in Radiation Therapy*, pages 64–65. Stockport, UK: Handley Printers Ltd, 1994.

[39] Peter WJ Voet. *Automation of Contouring and Planning in Radiotherapy*. PhD thesis, Erasmus MC: University Medical Center Rotterdam, 2014.

[40] Benedick A Fraass. Differences between plan evaluation and the optimization problem statement, and the difference it makes. *Operations Research Applications in Radiotherapy (ORART)*, 2002.

[41] Joseph O Deasy. The IMRT optimization problem statement: IMRT software design goals. *Operations Research Applications in Radiotherapy (ORART)*, 2002.

[42] Mark Langer, Eva K Lee, Joseph O Deasy, Ronald L Rardin, and James A Deye. Operations research applied to radiotherapy, an NCI–NSF–sponsored workshop February 7–9, 2002. *International Journal of Radiation Oncology* Biology* Physics*, 57(3):762–768, 2003.

[43] Thomas Bortfeld and Robert Jeraj. The physical basis and future of radiation therapy. *The British Journal of Radiology*, 84(1002):485–498, June 2011.

[44] Kyung-Wook Jee, Daniel L McShan, and Benedick A Fraass. Lexicographic ordering: intuitive multicriteria optimization for IMRT. *Physics in Medicine and Biology*, 52(7):1845, 2007.

[45] Jan J Wilkens, James R Alaly, Konstantin Zakarian, Wade L Thorstad, and Joseph O Deasy. IMRT treatment planning based on prioritizing prescription goals. *Physics in Medicine and Biology*, 52(6):1675, 2007.

[46] Sebastiaan Breedveld, Pascal RM Storchi, Marleen Keijzer, Arnold W Heemink, and Ben JM Heijmen. A novel approach to multi-criteria inverse planning for IMRT. *Physics in Medicine and Biology*, 52(20):6339, 2007.

[47] Peter WJ Voet, Sebastiaan Breedveld, Maarten LP Dirkx, Peter C Levendag, and Ben JM Heijmen. Integrated multicriterial optimization of beam angles and intensity

profiles for coplanar and noncoplanar head and neck IMRT and implications for VMAT. *Medical Physics*, 39(8):4858–4865, 2012.

[48] Peter WJ Voet, Maarten LP Dirkx, Sebastiaan Breedveld, Dennie Fransen, Peter C Levendag, and Ben JM Heijmen. Toward fully automated multicriterial plan generation: a prospective clinical study. *International Journal of Radiation Oncology* Biology* Physics*, 85(3):866–872, 2013.

[49] Peter WJ Voet, Maarten LP Dirkx, Sebastiaan Breedveld, Abrahim Al-Mamgani, Luca Incrocci, and Ben JM Heijmen. Fully automated volumetric modulated arc therapy plan generation for prostate cancer patients. *International Journal of Radiation Oncology* Biology* Physics*, 88(5):1175–1179, 2014.

[50] Dylan Jones and Mehrdad Tamiz. *Practical goal programming*, volume 141. Springer, 2010.

[51] Vanessa H Clark, Yixin Chen, Jan Wilkens, J R Alaly, K Zakaryan, and Joseph O Deasy. IMRT treatment planning for prostate cancer using prioritized prescription optimization and mean-tail-dose functions. *Linear Algebra and its Applications*, 428(5):1345–1364, 2008.

[52] Martha M Matuszak, Edward W Larsen, Kyung-Wook Jee, Daniel L McShan, and Benedick A Fraass. Adaptive diffusion smoothing: A diffusion-based method to reduce IMRT field complexity. *Medical Physics*, 35(4):1532–1546, 2008.

[53] Aaron C Spalding, Kyung-Wook Jee, Karen Vineberg, Marla Jablonowski, Benedick A Fraass, Charlie C Pan, Theodore S Lawrence, Randall K Ten Haken, and Edgar Ben-Josef. Potential for dose-escalation and reduction of risk in pancreatic

cancer using IMRT optimization with lexicographic ordering and gEUD-based cost functions. *Medical Physics*, 34(2):521–529, 2007.

[54] Troy Long, Martha M Matuszak, Mary Feng, Benedick A Fraass, Randall K Ten Haken, and H Edwin Romeijn. Sensitivity analysis for lexicographic ordering in radiation therapy treatment planning. *Medical Physics*, 39(6):3445–3455, 2012.

[55] Wei Liu, Steven J Frank, Xiaoqiang Li, Yupeng Li, Peter C Park, Lei Dong, X Ronald Zhu, and Radhe Mohan. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. *Medical Physics*, 40(5):051711, 2013.

[56] Albin Fredriksson. A characterization of robust radiation therapy treatment planning methodsfrom expected value to worst case optimization. *Medical Physics*, 39(8):5169–5181, 2012.

[57] Arinbjörn Olafsson and Stephen J Wright. Efficient schemes for robust IMRT treatment planning. *Physics in Medicine and Biology*, 51(21):5621, 2006.

[58] Millie Chu, Yuriy Zinchenko, Shane G Henderson, and Michael B Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Physics in Medicine and Biology*, 50(23):5463, 2005.

[59] Christoph Baum, Markus Alber, Mattias Birkner, and Fridtjof Nüsslin. Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities. *Radiotherapy and Oncology*, 78(1):27–35, 2006.

[60] Di Yan. Adaptive radiotherapy: merging principle into clinical practice. *Seminars in Radiation Oncology*, 20(2):79–83, April 2010.

[61] Qiuwen Wu, Yuwei Chi, Peter Y Chen, Daniel J Krauss, Di Yan, and Alvaro Martinez. Adaptive replanning strategies accounting for shrinkage in head and neck IMRT. *International Journal of Radiation Oncology* Biology* Physics*, 75(3):924–932, 2009.

[62] Di Yan, Frank Vicini, John Wong, and Alvaro Martinez. Adaptive radiation therapy. *Physics in Medicine and Biology*, 42(1):123, 1997.

[63] Jonathan G Li and Lei Xing. Inverse planning incorporating organ motion. *Medical Physics*, 27(7):1573–1578, 2000.

[64] B Sobotta, M Söhn, and M Alber. Robust optimization based upon statistical theory. *Medical Physics*, 37(8):4019–4028, 2010.

[65] Douglas Jones. ICRU report 50-prescribing, recording and reporting photon beam therapy. *Medical Physics*, 21(6):833–834, 1994.

[66] Søren M Bentzen, Louis S Constine, Joseph O Deasy, Avi Eisbruch, Andrew Jackson, Lawrence B Marks, Randall K Ten Haken, and Ellen D Yorke. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. *International Journal of Radiation Oncology* Biology* Physics*, 76(3):S3–S9, 2010.

[67] Hugo JWL Aerts, Angela AW van Baardwijk, Steven F Petit, Claudia Offermann, Judith van Loon, Ruud Houben, Anne-Marie C Dingemans, Rinus Wanders, Liesbeth Boersma, Jacques Borger, et al. Identification of residual metabolic-active areas within individual NSCLC tumours using a pre-radiotherapy Fluorodeoxyglucose-PET-CT scan. *Radiotherapy and Oncology*, 91(3):386–392, 2009.

[68] Michael A Deveau, Stephen R Bowen, David C Westerly, and Robert Jeraj. Feasibility and sensitivity study of helical tomotherapy for dose painting plans. *Acta Oncologica*, 49(7):991–996, 2010.

[69] Joep C Stroom and Ben JM Heijmen. Limitations of the planning organ at risk volume (PRV) concept. *International Journal of Radiation Oncology* Biology* Physics*, 66(1):279–286, 2006.

[70] Kyung-Wook Jee, Daniel L McShan, and Benedick A Fraass. Lexicographic ordering: intuitive multicriteria optimization for IMRT. *Physics in Medicine and Biology*, 52(7):1845, 2007.

[71] Sebastiaan Breedveld, Pascal RM Storchi, Peter WJ Voet, and Ben JM Heijmen. iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. *Medical Physics*, 39(2):951–963, 2012.

[72] Joep C Stroom, Hans CJ de Boer, Henk Huizenga, and Andries G Visser. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *International Journal of Radiation Oncology* Biology* Physics*, 43(4):905–919, 1999.

[73] Joseph A Moore, John J Gordon, Mitchell S Anscher, and Jeffrey V Siebers. Comparisons of treatment optimization directly incorporating random patient setup uncertainty with a margin-based approach. *Medical Physics*, 36(9):3880–3890, 2009.

[74] Kevin L Moore, R Scott Brame, Daniel A Low, and Sasa Mutic. Quantitative metrics for assessing plan quality. *Seminars in Radiation Oncology*, 22(1):62–69, January 2012.

[75] Román Bohoslavsky, Marnix G Witte, Tomas M Janssen, and Marcel Van Herk. Probabilistic objective functions for margin-less IMRT planning. *Physics in Medicine and Biology*, 58(11):3563, 2013.

[76] Jan Unkelbach and Uwe Oelfke. Incorporating organ movements in inverse planning: assessing dose uncertainties by bayesian inference. *Physics in Medicine and Biology*, 50(1):121, 2005.

[77] Jan Unkelbach and Uwe Oelfke. Incorporating organ movements in IMRT treatment planning for prostate cancer: minimizing uncertainties in the inverse planning process. *Medical Physics*, 32(8):2471–2483, 2005.

[78] Jan Unkelbach and Uwe Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Physics in Medicine and Biology*, 49(17):4005, 2004.

[79] Joseph Leong. Implementation of random positioning error in computerised radiation treatment planning systems as a result of fractionation. *Physics in Medicine and Biology*, 32(3):327, 1987.

[80] Joseph O Deasy, Angel I Blanco, and Vanessa H Clark. CERR: a computational environment for radiotherapy research. *Medical Physics*, 30(5):979–985, 2003.

[81] HSL. A collection of Fortran codes for large-scale scientific computation. *See http://www.hsl.rl.ac.uk*, 2011.

[82] Joseph A Moore, J James Gordon, Mitchell Anscher, Joaquin Silva, and Jeffrey V Siebers. Comparisons of treatment optimization directly incorporating system-

atic patient setup uncertainty with a margin-based approach. *Medical Physics*, 39(2):1102–1111, 2012.

[83] R Mohan, G S Mageras, B Baldwin, L J Brewster, G J Kutcher, S Leibel, C M Burman, C C Ling, and Z Fuks. Clinically relevant optimization of 3-D conformal treatments. *Medical Physics*, 19(4):933–944, 1992.

[84] Andrzej Niemierko, Marcia Urie, and Michael Goitein. Optimization of 3D radiation therapy with both physical and biological end points and constraints. *International Journal of Radiation Oncology* Biology* Physics*, 23(1):99–108, 1992.

[85] Robert E Drzymala, Michael D Holman, Di Yan, William B Harms, Nilesh L Jain, Michael G Kahn, Bahman Emami, and James A Purdy. Integrated software tools for the evaluation of radiotherapy treatment plans. *International Journal of Radiation Oncology* Biology* Physics*, 30(4):909–919, 1994.

[86] Christian Thieke, Thomas Bortfeld, and Karl-Heinz Küfer. Characterization of dose distributions through the max and mean dose concept. *Acta Oncologica*, 41(2):158–161, 2002.

[87] I El Naqa, G Suneja, PE Lindsay, A J Hope, J R Alaly, M Vicic, JD Bradley, Aditya Apte, and Joseph O Deasy. Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. *Physics in medicine and biology*, 51(22):5719, 2006.

[88] Robert V Hogg and Allen T Craig. *Introduction to mathematical statistics*. Macmillan, New York, 1978.

[89] Andrzej Niemierko. A generalized concept of equivalent uniform dose (EUD). *Medical Physics*, 26(6):1100, 1999.

[90] Felix Hausdorff. Dimension und äußeres maß. *Mathematische Annalen*, 79(1-2):157–179, 1918.

[91] Milton Abramowitz and Irene A Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Courier Dover Publications, 1964.

[92] Beong Choi and Joseph O Deasy. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Physics in Medicine and Biology*, 47(20):3579, 2002.

[93] Thomas F Coleman and Yuying Li. A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6(4):1040–1058, 1996.