Optimization under Uncertainty in Radiation Therapy

by

Timothy Ching-Yee Chan

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2007

Author .....
Sloan School of Management
May 17, 2007

Certified by .....
John N. Tsitsiklis
Clarence J. Lebel Professor of Electrical Engineering
Thesis Supervisor

Certified by .....
Thomas Bortfeld
Associate Professor of Medicine, Harvard Medical School
Thesis Supervisor

Accepted by .....
Dimitris J. Bertsimas
Boeing Professor of Operations Research
Co-Director, Operations Research Center

Optimization under Uncertainty in Radiation Therapy

by

Timothy Ching-Yee Chan

Submitted to the Sloan School of Management
on May 17, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

In the context of patient care for life-threatening illnesses, the presence of uncertainty may compromise the quality of a treatment. In this thesis, we investigate robust approaches to managing uncertainty in radiation therapy treatments for cancer. In the first part of the thesis, we study the effect of breathing motion uncertainty on intensity-modulated radiation therapy treatments of a lung tumor. We construct a robust framework that generalizes current mathematical programming formulations that account for motion. This framework gives insight into the trade-off between sparing the healthy tissues and ensuring that the tumor receives sufficient dose. With this trade-off in mind, we show that our robust solution outperforms a nominal (no uncertainty) solution and a margin (worst-case) solution on a clinical case.

Next, we perform an in-depth study into the structure of different intensity maps that were witnessed in the first part of the thesis. We consider parameterized intensity maps and investigate their ability to deliver a sufficient dose to the tumor in the presence of motion that follows a Gaussian distribution. We characterize the structure of optimal intensity maps in terms of certain conditions on the problem parameters.

Finally, in the last part of the thesis, we study intensity-modulated proton therapy under uncertainty in the location of maximum dose deposited by the beamlets of radiation. We provide a robust formulation for the optimization of proton-based treatments and show that it outperforms traditional formulations in the face of uncertainty. In our computational experiments, we see evidence that optimal robust solutions use the physical characteristics of the proton beam to create dose distributions that are far less sensitive to the underlying uncertainty.

Thesis Supervisor: John N. Tsitsiklis
Title: Clarence J. Lebel Professor of Electrical Engineering

Thesis Supervisor: Thomas Bortfeld
Title: Associate Professor of Medicine, Harvard Medical School

Acknowledgments

First off, I would like to give my sincerest and everlasting thanks to JNT for being the second best thing to happen to me while at MIT. He is the most patient, thoughtful, dedicated, rigorous, and insightful person I know, and I could not have made this journey without him. His work ethic and humor never ceased to inspire me. He is my academic aspiration.

I owe a great debt of gratitude to TRB for introducing me to radiation therapy research, and for being a supportive co-advisor throughout my degree. He helped me to never lose sight of the practical and relevant motivations on which my research is based. I would also like to acknowledge EF, with whom I worked during my first two years and who helped me get off the ground at MIT.

I am extremely grateful to DB and RF for serving on my thesis committee. They both encouraged and challenged me to improve my thesis in ways that I had not considered. They, along with GP and AB, also served as wonderful mentors during my job search. I had unforgettable teaching experiences at MIT, and will fondly remember sparring with AB. I would like to thank the OR faculty for the effort that they put into making the ORC a special place on campus, and especially JBO for his tireless dedication to the Center. The staff at the ORC (PM, LR, AC, VM) was a well-oiled machine, and helped me on countless occasions.

I am very thankful for the good friends that I made at the ORC: GR, ABN, KMT, EA, MVB, NU, LJC, DC, RS, AS, SM, PX, LK, CS, PG, YLT, KK, SG, TW, JD, and PH. I also had the wonderful experience of living in SP for most of my graduate career. Thanks to RB, KS, DCO, IB, AK_1, YGW, LLD, RGM, DM, AK_2, and RT for making SP my home, and not just a residence. I would also like to thank my wonderful friends and colleagues at MGH: DLC, AT, JU, BCM, TH, GCS, CV, and AK_3. Thanks especially to DLC for helping me a lot when I was starting out.

I could not have made it to MIT without the support of and guidance from professors from my undergraduate institution, UBC. In particular, I would like to thank RA, TM, and MQ for encouraging me to come to MIT, as well as MP, who I

did not know until after I left, but who has become a valuable mentor since.

I am forever grateful to my parents, GMC and EMC, for their unconditional love and encouragement. They have given me every opportunity that they never had. Finally, we have arrived at "the best thing to happen to me at MIT." That title belongs to LCC. Her friendship, compassion and support have helped me through many challenges during my graduate career. I feel truly blessed to have found a kindred spirit at MIT, who shares my passions and outlook in life.

The research presented in this thesis was supported in part by the National Science Foundation under grant ECS-0312921, the United States-Israel Binational Science Foundation (BSF) under grant 2003275, the Natural Sciences and Engineering Research Council of Canada, and the Hugh Hampton Young Memorial Fund Fellowship.

Contents

1 Introduction **15**

2 Robust Intensity-Modulated Radiation Therapy **21**

2.1 IMRT Background 23

2.2 Motion uncertainty in lung tumors 24

2.3 Model of uncertainty 26

2.4 Formulations 28

2.4.1 Nominal formulation 28

2.4.2 Robust formulation 31

2.4.3 Margin formulation 37

2.4.4 A delayed constraint generation approach to solving the robust formulation 41

2.5 Patient Data and PDFs 42

2.5.1 Creating the uncertainty set from patient data 47

2.6 A one-dimensional case 48

2.6.1 Modifications for the one-dimensional case 48

2.6.2 Results for the one-dimensional case 51

2.7 Results from a clinical lung case 56

2.8 A Pareto viewpoint 66

2.9 Conclusions 67

3 Optimal Parameterized Intensity Maps **69**

3.1 Optimal margins 71

3.1.1 Summary and Highlights . . . 73
3.1.2 Analysis . . . 73
3.2 Optimal margins with robustness . . . 84
3.2.1 Summary and Highlights . . . 84
3.2.2 Analysis . . . 85
3.3 Optimal 3D margins . . . 90
3.3.1 Summary and Highlights . . . 90
3.3.2 Analysis . . . 91
3.4 Optimal 3D margins with robustness . . . 96
3.4.1 Summary and Highlights . . . 97
3.4.2 Analysis . . . 98
3.5 Optimal edge-enhancements . . . 106
3.5.1 Summary and Highlights . . . 108
3.5.2 Analysis . . . 108
3.6 Optimal edge-enhancements with robustness . . . 134
3.6.1 Summary and Highlights . . . 135
3.6.2 Analysis . . . 135
3.7 Conclusions . . . 138

4 Robust Intensity-Modulated Proton Therapy 141
4.1 IMPT background . . . 142
4.2 Range uncertainty . . . 144
4.3 Model of Uncertainty . . . 145
4.4 Problem setup . . . 146
4.5 Formulations . . . 148
4.5.1 Nominal Formulation . . . 149
4.5.2 Robust formulation . . . 150
4.6 Results from a phantom study . . . 158
4.7 Conclusions . . . 166

5 Conclusions 173

List of Figures

2-1 An example RPM trace. 45
2-2 The pdf corresponding to the RPM trace in Figure 2-1. 45
2-3 Position of the tumor centroid in the ten phases of the breathing cycle. Phases 1, 3, 5, 7, 9 are labeled 46
2-4 The pdfs used in the nominal, margin and robust formulation illustrations. 53
2-5 Dose distribution of nominal solution using the pdfs from Figure 2-4. 53
2-6 Dose distribution of margin solution using the pdfs from Figure 2-4. . 54
2-7 The pdfs and error bars used in the robust formulation illustrations. . 54
2-8 Dose distribution of robust solution using the pdfs and error bars from Figure 2-7. 55
2-9 Intensity map corresponding to robust solution. 55
2-10 Dose distribution corresponding to the robust solution when the realized pdf is the same as the nominal one. The white outline denotes the GTV, while the green outline denotes the CTV. 59
2-11 Nominal solution DVHs for 78 realized pdfs in experiment #1. 60
2-12 Robust solution DVHs for 78 realized pdfs in experiment #1. 61
2-13 Margin solution DVHs for 78 realized pdfs in experiment #1. 61
2-14 The nominal, robust, and margin solution DVHs for one realized pdf in experiment #1. 62
2-15 Nominal solution DVHs for 52 realized pdfs in experiment #2. 63
2-16 Robust solution DVHs for 52 realized pdfs in experiment #2. 64
2-17 Margin solution DVHs for 52 realized pdfs in experiment #2. 64

2-18 The nominal, robust, and margin solution DVHs for one realized pdf in experiment #2. 65

2-19 An approximation of the Pareto surface generated by the robust formulation. 66

3-1 Dose delivered using a margin and scaling. 72

3-2 The ratio of the optimal margin size to σ as a function of the ratio of the tumor size to σ for a 1D tumor. 82

3-3 A two-dimensional slice of the three-dimensional static intensity map. 91

3-4 The ratio of the optimal margin size to σ as a function of the ratio of tumor size to σ for a 3D tumor. 97

3-5 The function $h_\alpha(u)$ for different values of α. 103

3-6 The ratio of the optimal margin size to σ as a function of the ratio of tumor size to σ for a 3D tumor, for different values of α. 106

3-7 An edge-enhanced intensity map. 107

3-8 The ratio of the optimal margin size to σ in an edge-enhanced intensity map as a function of the ratio of the tumor size to σ for a 1D tumor. 133

3-9 The ratio of the optimal l^* to σ in an edge-enhanced intensity map as a function of the ratio of the tumor size to σ for a 1D tumor. 134

3-10 The ratio of the optimal h^* to σ in an edge-enhanced intensity map as a function of the ratio of the tumor size to σ for a 1D tumor. 134

3-11 Dose delivered as a function of $\tilde{\sigma}$ for $t = 4, m = 1, l = 0.2, h = 2$. . . . 137

3-12 Dose delivered as a function of $\tilde{\sigma}$ for $t = 4, m = 5, l = 0.2, h = 2$. . . . 137

4-1 The depth-dose curve for a photon and proton. 143

4-2 The RTOG benchmark phantom and the beam directions used. . . . 147

4-3 The nominal treatment plan. (a) Dose distribution assuming each beamlet realizes its nominal range; (c/e) Dose distribution assuming each beamlet's range is shortened/lengthened by 5 mm; (b/d) Dose distribution of the individual beam at 45°/0° assuming each beamlet realizes its nominal range; (f) The maximum deviation of the delivered dose from the prescribed dose over all 11 ranges. 168
4-4 The robust treatment plan. (a) Dose distribution assuming each beamlet realizes its nominal range; (c/e) Dose distribution assuming each beamlet's range is shortened/lengthened by 5 mm; (b/d) Dose distribution of the individual beam at 45°/0° assuming each beamlet realizes its nominal range; (f) The maximum deviation of the delivered dose from the prescribed dose over all 11 ranges. 169
4-5 DVHs of the nominal (a) and robust (b) solution under three scenarios: all beamlets realize their nominal range, shortest range, or longest range. 170
4-6 DVHs of the nominal and robust solutions under three scenarios: all beamlets realize their nominal range (a), shortest range (b), or longest range (c). 171

List of Tables

2.1 RPM data used in the clinical lung case. 44

2.2 Comparing formulations on integral dose delivered to phantom and integral dose delivered to normal tissue under the realized pdf. The relative amount of dose delivered (%) by the various formulations is normalized to the margin formulation. 55

2.3 Basic information regarding the three formulations. 58

2.4 Comparing the nominal, robust, and margin formulations using the average of the 78 trials in experiment #1. This experiment showed the least difference between the nominal and robust solutions in terms of tumor coverage. The left subcolumns represent values in Gy, while the right subcolumns are percentages. The first row percentages are relative to 72 Gy, which is the desired dose level, while the next two rows are quoted as a percentage of the dose delivered by the margin solution. 62

2.5 Comparing the nominal, robust, and margin formulations using the average of 52 pdfs in experiment #2. This experiment showed the greatest difference between the nominal and robust solutions in terms of tumor coverage. The left subcolumns represent values in Gy, while the right subcolumns are percentages. The first row percentages are relative to 72 Gy, which is the desired dose level, while the next two rows are quoted as a percentage of the dose delivered by the margin solution. 65

3.1 The optimal margin size (m^*) for particular values of the tumor size (t), relative to the size of the standard deviation of the Gaussian pdf ($\sigma = 1$ in this case). The corresponding intensity scaling factors ($\gamma^*_{m^*}$) are also shown. 81

3.2 The optimal margin size (m^*) for particular values of the tumor size (t), relative to the size of the standard deviation of the Gaussian pdf ($\sigma = 1$ in this case), in the 3D nominal problem. The corresponding intensity scaling factors ($\gamma^*_{m^*}$) are also shown. 96

3.3 The relationship between α (the multiplicative factor in the relation $t = \alpha\hat{\mu}_1$) and m^* (the optimal margin size). A check (✓) indicates that the outcome is possible, while a cross ($\times$) indicates that the outcome is not possible. 105

3.4 The optimal l^* and h^* for particular values of the tumor size, relative to the size of the standard deviation of the Gaussian pdf. 130

4.1 Formulation (4.6). 155

4.2 Formulation (4.14). 156

4.3 Basic information regarding the nominal and robust formulations. . . 161

Chapter 1

Introduction

"I'm not sure."

– Werner Heisenberg

A famous quote states "The only things that are certain in life are death and taxes." However, uncertainty is also a certainty. In the context of treatments of life-threatening diseases, uncertainty is a challenge that needs to be properly managed. In this thesis, we consider the use of radiation therapy to treat cancer, and more specifically, the effects that uncertainty has on the quality of the resulting treatment.

There are many types of cancer treatments. Surgery is used to excise tumors, or parts of tumors, from specific locations directly using a knife. Chemotherapy uses chemical substances to treat cancer indirectly, since these drugs typically target all rapidly dividing cells, which include cancer cells as well as innocuous ones such as cells responsible for hair growth. Radiation therapy uses radiation to kill tumor cells. The mechanism for this type of cell death involves the radiation doing permanent damage to the DNA of the tumor cells. However, radiation will also damage healthy cells. Hence, the goal of a radiation therapy treatment is to deliver sufficient dose to the tumor, while ensuring that the healthy tissue around the tumor is spared as much as possible.

Two common types of radiation therapy are external beam radiation therapy and brachytherapy. In this thesis, we focus on the former, but let us briefly describe the

latter. Brachytherapy is a method that involves implanting small radioactive seeds into the patient (in and around a tumor site) thereby localizing the delivered radiation as much as possible. As an optimization problem, the location of these seeds and/or their duration at those positions are the typical decision variables [1, 49].

In external beam radiation therapy, radiation is delivered from outside the body using a linear accelerator. The agent of traditional external beam radiation therapy is a stream of photons, however, advanced treatments using protons are becoming more common. Our work in this thesis spans both of these treatment modalities: we discuss photon-based treatments in Chapters 2 and 3, and proton-based treatments in Chapter 4. Subsequent references to "radiation therapy" in this thesis refer to external beam treatments using photons. Chapter 4 is the only place where we focus specifically on protons.

Since radiation damages healthy cells, external beam treatments aim to deliver radiation from many different angles around the patient. The beams of radiation are arranged in such a way that the tumor lies within their intersection, while the surrounding healthy tissue "sees" as little radiation as possible. A typical radiation therapy treatment is carried out over the course of four to six weeks, on an approximately daily basis. This property of the treatment is known as "fractionation," and a "fraction" refers to one treatment session. For example, over a six week time horizon, a treatment may span 30 fractions. During a fractionated treatment over n days, roughly "$1/n$" of the treatment is delivered during each fraction. That is, the total amount of radiation that is to be delivered is split into smaller portions that are administered over the entire treatment horizon. This leverages the "therapeutic advantage" that healthy cells have over cancerous ones. Cancer cells do not possess the capable and robust repair mechanism that healthy cells have. Therefore, by spreading out the treatment over many days, healthy cells have a chance to recover between treatment sessions.

In order to construct a radiation therapy treatment, the starting point is a picture of the internal anatomy. This is obtained using computed tomography (CT). Once CT scans are taken, a physician will delineate (on a computer) the target and critical

structures. Critical structures are healthy organs, such as the spinal cord or prostate, that need to be spared as much as possible, and have higher priority than other normal, non-cancerous tissue. Along with the outlines of these structures, the physician will also provide rough objectives and constraints to the treatment planner, who will then use all of this information to create a treatment plan. The treatment planner works with software that uses an optimization routine to find a treatment plan that best satisfies the physician's requirements. Once the doctor and planner converge on a treatment, the treatment is then delivered over a suitable time horizon. Currently, during treatment, new data is rarely re-acquired due to resource constraints. However, future technological advances will provide direct, visual feedback during treatments, allowing treatments to become more dynamic.

The concept of optimization plays a significant role in the calculation of radiation therapy treatment plans. There are many decision variables that can be adjusted in order to arrive at a suitable plan. The main variables that are typically considered are the angles from which to deliver the radiation, and the intensity (or how long the beam is turned "on") of the radiation beams. Beam angle optimization has been studied using heuristic [9], integer programming [27], and linear programming [15] approaches. However, the nonconvex nature of the problem [50] makes it difficult to solve in practice. Therefore, a treatment planner typically chooses beam directions manually based on past experience and the current patient's situation, and then optimizes the intensities of the beams, given these pre-determined orientations. This is the approach we take in this thesis. Our focus will be the optimization of beam intensities, assuming that the beam directions are given. In particular, we focus on a technique known as intensity-modulated radiation therapy (IMRT), which is described more thoroughly in the next chapter. Basically, in IMRT, a beam of radiation is divided into many small beamlets, each with individually adjustable intensities. This results in many more optimization variables than non-intensity-modulated treatments. The resulting solution is known as an "intensity map," since it resembles a topological map where the height of a point on the map matches the intensity of the corresponding beamlet.

The topic of beam-based or beamlet-based optimization in radiation therapy has

received a lot of attention from the optimization community. Examples are [45, 20] for linear, [17] for nonlinear, and [17, 28, 43] for mixed integer programming approaches. The references [48, 44] provide comprehensive reviews of the existing literature. Interestingly, the first appearance of linear programming applied to radiation therapy treatment planning in the literature occurred in the 1960's [25, 2]. Nowadays, radiation therapy is opening new doors for researchers from the optimization community to develop models and algorithms for practical problems in the field of health care.

Recently, much effort has been devoted to incorporating treatment uncertainties in a sophisticated manner into the underlying optimization problem. In practice, the concept of a "margin" has existed for a long time, with many different definitions to address different types of uncertainties. A margin is an artificial, volumetric expansion around the tumor, which is treated as if it is an extension of the tumor. The Gross Tumor Volume (GTV) describes the tumor mass that is visible on the CT scan. To protect for tumor growth that may not be visible on the scan, a margin is added around the GTV. The Clinical Target Volume (CTV) is a superset of the GTV that includes this margin. The Internal Target Volume (ITV) consists of the union of the CTV in various positions due to internal motion (such as breathing-induced motion). Finally, the Planning Target Volume (PTV) is a superset of the ITV that includes an additional margin to account for other uncertainties such as setup error (the patient may be set up on the treatment table slightly differently from fraction to fraction). The anatomy contained within the PTV is what is finally irradiated to the level that the tumor is scheduled to receive. From its construction, it should be clear that a large fraction of the PTV may contain healthy cells. This motivates "margin reduction" research, which aims to reduce the margins around the tumor, while ensuring the tumor receives adequate dose in the face of uncertainty. Our initial research in this area is presented in [13], which included results on a one-dimensional problem. The details of this work, along with further studies on a clinical three-dimensional lung case, are presented in the next chapter.

The goal of introducing a robust optimization framework in radiation therapy treatment planning is to enable a treatment planner to deliver treatments that are

more aggressive than traditional margin approaches, without sacrificing protection from uncertainty. As a preview, the robust formulation in Chapter 2 actually contains the ITV margin approach as a special case. Therefore, it will only produce a margin when appropriate, and otherwise, it will replace a margin-like solution with a more aggressive one. Other recent research in the area of robust optimization applied to radiation therapy can be found in [14, 39]. In these papers, the authors consider different types of uncertainty (e.g. patient positioning uncertainty) than we do, which results in different robust formulations (e.g. second-order cone programs).

Overall, the presence of uncertainty significantly affects the quality of a treatment. Our goals in this thesis are to derive robust radiation therapy treatments that effectively manage uncertainty, and to understand the effects of uncertainty on the structure of our solutions. The structure of the thesis is as follows.

In Chapter 2 we derive a linear robust formulation for the treatment of tumors that are subject to motion uncertainty. In particular, we focus on uncertainty induced by irregular breathing motion, and its effect on the treatment of a lung tumor. We implement our robust approach on both an idealized patient geometry, as well as a full three-dimensional clinical case. We compare the performance of our robust solution with other common formulations for this problem. Finally, we offer a multi-objective optimization viewpoint of our robust approach.

In Chapter 3, motivated by the structure of the robust solution from Chapter 2, we conduct a mathematical analysis of the structure of optimal intensity maps under uncertainty. We consider a motion distribution that is Gaussian, and analyze the shape of the optimal intensity maps for this motion. Then, we extend this work to the robust setting by considering the effect of uncertainty in the mean and variance of the Gaussian on the resulting optimized solution.

In Chapter 4, we consider radiation therapy using protons, as opposed to photons, which were the focus of the previous two chapters. We derive a linear robust formulation where the uncertainty is related to the location of the maximum dose deposited by the stream of protons. As in Chapter 2, the optimal robust solution has a very specific and intuitive structure. We demonstrate the advantages of such an approach

over a non-robust approach, using a standard patient geometry.

We view our contributions as follows:

1. We present the first robust optimization approach to IMRT treatments subject to breathing motion uncertainty. This formulation retains tractability since the robust counterpart, like its nominal predecessor, is a linear program. Our robust framework generalizes state-of-the-art formulations that are used in both theory and practice for the management of motion (and motion uncertainty) during treatments, and can be extended to mitigate other types of uncertainty.

2. We provide mathematical and empirical evidence that the structure of optimal robust solutions satisfies certain properties that are intuitive to practitioners from the medical physics community. We conduct an analytical evaluation of the structure of certain margin and intensity-modulated intensity maps, and derive novel bounds on the relevant parameters that classify the solutions according to their structural properties.

3. We present the first robust optimization formulation for intensity-modulated proton therapy under uncertainty of any sort. We produce intuitive solutions that make explicit use of the physical properties of the proton beams to manage uncertainty without being overly conservative.

Chapter 2

Robust Intensity-Modulated Radiation Therapy

> "Clearly any real body must have extension in *four* directions: it must have Length, Breadth, Thickness, and–Duration."
>
> – H. G. Wells, *The Time Machine*

In this chapter, we study the use of intensity-modulated radiation therapy to treat moving lung tumors, when there is uncertainty in the motion itself. Even if we have a description of the tumor motion during the planning stages of the treatment, the realized motion during treatment delivery may be quite different than what was previously measured. This motivates us to consider a robust approach to managing motion uncertainty. Determining an acceptable treatment plan is primarily about balancing various inherent trade-offs. In this case, we aim to strike a balance between the sparing of the healthy tissue and the effective treatment of the tumor, given the existence of uncertainty.

Recent research has highlighted the topic of robustness in the optimization of beamlet intensities in external beam radiation therapy. Both [3, 56] use probabilities in the objective function to generate solutions that are less sensitive to interfraction (between fractions) uncertainty. The combination of the need for more rigorous optimization approaches to deal with clinical uncertainties, together with advances in

deriving tractable robust counterparts for uncertain mathematical programs [4, 5, 6], has led to a fair amount of recent interest in the use of robust optimization methods in radiation therapy treatment planning. In [14], the authors modeled uncertainty in the dose delivered due to shifts of the anatomy over many fractions. The robust counterpart of their nominal linear program was a second-order cone program. The work presented in [39] also considered interfraction positional uncertainty, but added uncertainty in the dose calculation (how much dose is deposited at a point in the body from the beams of radiation) as well. Their robust counterpart was also a second-order cone program. Our research differs primarily in the type of uncertainty considered (intrafraction motion uncertainty; that is, motion within a fraction) and the complexity of the resulting robust counterpart (a linear program).

Our overall goals in this chapter are to understand the effects of motion uncertainty on the quality of a treatment, and to derive a framework that creates solutions which are insensitive to this uncertainty. We propose a novel robust optimization formulation that focuses specifically on motion uncertainty and we demonstrate its effectiveness with clinical data.

The rest of this chapter is organized as follows. Section 2.1 introduces IMRT optimization and Section 2.2 provides background on motion effects in the treatment of lung cancer. Section 2.3 describes our model of motion uncertainty. In Section 2.4, we develop our robust formulation and present other formulations that will be compared to the robust formulation. Section 2.5 describes the motion data that is used in the optimization formulations. In Sections 2.6 and 2.7, we illustrate the results obtained from our robust formulation and compare them to alternative approaches on both a simple 1D geometry and a more complex 3D clinical lung case, respectively. In Section 2.8, we discuss trade-offs from a Pareto perspective and show how the robust formulation effectively addresses this topic. Finally, we present our conclusions in Section 2.9.

2.1 IMRT Background

In recent years, intensity-modulated radiation therapy (IMRT) has garnered much research and clinical attention, and is now in clinical use at most radiation oncology centers in the United States [36]. Unlike traditional conformal radiation therapy where the beam of radiation has a uniform intensity, we can think of IMRT as partitioning each beam of radiation into a large set of "beamlets" (e.g. 5 mm $\times$ 5 mm in size), which have individually adjustable intensities; these beamlet intensities are the decision variables in our underlying optimization problem. The flexibility to deliver intensity that is non-uniform across the beam can enhance the conformity of the dose delivered to the tumor and the sparing of healthy tissue, especially for complex shapes of the target volume. The calculation of the "intensity maps" for each beam is generally accomplished with "inverse" treatment planning systems that employ optimization approaches. See [22, 59, 10] for an overview of IMRT technology.

The delivery of intensity modulated beams has been accomplished in a number of ways, but the multileaf collimator (MLC) is the most commonly used apparatus [22]. The MLC is a mechanical device attached to the head of the beam. It is comprised of two sets of "leaves" that resemble teeth, and can shape the radiation exiting the linear accelerator by positioning the leaves to block some parts of the beam. Changing the shape of the beam and varying the time that the beam is on will ultimately affect the profile of the intensity map that is delivered. For example, parts of the beam that were blocked for a greater fraction of the time will ultimately deliver a lower intensity over the total time the beam is on.

Intensity-modulated radiation therapy is a valuable technique since it is able to produce steeper dose gradients (how quickly the dose falls off from areas of high dose to areas of low dose) as compared to conformal radiation therapy. However, these steeper dose gradients also make this type of therapy more vulnerable to uncertainties. In the next section, we discuss some general uncertainties, and then focus on motion specifically.

2.2 Motion uncertainty in lung tumors

There are many sources of uncertainty that need to be taken into account in the course of the treatment planning process. The segmentation of the tumor and the critical structures based on medical images is inherently uncertain and error prone. Furthermore, patient positioning uncertainties are relevant because the patient needs to be set up in the same position every day over the course of the treatment. Motion effects constitute another class of uncertainty. In particular, there are two types of motion: interfraction and intrafraction motion. Interfraction motion refers to motion between treatment sessions (also known as "fractions"). An example of this type of motion comes from the variations in the location of a prostate tumor due to differences in rectal and bladder filling from day to day. Intrafraction motion refers to motion that occurs during a treatment session, such as breathing motion.

This work draws motivation from lung cancer, and we will focus on breathing motion as the source of uncertainty. As a patient breathes during treatment, a tumor in the lung will be moving (possibly irregularly), and we would like to find a method that reliably delivers a sufficient amount of dose to the tumor, while sparing the healthy tissue as much as possible. A detailed review on the topic of motion in radiation therapy is given in [26]. The dosimetric effects of respiratory motion in IMRT were studied in [8], and it was found that, to first order, the effects can be described as a "blurring" of the spatial dose distribution. In particular, steep dose gradients are washed out due to the motion, whereas secondary effects such as the interplay between breathing motion and the motion of the delivery device (MLC) can often be neglected. An important consequence of dose blurring due to motion is that, without compensation, motion leads to regions of underdose ("cold spots") near the edge of the target volume. Mathematically, the dose blurring effect can be modeled as a convolution of the static spatial dose distribution with a probability density function (pdf) that describes the motion [34, 16, 11]. This pdf specifies the relative amount of time the tumor (and other tissue) spends in different locations or, in an alternative interpretation, the relative amount of time spent in different phases

of the breathing cycle.

A number of strategies have been proposed to reduce the effects of motion ([24] provides a comprehensive review). The methods fall into two main categories: (i) reduction of the amplitude of the motion relative to the treatment device (e.g., breath hold, gating, tumor tracking), and (ii) reduction of the dosimetric consequences of the motion. Breath hold is self-explanatory. The gating method turns off the beam when the tumor leaves a specified window during the motion cycle, with the aim of delivering radiation only when the tumor is close to a certain position. Tumor tracking involves tracking the tumor movement in real time with the beam of radiation. The methods in (i) can be somewhat unreliable and difficult for the patient to tolerate (breath hold), technologically demanding (gating and tracking), or may prolong the treatment time (breath hold and gating). The most common method, which belongs to the latter category (ii) is the use of a "margin", which is an expanded region around the nominal position of the tumor in which we are confident the tumor will remain. Then, by treating this margin area with a high uniform dose, we ensure that the tumor receives a sufficient amount of dose, no matter how it moves within this margin. Of course, the main disadvantage of using a margin is that the healthy tissue surrounding the tumor will receive more dose than necessary. We can think of the use of a margin as a worst-case approach, in which we use some bounds on the motion, but assume no additional information on how the tumor will move within these bounds. We show later that the margin is indeed the correct approach in this case, and define more precisely the worst-case that the margin is protecting against.

A recent development that also belongs to category (ii) is that of "4D" optimization, which aims at undoing (de-convolving) the blurring effect of the motion. The basic idea is to hit the tumor edges harder (with higher intensities), and thus compensate for the blurring effect and possible dose cold spots [54, 62]. These recent studies showed promising results but relied on the assumption that the breathing motion realized during every treatment fraction was exactly the same as the motion exhibited during the planning session. Hence, this approach assumes "perfect information" or "no uncertainty" on the motion pdf, and is appropriate if we can be certain that the

patient will breathe in the same, predictable way during each fraction. On the other hand, it was shown in [47] that the quality of the resulting dose distribution can be seriously degraded if this assumption is violated. Furthermore, the subject of changes in respiratory patterns between treatment simulation and treatment delivery was the very first in a list of topics recommended for further research, in a recent report on the management of respiratory motion in radiation oncology by a group of experts from the American Association of Physicists in Medicine [24].

Our work in this chapter strives to generalize and improve upon the two approaches we mentioned in category (ii). We build off the motion pdf approach by relaxing the rather strong assumption of highly regular and reproducible breathing. In particular, we develop a model of uncertainty in the motion pdf, and incorporate it into a robust optimization framework that can produce treatment plans that are robust to this uncertainty; that is, plans that are guaranteed to be feasible under any realization of the uncertainty. It is important to distinguish between uncertainty *due* to motion and uncertainty *in* the motion itself. If the motion is perfectly regular and predictable, then we say that there might be uncertainty due to motion. However, this uncertainty can be eliminated by observing the motion and building the corresponding pdf. Since the motion is predictable in this case, knowing the pdf before treatment essentially eliminates the uncertainty *due* to motion in subsequent fractions. On the other hand, our work focuses on uncertainty *in* the motion itself, and moreover, considers variations in the motion from the time data is acquired to the time that the patient is treated. So even if we have a pdf that perfectly describes the motion prior to treatment, there is no guarantee that this will be the same pdf that is realized during treatment. This is the uncertainty to which we would like to be robust.

2.3 Model of uncertainty

The presence of motion during radiation delivery spreads out the dose to areas surrounding the target region. In order to compensate for this blurring effect, the "mo-

tion pdf" approach aims to produce a dose distribution which, when convolved with the assumed pdf, recovers the desired distribution. Naturally, this method relies on *a priori* knowledge of what the motion pdf will be during treatment. If the realized motion is quite different from what was assumed, then convolving the realized motion pdf with a dose distribution optimized for a different pdf is likely to result in an unacceptable dose distribution with hot and cold spots. This motivates the need for an approach that can mitigate the uncertainty in the pdf to be realized during treatment.

Our framework involves a finite set X, which corresponds to the set of possible phases of the breathing cycle. A *(motion) pdf* is defined as a nonnegative real function $f : X \to \mathbb{R}$, such that $\sum_{x \in X} f(x) = 1$. Our starting point is a *nominal pdf*, denoted by p, which is constructed on the basis of data taken during the planning session. Regarding the actual pdf, $\tilde{p}$, to be realized during treatment, we assume that it can differ from the nominal pdf, p, on a subset U of the domain X, and satisfies

$$p(x) - \underline{p}(x) \leq \tilde{p}(x) \leq p(x) + \overline{p}(x), \qquad \forall\, x \in U,$$

where the functions $\underline{p}$ and $\overline{p}$ provide bounds on the difference between the nominal and the realized pdf during treatment. Without loss of generality, we assume that $p - \underline{p} \geq 0$ and $p + \overline{p} \leq 1$. Thus, the set P_U, of possible realized pdfs, called the *uncertainty set*, is defined as

$$\begin{aligned} P_U = \big\{ \tilde{p} \in \mathbb{R}^{|X|} \ : \tilde{p}(x) \in [p(x) - \underline{p}(x), p(x) + \overline{p}(x)],\ \forall\, x \in U; \\ \tilde{p}(x) = p(x),\ \forall\, x \in X \setminus U;\ \sum_{x \in X} \tilde{p}(x) = 1 \big\}. \end{aligned} \tag{2.1}$$

Note that the set U is somewhat redundant, since it can be accounted for by setting $\overline{p}(x) = \underline{p}(x) = 0$ for $x \in X \setminus U$, but we will keep it for clarity of exposition. We will refer to the upper and lower bounds on $\tilde{p}$ as "error bars."

A treatment plan will be said to be *robust* if all of the constraints in our subsequent formulation are satisfied (cf. Section 2.4), no matter which pdf from the set P_U is

realized. Thus, P_U is to be interpreted as the set of all pdfs that we are protecting against in our optimization. If one views this approach as being overly conservative (since it permits certain implausible, highly oscillatory pdfs), one can introduce simple (linear) "smoothness" constraints into the definition of P_U, such as

$$|\tilde{p}(x) - \tilde{p}(y)| \leq \epsilon, \quad \text{if } |x - y| \leq \delta,$$

for suitable $\epsilon, \delta > 0$.

One of the main challenges here is to keep the set P_U large enough so that we can protect against realistic variations that the patient may exhibit, without losing all information about the patient's particular breathing patterns, and just delivering a margin. The practical construction of this set from real data is discussed in Section 2.5.1.

2.4 Formulations

This section introduces the different formulations that will be investigated. In our formulations, the goal is to minimize the total dose delivered, while ensuring the tumor receives an adequate dose. In Section 2.4.1, we start with the basic problem of finding the optimal intensities of the beamlets in the IMRT planning problem when there is no motion and no uncertainty. We then continue with a nominal formulation (a linear program) that takes into account a nominal motion pdf. In Section 2.4.2, we present a robust formulation, in which the possible tumor motion is modeled by the uncertainty set P_U, and demonstrate that it can also be cast as a linear program. Finally, in Section 2.4.3, we describe the classical margin approach and explore its relation with our robust formulation.

2.4.1 Nominal formulation

We assume that the treatment will be delivered using beamlets belonging to a given set $\mathcal{B}$, and we let w_b be the weight or intensity of beamlet b (related to the amount of

time the beamlet is "on"). We are interested in a *phantom*, represented by a set $\mathcal{V}$ of *voxels*. A voxel is a small volume-element used to represent a specific location in the patient. A phantom is a physical model of the human body or any of its parts, and here it is synonymous with the set of voxels that represent the tumor and all healthy tissue. Let $\mathcal{T}$ be the set of voxels in the tumor, and let $\mathcal{N}$ be the set of voxels in the normal (non-tumor) tissue. Furthermore, let $m_{\mathcal{T}} = |\mathcal{T}|$, $n = |\mathcal{B}|$, and $m_U = |U|$. Let $D_{v,b}$ be the dose that voxel v receives per unit of intensity of beamlet b, and let θ_v be the prescribed dose that voxel v should receive. The following formulation aims at minimizing the total delivered dose to the healthy tissue, in the absence of motion, and will be referred to as the *basic problem*.

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v \in \mathcal{N}} \sum_{b \in \mathcal{B}} D_{v,b} w_b \\ \text{subject to} \quad & \sum_{b \in \mathcal{B}} D_{v,b} w_b \geq \theta_v, \quad \forall\, v \in \mathcal{T}, \\ & w_b \geq 0, \quad \forall\, b \in \mathcal{B}. \end{aligned} \tag{2.2}$$

In practice, the prescribed dose is often uniform over the tumor, in which case we have $\theta_v = \theta$ for all $v \in \mathcal{T}$.

The next step is to incorporate motion (but not yet motion uncertainty) into (2.2), using the nominal pdf, p, described in Section 2.3. In the presence of motion, we need to fix some nominal reference frame in relation to which the voxels will be moving. For example, this reference frame could correspond to the position of the internal anatomy when a patient is in the full exhale phase of breathing. We also need to adjust the "matrix" D, in order to account for the motion. For that purpose, we create a "matrix" Δ, with components $\Delta_{v,x,b}$ that describe the dose delivered to voxel v, when the patient is in breathing phase x, per unit intensity of beamlet b. Without loss of generality, we can assume that the nominal reference frame corresponds to breathing phase 0 and the position of the anatomy in the basic problem (i.e., in the absence of motion). Hence, $\Delta_{v,x,b} = D_{v,b}$ when $x = 0$. With this picture in mind, $p(x)$ is the probability that the patient is in breathing phase x. This definition facilitates

the incorporation of non-rigid-body motion into the problem formulation, since $\Delta_{v,x,b}$ is computed for snapshots of the anatomy in each phase (as opposed to extrapolating based on rigid-body motion). Note that rigid-body motion can be cast as a special case of this definition, where x corresponds to displacement from a nominal position so that, again, $x = 0$ corresponds to the nominal position.

Adjusting for motion, as described by the nominal pdf, we obtain the following *nominal problem*.

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{N}}\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \\ \text{subject to} \quad & \sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \geq \theta_v, \quad \forall\, v\in\mathcal{T}, \\ & w_b \geq 0, \quad \forall\, b\in\mathcal{B}. \end{aligned} \tag{2.3}$$

This formulation is a straightforward extension of (2.2), in which the coefficients $D_{v,b}$, are replaced with

$$\sum_{x\in X}\Delta_{v,x,b}p(x)$$

in order to account for the motion as described by p. Note that the pdf used in the objective function and constraints does not correspond to just one fraction, but refers to the average pdf over the entire course of treatment. While it is true that during one fraction we do not expect the dose delivered to be well-approximated by the expression in the objective function, we do expect convergence of the delivered dose to the expected delivered dose (represented by the pdf p) after many treatment fractions [11].

Formulation (2.3) is the appropriate formulation when there is "perfect information," in the sense that the breathing-induced motion is accurately represented by p. However, if we use this formulation and the patient breathes according to a different pdf, the lower bound constraints on the tumor dose are likely to be violated. This motivates the inclusion of the model of uncertainty, as described in Section 2.3, into formulation (2.3).

2.4.2 Robust formulation

In the presence of uncertainty, we want the tumor to receive the required dose no matter which pdf within the uncertainty set is realized during treatment. This desire is captured by requiring the lower bound constraint in formulation (2.3) to be satisfied when p is replaced by $\tilde{p}$, for every $\tilde{p} \in P_U$. This leads to the following *robust problem*.

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{N}}\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \\ \text{subject to} \quad & \sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}\tilde{p}(x)w_b \geq \theta_v, \quad \forall\, v\in\mathcal{T},\ \forall\,\tilde{p}\in P_U, \\ & w_b \geq 0, \quad \forall\, b\in\mathcal{B}. \end{aligned} \tag{2.4}$$

Note that the uncertainty is in the pdf, and not in the Δ matrix. This corresponds to the assumption that the possible geometrical configurations of the anatomy (i.e., path of the motion trajectory), due to motion, do not change between treatment planning and treatment delivery, but that the fraction of time spent in each configuration (i.e., breathing phase or location on the trajectory) is uncertain.

Furthermore, notice that the uncertainty in p is accounted for in the constraints and not in the objective. To gain some understanding of the implications, note that one could just minimize the total dose delivered to the anatomy in a particular phase x. However, the choice of x would have been arbitrary and subjective. Instead, our goal is to use a reasonable objective function that approximates well the total dose delivered to the healthy tissue. By using p in the objective, we are essentially averaging the dose delivered over the various phases, since this objective function is a convex combination of the corresponding objective functions associated with individual phases. Different p in the objective will indeed redistribute the dose between voxels, but will not change the integral dose appreciably, since all pdfs integrate to one.

There are other types of objective functions that would require accounting for the uncertainty in the pdf. For example, if we are interested in an objective such as minimizing the maximum dose delivered to the healthy tissue, we would replace the

objective with an auxiliary variable z, and introduce to formulation (2.4) additional constraints, of the form

$$z \geq \sum_{b\in\mathcal{B}} \sum_{x\in X} \Delta_{v,x,b}\tilde{p}(x)w_b, \quad \forall\, v \in \mathcal{N},\ \forall\, \tilde{p} \in P_U. \tag{2.5}$$

These constraints have essentially the same structure as the lower bound constraints in the robust problem, and the analysis provided in the rest of this section (i.e., the derivation of a finite LP) still applies to formulation (2.4) augmented with constraints (2.5).

Formulation (2.4) is not a linear program due to the infinite number of constraints. In what follows, we transform (2.4) into an equivalent linear program, and also develop some insights on its structure. It should be clear that (2.4) will become a linear program since the uncertainty set, P_U, is a polyhedron.

Proposition 2.1. *Formulation* (2.4) *is equivalent to*

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{N}} \sum_{b\in\mathcal{B}} \sum_{x\in X} \Delta_{v,x,b}p(x)w_b \\ \text{subject to} \quad & \sum_{b\in\mathcal{B}} \sum_{x\in X} \Delta_{v,x,b}p(x)w_b + \beta_v(\boldsymbol{w}) \geq \theta_v, \quad \forall\, v \in \mathcal{T}, \\ & w_b \geq 0, \quad \forall\, b \in \mathcal{B}. \end{aligned} \tag{2.6}$$

where

$$\begin{aligned} \beta_v(\boldsymbol{w}) = \min_{\hat{p}} \quad & \sum_{b\in\mathcal{B}} \sum_{x\in U} \Delta_{v,x,b}\hat{p}(x)w_b \\ \text{subject to} \quad & \sum_{x\in U} \hat{p}(x) = 0, \\ & -\underline{p}(x) \leq \hat{p}(x) \leq \overline{p}(x), \quad \forall\, x \in U. \end{aligned} \tag{2.7}$$

Proof: Any $\tilde{p} \in P_U$ can be written in the form $\tilde{p} = p + \hat{p}$, where p is the nominal pdf, $\hat{p} \in \hat{P}_U$, and

$$\hat{P}_U = \Big\{\hat{p} \in \mathbb{R}^{|X|} \ : \hat{p}(x) \in [-\underline{p}(x), \overline{p}(x)],\ \forall\, x \in U; \\ \hat{p}(x) = 0,\ \forall\, x \in X \setminus U;\ \sum_{x\in X} \hat{p}(x) = 0\Big\}.$$

Then, for any $v \in \mathcal{T}$, the corresponding constraint in (2.4) can be rewritten as

$$\min_{\tilde{p} \in P_U} \sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} \tilde{p}(x) w_b \geq \theta_v,$$

or

$$\sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} p(x) w_b + \min_{\hat{p} \in \hat{P}_U} \sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} \hat{p}(x) w_b \geq \theta_v.$$

□

The quantity $\beta_v(\boldsymbol{w})$ arises because we are protecting against worst-case (in terms of underdosing the tumor) motion within our uncertainty set P_U. To better understand the robust formulation, we now investigate the structure of the solution of problem (2.7).

Proposition 2.2. *Let*

$$d_{v,x}(\boldsymbol{w}) := \sum_{b \in \mathcal{B}} \Delta_{v,x,b} w_b.$$

For any v and $\boldsymbol{w}$, we introduce an ordering $x(1), \ldots, x(m_U)$ of the elements of X such that $d_{v,x(1)}(\boldsymbol{w}) \leq \cdots \leq d_{v,x(m_U)}(\boldsymbol{w})$. Then, an optimal solution to (2.7) is of the form

$$\hat{p}^*(x(j)) = \begin{cases} \overline{p}(x(j)), & j < j^*, \\ \sum_{j > j^*} \underline{p}(x(j)) - \sum_{j < j^*} \overline{p}(x(j)), & j = j^*, \\ -\underline{p}(x(j)), & j > j^*, \end{cases}$$

where j^* satisfies the inequalities

$$\sum_{j \geq j^*} \underline{p}(x(j)) - \sum_{j < j^*} \overline{p}(x(j)) \geq 0,$$
$$\sum_{j \geq j^*+1} \underline{p}(x(j)) - \sum_{j < j^*+1} \overline{p}(x(j)) < 0.$$

Proof: Let λ be the optimal dual variable corresponding to the equality constraint

in (2.7). Then there exists a $\hat{p}$ that is an optimal solution to

$$\begin{aligned} \underset{\hat{p}}{\text{minimize}} \quad & \sum_{x \in U} \hat{p}(x)(d_{v,x}(\boldsymbol{w}) - \lambda) \\ \text{subject to} \quad & -\underline{p}(x) \leq \hat{p}(x) \leq \overline{p}(x), \quad \forall\, x \in U, \end{aligned} \tag{2.8}$$

that is also an optimal solution for the original problem (2.7). In such an optimal solution, $\hat{p}(x)$ must be set to its lower (respectively, upper) bound $-\underline{p}(x)$ (respectively, $\overline{p}(x)$), for every x for which $d_{v,x}(\boldsymbol{w}) > \lambda$ (respectively, $d_{v,x}(\boldsymbol{w}) < \lambda$). □

Corollary 2.1. *The optimal value of* (2.7) *is equal to*

$$\begin{aligned} \beta_v(\boldsymbol{w}) = & -\sum_{b \in \mathcal{B}} \sum_{j > j^*} (\Delta_{v,x(j),b} - \Delta_{v,x(j^*),b}) \underline{p}(x(j)) w_b \\ & -\sum_{b \in \mathcal{B}} \sum_{j < j^*} (\Delta_{v,x(j^*),b} - \Delta_{v,x(j),b}) \overline{p}(x(j)) w_b. \end{aligned} \tag{2.9}$$

The expression for $\beta_v(\boldsymbol{w})$ in (2.9) gives insight into the structure of the robust dose distribution generated by the solution to the robust formulation (2.4). Let $\boldsymbol{w}$ be an optimal solution to the nominal problem and consider the resulting dose delivered

$$\sum_{b \in \mathcal{B}} D_{v,b} w_b,$$

not to a particular anatomical voxel, but to the point in space where voxel v resides in its nominal position; this is the static dose distribution corresponding to the basic problem. Naturally, there may be regions of high static dose and regions of low static dose. Now imagine that a particular voxel v moves within this dose distribution, due to breathing motion. The solution $\boldsymbol{w}$ assumes that v will move according to the nominal p, "acquiring" just the right amount of dose from both the high and low dose parts of the static dose distribution. However, if v moves according to a different pdf, we must protect against the scenario where v spends more time in the low dose locations and less time in the high dose locations than nominally assumed. By protecting against this situation, which is exactly what $\beta_v(\boldsymbol{w})$ represents in (2.9),

we are ensuring feasibility over the entire uncertainty set. Practically, a robust dose distribution will be smoother (fewer peaks and valleys), as compared to a nominal dose distribution, to ensure that voxels that move according to a different pdf still receive enough dose. See Section 2.6.1 for a related discussion on the concept of a dose distribution fixed in space.

In order to create a tractable formulation that is equivalent to (2.4), we need the following proposition, which is essentially a *theorem of the alternative*.

Proposition 2.3. *For a given v and z, we have $\beta_v(\boldsymbol{w}) \geq z$ if and only if there exist q_v and $\mathbf{r}_v$ such that*

$$\begin{aligned}
&\sum_{x\in U} \underline{p}(x) q_v - \sum_{x\in U} r_{v,x} - \sum_{b\in\mathcal{B}} \sum_{x\in U} \Delta_{v,x,b} \underline{p}(x) w_b \geq z, \\
&(\overline{p}(x) + \underline{p}(x)) q_v - r_{v,x} \leq \sum_{b\in\mathcal{B}} \Delta_{v,x,b} (\overline{p}(x) + \underline{p}(x)) w_b, \quad \forall\, x \in U, \\
&q_v \text{ free}, \\
&r_{v,x} \geq 0, \quad \forall\, x \in U.
\end{aligned} \tag{2.10}$$

Proof: The dual of (2.7) is

$$\begin{aligned}
\underset{q_v, \mathbf{r}_v, \mathbf{s}_v}{\text{maximize}} \quad & \sum_{x\in U} \overline{p}(x) r_{v,x} + \sum_{x\in U} \underline{p}(x) s_{v,x} \\
\text{subject to} \quad & q_v + r_{v,x} + s_{v,x} = \sum_{b\in\mathcal{B}} \Delta_{v,x,b} w_b, \quad \forall\, x \in U, \\
& q_v \text{ free}, \\
& r_{v,x} \leq 0, \quad \forall\, x \in U, \\
& s_{v,x} \geq 0, \quad \forall\, x \in U.
\end{aligned} \tag{2.11}$$

Using the equality constraint to eliminate the $s_{v,x}$ variables, and replacing the vari-

ables $r_{v,x}$ with $-r_{v,x}/(\overline{p}(x)+\underline{p}(x))$, we arrive at an equivalent formulation

$$\begin{aligned}
\underset{q_v,\mathbf{r}_v,}{\text{maximize}} \quad & \sum_{x\in U}\underline{p}(x)q_v - \sum_{x\in U} r_{v,x} - \sum_{b\in\mathcal{B}}\sum_{x\in U}\Delta_{v,x,b}\underline{p}(x)w_b \\
\text{subject to} \quad & (\overline{p}(x)+\underline{p}(x))q_v - r_{v,x} \le \sum_{b\in\mathcal{B}}\Delta_{v,x,b}(\overline{p}(x)+\underline{p}(x))w_b, \quad \forall\, x\in U, \\
& q_v \text{ free}, \\
& r_{v,x}\ge 0, \quad \forall\, x\in U.
\end{aligned} \tag{2.12}$$

To prove the forward implication, suppose $\beta_v(\boldsymbol{w})\ge z$. Then the minimization problem (2.7) is bounded (and feasible), so there exists an optimal solution, $\hat{p}^*$, with cost

$$\sum_{b\in\mathcal{B}}\sum_{x\in U}\Delta_{v,x,b}\hat{p}^*(x)w_b \ge z.$$

By strong duality of linear programming, there exists an optimal solution $(q_v^*,\mathbf{r}_v^*)$ to (2.12) with cost

$$\sum_{x\in U}\underline{p}(x)q_v^* - \sum_{x\in U} r_{v,x}^* - \sum_{b\in\mathcal{B}}\sum_{x\in U}\Delta_{v,x,b}\underline{p}(x)w_b = \sum_{b\in\mathcal{B}}\sum_{x\in U}\Delta_{v,x,b}\hat{p}^*(x)w_b \ge z$$

and which therefore satisfies (2.10). For the reverse implication, let $(q_v,\mathbf{r}_v)$ satisfy (2.10). Then (2.12) is feasible and has cost at least z. All feasible $\hat{p}$ for the primal satisfy

$$\sum_{b\in\mathcal{B}}\sum_{x\in U}\Delta_{v,x,b}\hat{p}(x)w_b \ge \sum_{x\in U}\underline{p}(x)q_v - \sum_{x\in U} r_{v,x} - \sum_{b\in\mathcal{B}}\sum_{x\in U}\Delta_{v,x,b}\underline{p}(x)w_b \ge z$$

by weak duality of linear programming. Hence, the optimal cost of the primal satisfies $\beta_v(\boldsymbol{w})\ge z$. □

We now use the above proposition to transform (2.4) into a linear program.

Theorem 2.1. *Formulation* (2.6) *is equivalent to the following linear program:*

$$
\begin{aligned}
\underset{\boldsymbol{w},\mathbf{q},\mathbf{r}}{\text{minimize}} \quad & \sum_{v\in\mathcal{N}}\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \\
\text{subject to} \quad & \sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b - \sum_{b}\sum_{x\in U}\Delta_{v,x,b}\underline{p}(x)w_b \\
& \quad + \sum_{x\in U}\underline{p}(x)q_v - \sum_{x\in U} r_{v,x} \geq \theta_v, \quad \forall\, v\in\mathcal{T}, \\
& (\overline{p}(x)+\underline{p}(x))q_v - r_{v,x} \leq \sum_{b}\Delta_{v,x,b}(\overline{p}(x)+\underline{p}(x))w_b, \quad \forall\, v\in\mathcal{T}, \forall\, x\in U, \\
& q_v \text{ free}, \quad \forall\, v\in\mathcal{T}, \\
& r_{v,x}\geq 0, \quad \forall\, v\in\mathcal{T}, \forall\, x\in U, \\
& w_b \geq 0, \quad \forall\, b\in\mathcal{B}.
\end{aligned}
\tag{2.13}
$$

Proof: Invoke Proposition 2.3 with $z = \theta_v - \sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b$. □

Formulation (2.13) provides us with a linear program that takes into account the uncertainty set in a tractable manner. By adjusting the size of the uncertainty set in our model, we will affect the conservativeness of the resulting robust solution. We observe that whereas the nominal formulation (2.3) has $m_{\mathcal{T}}$ constraints and n variables, the robust formulation (2.13) has $m_{\mathcal{T}}m_U + m_{\mathcal{T}}$ constraints and $m_{\mathcal{T}}m_U + m_{\mathcal{T}} + n$ variables. Although the problem now has a larger size it is still a linear program and, therefore, tractable for realistic values of m_U and $m_{\mathcal{T}}$. See Sections 2.6 and 2.7 for the exact problem sizes and running times of our experiments.

2.4.3 Margin formulation

Independent from the development of our robust methodology, we formulate the problem of ensuring enough dose to the tumor in the absolute worst-case, without any information on the motion except for bounds on the amplitude. This type of formulation is what we will call a *margin* formulation. This name comes from the medical physics community, where a margin (an expansion of the irradiated area) is used to

account for uncertainty and ensure that the tumor receives a sufficient dose. It is important to note that other studies aimed at calculating optimal margin sizes [57, 16] implicitly assume some distribution for an unknown pdf. This results in smaller margin sizes and less conservative solutions as compared to the margin approach described here, which aims to deliver a uniform dose to the tumor in the absence of a particular distribution for the pdf. In this section, we illustrate the connection between the margin concept and our robust formulation.

The robust formulation proposed in the previous section allows us to control the robustness of our solution depending on the amount of uncertainty present. In one extreme, one assumes perfect knowledge of the motion pdf, which is equivalent to the nominal formulation. At the opposite extreme, one assumes that there is no information on the motion (other than the possible range or amplitude of motion, which is captured by the set X), and protects against the absolute worst-case, which is exactly the margin approach. Intuitively, we expect to be able to replicate these formulations by using the robust formulation with an appropriate uncertainty set. We will show that the robust formulation builds a bridge between these two extreme solutions and gives us the flexibility to craft solutions anywhere in between.

First, we need to more precisely define a margin. In fact, there are two types of margins to consider: a "dosimetric margin" and a "geometric margin." The difference between these two concepts reflects the difference between using anatomical voxels to account for the tumor and other structures versus voxels that are fixed in space (based on a coordinate system independent of the patient). A dosimetric margin is the minimal (in terms of total dose delivered) dose distribution that, when added voxel by voxel to the basic static dose distribution, delivers a sufficient amount of dose to the tumor throughout each phase of the breathing cycle. A geometric margin (similar to an Internal Target Volume (ITV) in the medical physics literature) is constructed by taking the union of the tumor voxels throughout the breathing cycle, and subtracting away the nominal tumor voxels (the voxels representing the tumor in its nominal position). Under some mild technical conditions (cf. Section 2.6.1), we can show that these two margin concepts are equivalent. In practice, however,

the dosimetric margin is the more meaningful concept since it is able to account for the fact that the motion itself may affect the dose delivered to a point in space. For example, as a result of motion, the beam of radiation may pass through ribs or different amounts of air in the lung, which would affect the dose delivered. Hence, we focus on the dosimetric margin here. In [13], we focused on a geometric margin since the problem had a simplified one-dimensional geometry. We will provide more details about the geometric margin in Section 2.6.

We write the *dosimetric margin problem* as

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{V}}\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \\ \text{subject to} \quad & \sum_{b\in\mathcal{B}}\Delta_{v,x,b}w_b \geq \theta_v, \quad \forall\, v\in\mathcal{T},\ \forall\, x\in X, \\ & w_b \geq 0, \quad \forall\, b\in\mathcal{B}. \end{aligned} \tag{2.14}$$

This formulation ensures that every voxel in the tumor receives a sufficient dose in every phase of the breathing cycle. As mentioned before, the dose to the tumor is typically desired to be uniform, in which case we have $\theta_v = \theta$ for all $v \in \mathcal{T}$.

By considering the function of the model of uncertainty, it should be intuitive that the smallest uncertainty set (the set consisting only of p) replicates the nominal formulation situation, while the largest uncertainty set (setting $U = X, \overline{p} = 1-p, \underline{p} = p$) would give rise to a margin-like solution. The next proposition makes this intuition concrete, illustrating the fact that the robust formulation is a generalization of the margin and nominal formulations.

Proposition 2.4. *If $U = \emptyset$ (or equivalently, $\overline{p} = \underline{p} = 0$), then the robust formulation is equivalent to the nominal formulation. Alternatively, if $U = X$, $\overline{p} = 1 - p$, and $\underline{p} = p$, then the robust formulation is equivalent to the dosimetric margin formulation.*

Proof: If $U = \emptyset$, then $P_U = \{p\}$ and the robust formulation is equivalent to the nominal formulation. Next, we will prove the equivalence of the dosimetric margin

formulation and the robust formulation (2.4) with

$$P_X = \left\{ \tilde{p} \in \mathbb{R}^{|X|} \,:\, 0 \leq \tilde{p}(x) \leq 1, \forall\, x \in X;\ \sum_{x \in X} \tilde{p}(x) = 1 \right\}$$

in place of P_U:

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v \in \mathcal{V}} \sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} p(x) w_b \\ \text{subject to} \quad & \sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} \tilde{p}(x) w_b \geq \theta_v, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{p} \in P_X, \\ & w_b \geq 0, \quad \forall\, b \in \mathcal{B}. \end{aligned} \tag{2.15}$$

Let $\boldsymbol{w}$ be a feasible solution to the robust problem (2.15). Pick an arbitrary $\overline{x} \in X$ and define

$$p^*(x) = \begin{cases} 1, & \text{if } x = \overline{x}, \\ 0, & \text{otherwise.} \end{cases}$$

Then, for any $v \in \mathcal{T}$,

$$\sum_{b \in \mathcal{B}} \Delta_{v,\overline{x},b} w_b = \sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} p^*(x) w_b \geq \theta_v,$$

which demonstrates that $\boldsymbol{w}$ is feasible for the dosimetric margin problem (2.14). For the converse, assume $\boldsymbol{w}$ is feasible for (2.14), and let $\tilde{p} \in P_X$. For any $v \in \mathcal{T}$,

$$\sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} \tilde{p}(x) w_b \geq \sum_{x \in X} \theta \tilde{p}(x) = \theta_v.$$

□

We can think of the solution of the nominal formulation as the most risky because it takes at face value the information contained in the motion pdf. A solution to the margin formulation is the most conservative since it must protect against the largest uncertainty set. By using the robust formulation with an uncertainty set appropriate for the patient at hand, the solution will automatically be driven towards a margin

solution if the uncertainty set is big, a nominal solution if the uncertainty set is small, or something in between, otherwise.

The robust formulation also provides some insight into why using a margin may be a poor method in many cases. The margin solution is only optimal when $U = X, \overline{p} = 1-p$, and $\underline{p} = p$; that is, only if there is a possibility that the tumor will spend 100% of its time in any phase of the breathing cycle. This is especially dubious for extreme phases of the breathing cycle such as inhale and exhale. Thus, by gathering information on a patient's breathing and associated uncertainty set, we should always be able to improve upon a margin solution.

2.4.4 A delayed constraint generation approach to solving the robust formulation

In Section 2.4.2, we reformulated the robust formulation (2.4) as an equivalent linear program, which can then be solved using a standard linear programming solver. Alternatively, we can leave the robust formulation in its raw form and use constraint generation to solve it.

Note that P_U is a bounded polyhedron, so it has no extreme rays. Let $p^1, \ldots, p^I$ denote the extreme points of P_U. Then (2.4) may be rewritten as

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{N}} \sum_{b\in\mathcal{B}} \sum_{x\in X} \Delta_{v,x,b} p(x) w_b \\ \text{subject to} \quad & \sum_{b\in\mathcal{B}} \sum_{x\in X} \Delta_{v,x,b} p^i(x) w_b \geq \theta_v, \quad \forall\, v \in \mathcal{T},\ \forall\, i = 1, \ldots, I, \\ & w_b \geq 0, \quad \forall\, b \in \mathcal{B}. \end{aligned} \tag{2.16}$$

In this form, (2.16) can be solved using a standard delayed constraint generation technique outlined below. Let F^k denote formulation (2.16) with only k of the inequality constraints involving θ_v.

Delayed constraint generation for robust formulation

1. Solve F^k and let $\overline{\boldsymbol{w}}$ be an optimal solution.

2. For all $v \in \mathcal{T}$, solve the subproblem

$$\begin{aligned} f(v) = \min_{\tilde{p}} \quad & \sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} \tilde{p}(x) \overline{w}_b \\ \text{subject to} \quad & \tilde{p} \in P_U. \end{aligned} \tag{2.17}$$

   (a) If $f(v) \geq \theta_v$ for all $v \in \mathcal{T}$, terminate with the optimal solution $\overline{\boldsymbol{w}}$.

   (b) If $f(v) < \theta_v$ for some $v \in \mathcal{T}$, add the constraint

$$\sum_{b \in \mathcal{B}} \sum_{x \in X} \Delta_{v,x,b} p^i(x) w_b \geq \theta_v$$

   to F^k, where p^i is an optimal extreme point from (2.17).

3. Return to step 1.

A simple way to initialize the algorithm would be to solve the nominal problem in step 1. This algorithm could be employed to solve very large scale instances of clinical cases that are faced in practice. For the examples we considered in this chapter, the linear robust counterpart was solved fairly quickly, so we did not further explore constraint generation or alternative methods. However, a computational study using constraint generation would be an interesting topic for future research.

2.5 Patient Data and PDFs

Probability density function data are not immediately available from the planning session, but must be mined from other data that are collected. One type of data that is collected for patients that exhibit significant motion is data that track the movement of the tumor. The best way to obtain data on internal motion is to implant a marker into the tumor and continuously image the patient, tracking the motion of the marker. Often though, this method is not used due to possible complications associated with the implantation procedure or the inability of the facility to carry it out. A far less invasive alternative involves the use of an external marker placed on the patient's

abdomen. By recording the motion of this external marker, we obtain information about the patient's breathing pattern and corresponding internal motion. Then, by estimating anatomical amplitudes through imaging, we can use this external data as a surrogate for internal motion data. In the remainder of this section, we describe how we obtain pdf data to be included in our optimization runs for the three-dimensional clinical case described in Section 2.7. At the beginning of Section 2.6, we briefly highlight the differences in the pdf data used for the one-dimensional case.

The patient data used in this research were acquired from Massachusetts General Hospital. We used Varian's (Varian Medical Systems, Inc., Palo Alto, CA) real-time position management (RPM) system to gather data from five patients by recording their breathing patterns in every treatment fraction over the course of their entire treatments. RPM data are recorded by an infrared camera mounted in the treatment room, tracking the movement of an optical marker placed on the chest or abdomen of the patient. This camera records the respiratory motion of the patient by observing the up and down motion of the marker block. Much research has focused on correlating the external signal to internal motion [58, 19, 55], and we will assume, as many studies have shown to be the case, that the internal and external movement are highly correlated. In reality, there are uncertainties that need to be taken into account when using this method, but this is outside the scope of our work. Ideally, internal motion data would be used instead of this external data. Accordingly, we develop our framework with the understanding that future advances in medical imaging will greatly enhance the accuracy and availability of anatomical motion tracking, and will give us data that are indeed representative of internal motion.

The recent advance of breathing-correlated "4D" computed tomography imaging [40, 32, 23] in combination with deformable image registration [35, 12, 33] has made it possible to visualize tissue motion trajectories within the scanned 4D-CT volume of a patient, and could in principle provide an alternative to RPM data. However, while 4D-CT data are a useful tool to study the spatial variability of motion, they do not reflect temporal variability, which has a stronger influence on the delivery outcome. Furthermore, we do not acquire this internal data for each treat-

ment fraction due to time and resource constraints. Therefore, we focus on RPM data since they are easily obtained on different days and can be used to estimate the variability in breathing from fraction to fraction.

The amplitude of an RPM trace is not representative of any particular anatomical motion since its units depend on many external factors such as the exact position and tilt of the marker block on the patient's abdomen, and the distance from the block to the camera. Still, the RPM data are useful for measuring variations in the breathing patterns. Therefore, we scaled the average RPM amplitude for each trace to the average anatomical amplitude for the corresponding patient gathered from 4D-CT scans. The number of traces from each patient, along with their anatomical amplitudes from 4D-CT are listed in Table 2.1.

Table 2.1: RPM data used in the clinical lung case.

Patient number	Patient name	Number of traces	Anatomical amplitude (cm)
1	Cardiac-1	78	1.5
2	Liver-1	44	1
3	Lung-1	75	1
4	Cardiac-2	79	1
5	Lung-2	52	1

Each RPM trace was used to create a single pdf by integrating over the displacement axis of the RPM trace. This produces a function that describes the relative amount of time the tumor spends in various phases. Figure 2-1 shows an example RPM trace scaled to anatomical amplitude and Figure 2-2 shows the corresponding pdf. This method of generating the pdfs from motion data was first presented in [34].

For the 3D problem, the patient's breathing cycle was characterized by ten distinct phases. Ideally, we would have a Δ matrix associated with each one, but for this problem, we only had five Δ matrices that were calculated specifically for the anatomy of the patient at phases 1, 3, 5, 7 and 9 (phase 1 was inhale, while phases 5 and 7 were closest to exhale). The corresponding displacements of the centroid of the tumor for these five phases (relative to inhale) were 0 mm, 3.1 mm, 8.3 mm, 8.7 mm, and 5.7 mm, respectively. Figure 2-3 relates the phases to the displacement of the tumor

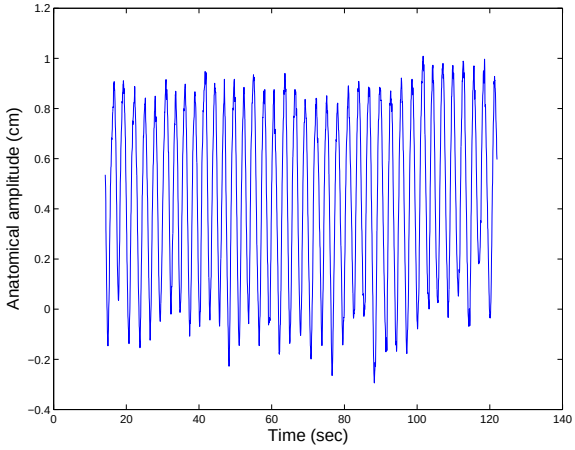


Figure 2-1: An example RPM trace.

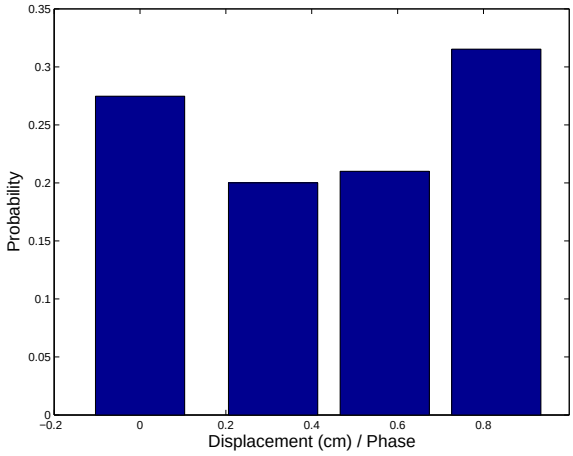


Figure 2-2: The pdf corresponding to the RPM trace in Figure 2-1.

centroid for this patient. This phase/displacement data and the Δ matrices are the same as in [54].

Since phases 5 and 7 were very close to exhale, we combined these two phases in the generation of the pdfs. Thus, the pdfs were constructed from the RPM data based on four phases centered at displacements 0 mm, 3.1 mm, 5.7 mm, and 8.5 mm (this last displacement, exhale, was obtained from taking the average of the phase 5 and 7 displacements). If we kept these phases separate, then phase 7 would only receive the probability mass corresponding to amplitudes above 8.5 mm, which would make it seem like the patient spends much less time in phase 7 than phase 5. The pdfs

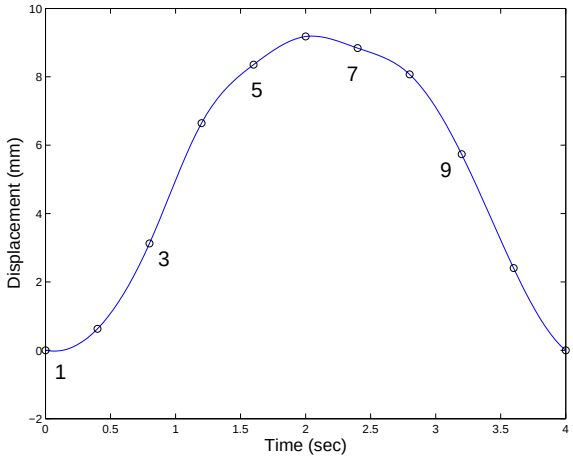


Figure 2-3: Position of the tumor centroid in the ten phases of the breathing cycle. Phases 1, 3, 5, 7, 9 are labeled

used in the optimization, however, had a support of five phases. This is due to the fact that the Δ matrices corresponding to each phase are different from one another. A pdf based on five phases was constructed by dividing the probability (from the four-phase pdf) of being at exhale (at 8.5 cm) equally between phase 5 and 7.

In order to make the pdfs from the other five patients useful for our optimization framework, we needed to bin the RPM data into the above bins, since they were the most appropriate for the $\Delta_{v,x,b}$ matrices we had. That is, we are given the matrices $\Delta_{v,x,b}$ for a fixed set of phases ($X = \{1, 3, 5, 7, 9\}$), and in order to use any pdfs generated from this process, we must tailor the support of these pdfs to match the set X of possible phases. The finer the set X is the more flexibility we have to precisely specify a pdf. However, the set of breathing phases X is typically small, resulting in somewhat coarse pdfs.

The error bars and associated uncertainty set are generated from a family of pdfs as shown in the following section. The idea is that once we have some information about the patient at hand in the form of a nominal pdf from RPM data, we can use information on breathing patterns from past patients to form an uncertainty set for this patient.

2.5.1 Creating the uncertainty set from patient data

The error bars in our model are formed from a family of pdfs as follows. Consider a family of pdfs $p_0, p_1, \ldots, p_k$ obtained from patient data (e.g., according to the method described above), where $p_0 := p$ is the nominal pdf of our current patient, p_1 is the nominal pdf of a past patient, and $p_2, \ldots, p_k$ are the pdfs of the past patient realized during his/her treatment fractions. We are interested in finding out how this past patient deviated from his/her nominal pdf on subsequent days, and using this as a measure of the relative deviation that our current patient exhibits from his/her nominal pdf. Of course, this relationship needs to be tested in practice in order to be useful for clinical implementation. However, the underlying assumption seems reasonable, as long as we build the model of uncertainty based on data from past patients that are appropriately similar to the patient at hand. What is deemed "appropriately similar" should be studied in future work.

Let us generate error bars for this past patient as follows. For every $x \in X$,

$$\underline{p_1}(x) := p_1(x) - \min_{i=1,\ldots,k} p_i(x) \tag{2.18}$$

and

$$\overline{p_1}(x) := \max_{i=1,\ldots,k} p_i(x) - p_1(x). \tag{2.19}$$

We define the relative minimum deviation below the nominal pdf p_1 to be $\underline{p_1}/p_1$, and the relative maximum deviation above the nominal pdf p_1 as $\overline{p_1}/(1-p_1)$. Note that the definition for the maximum takes into account that the fact that the realized pdfs are bounded above by 1. Using these quantities, we define $\overline{p}$ and $\underline{p}$ for the current patient as

$$\underline{p}(x) := \frac{\underline{p_1}(x)}{p_1(x)} p(x)$$

and

$$\overline{p}(x) := \frac{\overline{p_1}(x)}{1-p_1(x)}(1-p(x)),$$

for every $x \in X$. Finally, the upper and lower error bars of p can be written as

$$p(x) + \overline{p}(x) = p(x) + \frac{\overline{p_1}(x)}{1 - p_1(x)}(1 - p(x))$$

and

$$p(x) - \underline{p}(x) = p(x) - \frac{\underline{p_1}(x)}{p_1(x)}p(x),$$

respectively. It is straightforward to check that using these definitions for $\overline{p}$ and $\underline{p}$, $0 \leq p - \underline{p} \leq p + \overline{p} \leq 1$. The natural extension of the above procedure is to use multiple patients $j = 1, \ldots, m$, to create $\underline{p}$ and $\overline{p}$:

$$\underline{p}(x) := \max_{j=1,\ldots,m} \frac{\underline{p_j}(x)}{p_j(x)}p(x) \tag{2.20}$$

and

$$\overline{p}(x) := \max_{j=1,\ldots,m} \frac{\overline{p_j}(x)}{1 - p_j(x)}(1 - p(x)), \tag{2.21}$$

where $p_j, \overline{p_j}, \underline{p_j}$ are indexed by patient j.

2.6 A one-dimensional case

Before we proceed to the results on a three-dimensional case, we first consider a one-dimensional problem in order to build intuition. In fact, this 1D study was conducted before the 3D case, so there are slight differences in the formulations and data generation presented in the previous sections.

2.6.1 Modifications for the one-dimensional case

First, in the objective function of the various formulations, we replaced $\mathcal{N}$ with $\mathcal{V}$ (the set of all voxels in the phantom). This was necessary in the 1D case, since using $\mathcal{N}$ in the objective would lead to intensity maps that were infinitely high over the tumor in order to lower the dose to the left and right of the tumor (the solution would try to get an infinitely steep falloff of the dose over the edge of the tumor). This is

clearly not an issue in the 3D case since the beam of radiation would pass through healthy tissue on the way to the tumor. But in the 1D case, we could not directly account for this healthy tissue in front of the tumor, so instead, we minimized the dose delivered to the entire phantom.

The definition of the pdfs was also modified. In the 1D case, the pdfs were based on displacements, and not on phases. We did this because of the simple geometry in this case. We assumed that the dose delivered by some beamlet to voxel v at 0 cm, shifted by 1 mm to the right, would be the same dose delivered to the voxel 1 mm to the right of v, in the absence of motion. Thus, we were able to use one D matrix and simply adjust it for shifts left and right. These shifts were taken to be in increments of a whole voxel size, and allowed us to have pdfs based on a finer support, depending on the number of shifts to the left and right we considered. The error bars were also constructed slightly differently. For this case, we simply took the upper and lower envelopes of all pdfs available (as shown in equations (2.18) and (2.19)).

An important concept mentioned in Section 2.4.3 is that of the *static dose cloud assumption*, named so since the picture it depicts is that of a dose distribution fixed in space and a voxel acquiring dose as it moves in this cloud. This is a reasonable assumption for very simple geometries where movements will not greatly affect the dose delivered to a point in space, hence, we assumed the static dose cloud assumption for the 1D case. This is also integral to the idea of geometric voxels and how they are typically inadequate to describe the dose that anatomical voxels receive. This assumption is precisely stated below.

Assumption 2.1 (Static Dose Cloud)**.** *Under the static dose cloud assumption, a dose distribution is defined on voxels based on a treatment room coordinate system (geometric voxels), instead of on anatomical voxels. As anatomical voxels move within the treatment room, they will receive dose proportional to the dose in the geometric voxels they pass through. In particular, for a particular beamlet b, phase x, and anatomical voxel v,*

$$\Delta_{v,x,b} = D_{v',b}$$

where v' is the geometric voxel that coincides with the location of v in phase x. We assume that all shifts of the voxels are in unit increments of the size of a voxel.

Accordingly, the margin formulation in the one-dimensional case differed from formulation (2.14) as well. In the one-dimensional case, a geometric margin was considered, instead of a dosimetric margin. The geometric margin problem is written as

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{V}}\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \\ \text{subject to} \quad & \sum_{b\in\mathcal{B}} D_{v,b}w_b \geq \theta_v, \quad \forall\, v\in\mathcal{T}, \\ & \sum_{b\in\mathcal{B}} D_{v,b}w_b \geq \mu_v, \quad \forall\, v\in\mathcal{M} \\ & w_b \geq 0, \quad \forall\, b\in\mathcal{B}, \end{aligned} \tag{2.22}$$

where the geometric margin, $\mathcal{M}$, is

$$\mathcal{M} := \{v\in\mathcal{V}\setminus\mathcal{T} \mid v = \hat{v}+x \text{ for some } \hat{v}\in\mathcal{T}, x\in X\}.$$

Here, $\mathcal{V}$ and $\mathcal{T}$ are fixed in space, representing the voxels of the phantom and tumor in the nominal reference frame. The geometric margin, $\mathcal{M}$, is composed of treatment room voxels mapped back to the anatomical voxels. It turns out that the mapping of the geometric voxels to anatomical voxels is what limits the usefulness of the geometric margin, since the static dose cloud assumption is typically violated in practice. However, if this assumption holds, it turns out that the two margin formulations are equivalent.

Proposition 2.5. *Under Assumption 2.1, if $\theta_v = \theta$ for all $v\in\mathcal{T}$ and $\mu_v = \theta$ for all $v\in\mathcal{M}$, the geometric and dosimetric margin formulations are equivalent.*

Proof: Any $v\in\mathcal{T}\cup\mathcal{M}$ can be written as $\hat{v}+x$ for $\hat{v}\in\mathcal{T}$ and $x\in X$. Hence,

$$\sum_{b\in\mathcal{B}} D_{v,b}w_b = \sum_{b\in\mathcal{B}} \Delta_{\hat{v},x,b}w_b,$$

and the two formulations have the same feasible region. □

The bottom line is that an Internal Target Volume (what we denote as $\mathcal{T} \cup \mathcal{M}$) is not a good margin concept, unless the case at hand approximately satisfies Assumption 2.1 and the patient is a very irregular breather. Next, we demonstrate the results of the nominal, robust, and geometric margin formulations on a one-dimensional case.

2.6.2 Results for the one-dimensional case

In this section, we motivate the use of robustness by considering what happens when the nominal formulation is used, first in the absence of uncertainty, and then with a realized pdf that is different from the planned one. Next, we show that while the margin solution is robust to the uncertainty in the realized pdf, it significantly increases the dose to the normal tissue. Finally, we show how the robust solution produces a dose distribution that, unlike the nominal solution, can combat uncertainty, while delivering less dose than the margin solution. Numerical results are shown at the end to supplement the figures.

For the illustrations below, we considered a one-dimensional phantom containing a 10 cm tumor surrounded by 10 cm of normal tissue on either side, and one beam orthogonal to the direction of motion. Voxels were 0.2 cm wide and beamlets were 0.5 cm wide, resulting in a total of 151 voxels and 28 beamlets. As shown in Figure 2-4, the tumor motion is on the order of 2 cm peak to peak. The $D_{v,b}$ matrix was computed using error functions to model beamlet dose profiles according to [16], and the mathematical formulations were optimized using the barrier algorithm of ILOG's CPLEX 9.0 solver (ILOG, Inc., Gentilly, France) through a Matlab interface (The MathWorks, Inc., Natick, Massachusetts). Each run took a few seconds on a 1.5 GHz computer with 1 GB of memory. While the geometry of this example is one-dimensional, the illustrations below can be thought of as a slice of a two-dimensional phantom.

The pdfs and error bars were mined from actual patient data as described in Section 2.6.1. In our analysis, we used RPM data from four patients, totalling 95

traces. Each trace was baselined to its median value (so the resulting pdf has a median value of 0), and for simplicity, the average amplitudes of the RPM traces for the other three patients were scaled to approximately the average amplitude of the first patient. Although somewhat artificial, this was done to simulate the scenario where the population data is comparable in amplitude to the patient at hand. We used the planning RPM trace from the first patient to create the nominal pdf, the average of all 19 traces from the first patient to create the realized pdf, and all 95 traces (one pdf per trace) from the four patients to create the error bars. To create the pdfs, we used a bin size equal to the width of a voxel because we focused on shifts of voxels by integer multiples of the voxel width. Since we used RPM data as a surrogate for tumor motion data, the amplitudes were scaled to be representative of large tumor motion.

The reader should keep in mind the following situation as we examine the figures below. If a patient breathes somewhat regularly, then we would expect to see the "bowl-shaped" pdf in Figure 2-4 [34]. However, after many treatment days, the overall pdf may start to resemble a Gaussian distribution [18, 16] due to differences in breathing patterns from day to day ("realized pdf" in Figure 2-4). Thus, the principle behind applying the robust methodology in practice is as follows. If a patient breathes regularly during the planning session and we design the treatment plan using that pdf (nominal pdf), we will produce a plan with good tumor coverage and sharp dose gradients at the edge of the tumor, even if the cumulative pdf (realized pdf) over the course of the treatment is *significantly* different from the planned one. In the figures depicting the dose distributions of the three formulations, the pdf used in the planning for each is the nominal pdf in Figure 2-4, while the pdf that is realized is either the nominal (part (a) of each figure) or the realized (part (b)) pdf.

In Figure 2-5(a), we see the dose distribution from the nominal problem, using the same realized pdf as the nominal (planned) pdf. The tumor receives its prescribed dose in a largely homogeneous fashion, and there is a sharp falloff of the dose into the surrounding normal tissue. However, Figure 2-5(b) shows severe hot and cold spots that result from the dose distribution if the realized pdf is different from the nominal

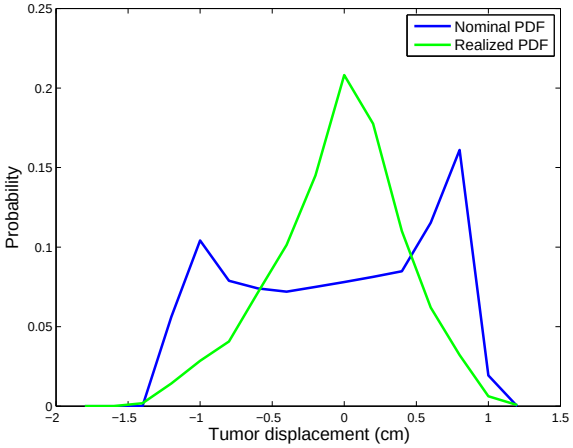
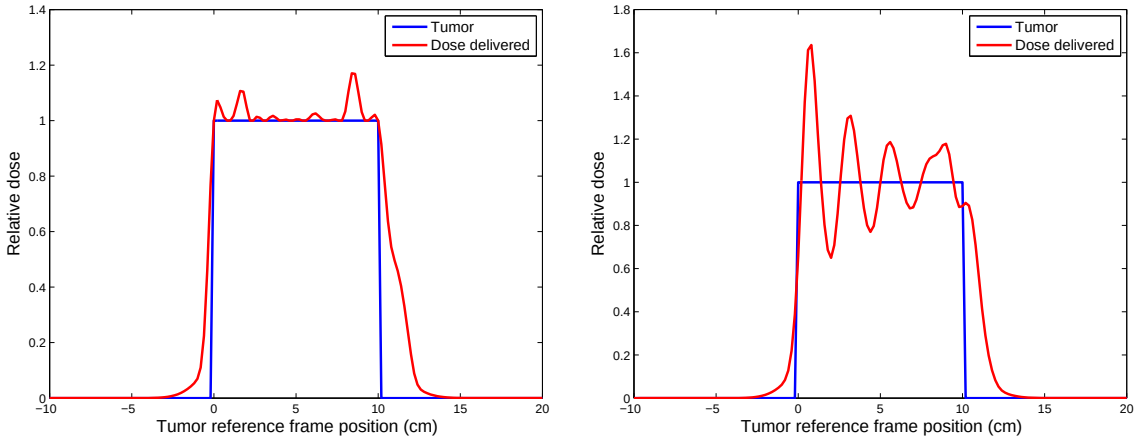


Figure 2-4: The pdfs used in the nominal, margin and robust formulation illustrations.

one.



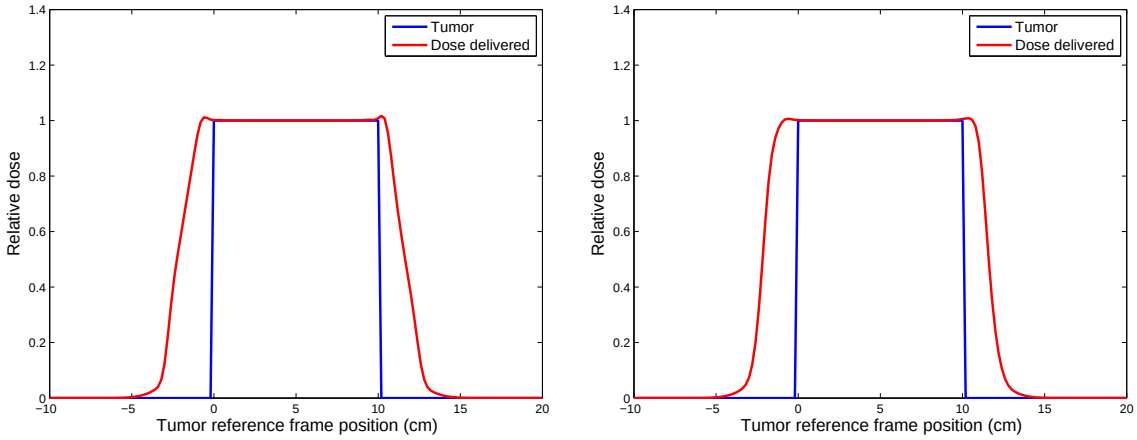
(a) Nominal pdf.

(b) Realized pdf.

Figure 2-5: Dose distribution of nominal solution using the pdfs from Figure 2-4.

Figures 2-6(a) and 2-6(b) show the corresponding figures when the margin approach is used. As expected, the tumor receives a sufficient, homogeneous dose in both cases, but the tails of the dose distribution are much broader and extend further into the normal tissue, indicating a significant increase in the dose delivered there.

We now describe the results of the robust formulation. The uncertainty region was set equal to X, and the error bars shown in Figure 2-7 depict the range of pdfs over which the robust solution is protecting. By comparing Figure 2-8(a) and



(a) Nominal pdf. (b) Realized pdf.

Figure 2-6: Dose distribution of margin solution using the pdfs from Figure 2-4.

Figure 2-8(b), we can see that the dose distribution from the robust plan is almost unchanged when a completely different pdf from the nominal pdf is realized. One interpretation of the robust dose distribution is that it is a combination of a margin-like distribution in the tumor to deal with uncertainty, and "horns" at the edge of the tumor to produce a sharp dose gradient, combining the features of both the margin and nominal solutions. As you can see in Figure 2-9, the intensity map (beamlet weights) corresponding to the robust solution is not overly complex.

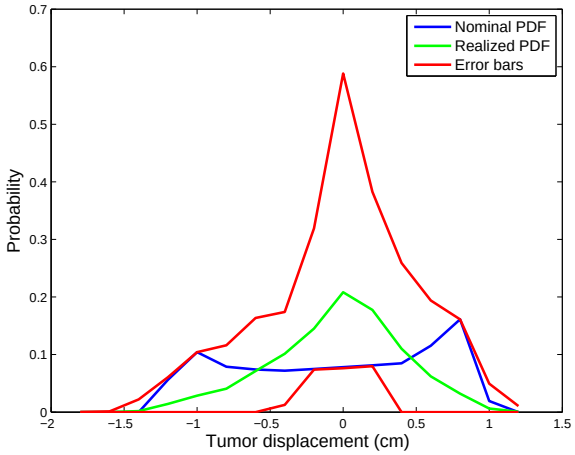
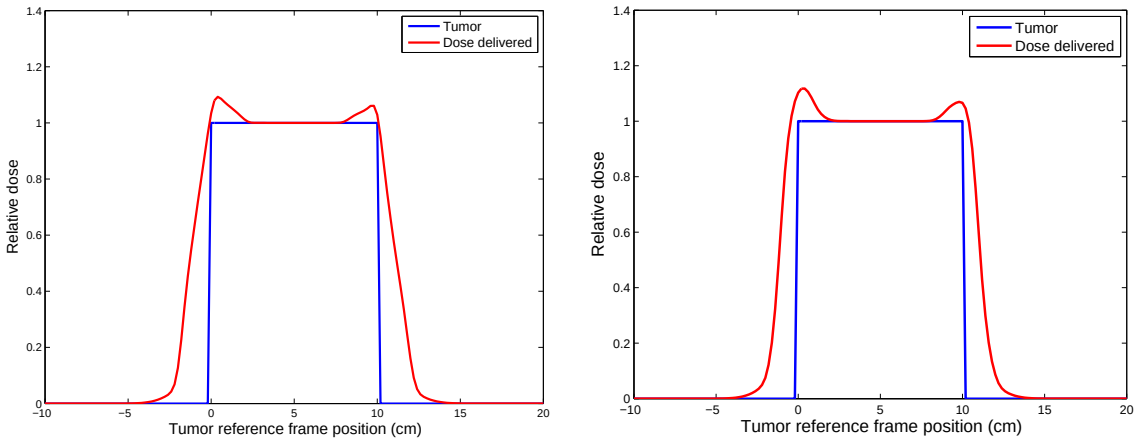


Figure 2-7: The pdfs and error bars used in the robust formulation illustrations.

Table 2.2 shows the numerical results corresponding to the previous figures. We



(a) Nominal pdf. (b) Realized pdf.

Figure 2-8: Dose distribution of robust solution using the pdfs and error bars from Figure 2-7.

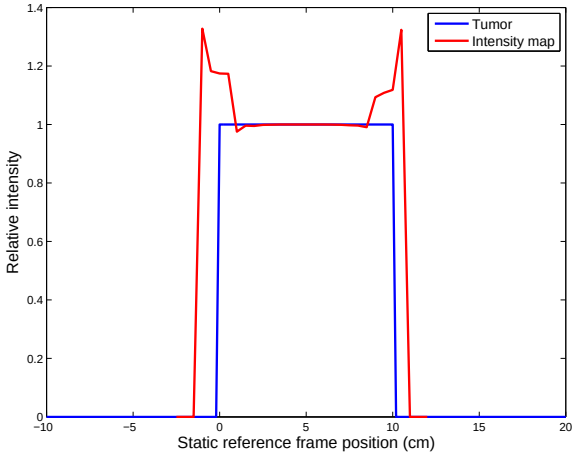


Figure 2-9: Intensity map corresponding to robust solution.

Table 2.2: Comparing formulations on integral dose delivered to phantom and integral dose delivered to normal tissue under the realized pdf. The relative amount of dose delivered (%) by the various formulations is normalized to the margin formulation.

	Nominal	Robust	Margin
Dose to phantom	85.29	91.45	100.00
Dose to normal tissue	31.41	61.97	100.00

can see that the robust solution delivers 8.55% less total dose and 38.03% less dose to the normal tissue than the corresponding margin solution. As expected, the nominal

solution performs the best in terms of dose delivered, but the tradeoff is that it has no capability to mitigate uncertainty. There is a price to be paid to have protection from uncertainty, but the point is that by using the robust solution, we can pay a substantially lower premium than the cost of the margin solution, while achieving the same level of robustness to uncertainty.

2.7 Results from a clinical lung case

In this section we present the results of using the three formulations (nominal, robust, and dosimetric margin) on a clinical problem, involving a tumor in the lower left lung. Our formulations are driven by a simple model where the major internal structures of interest are a tumor and the surrounding healthy lung tissue, which is a common representation of a lung cancer case. Since studies indicate that the mean dose delivered in the lung is well-correlated with the probability of normal tissue complication associated with treatment [51], minimizing the integral dose to the healthy tissue is an appropriate objective function.

Before implementing our formulations on a clinical case, we made one slight modification to account for realistic considerations. We added upper bound constraints to control the dosage fluctuations within the tumor. In particular, we constrained the maximum dose in the tumor to be no more than 10% above the minimum dose. Using a minimum dose of 72 Gy for the tumor meant that the maximum dose within the tumor did not exceed 79.2 Gy. This is slightly more stringent than the current standards which require that the dose in the tumor be no more than 5% below or 7% above the prescribed dose (a gap of 12%, or 12.6% above the minimum). We found that these modifications produced acceptable solutions, so we did not further refine the formulation. Of course, for very complex cases, constraints that limit the dose to certain healthy structures will be needed. These types of constraints can be added to our model without major revisions of the theory we presented. The robust

formulation used in the subsequent runs is shown in (2.23).

$$\begin{aligned}
\underset{\boldsymbol{w}}{\text{minimize}} \quad & \sum_{v\in\mathcal{N}}\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}p(x)w_b \\
\text{subject to} \quad & \sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}\tilde{p}(x)w_b \geq \theta_v, \quad \forall\, v\in\mathcal{T},\ \forall\,\tilde{p}\in P_U, \\
& \sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}\tilde{p}(x)w_b \leq \gamma\theta_v, \quad \forall\, v\in\mathcal{T},\ \forall\,\tilde{p}\in P_U, \\
& w_b \geq 0, \quad \forall\, b\in\mathcal{B}.
\end{aligned} \tag{2.23}$$

As mentioned above, $\gamma = 1.1$ and $\theta_v = 72$ for all $v \in \mathcal{T}$.

The anatomy was discretized into voxels of size 2.93 mm $\times$ 2.5 mm $\times$ 2.93 mm, resulting in a total of 110,275 voxels, 5495 of which were in the tumor. Radiation was delivered from five angles at 0° (directly anterior), 52°, 104°, 156°, and 208°, which resulted in a total of 1625 beamlets (of size 5 mm $\times$ 5 mm). We modeled the support of the motion pdfs using five phases of the breathing cycle. Our uncertainty set allowed the value of the realized pdf to differ from the nominal one in every phase x, which corresponded to setting $U = X$. The nominal pdf was taken from the planning session RPM of the patient to be treated (the one for whom we have the Δ matrices).

We conducted a series of experiments using pdf data derived from the five patients in Table 2.1. In each experiment, the pdfs corresponding to four patients were used to construct the uncertainty set (in the notation of equations (2.20)-(2.21), we have $m = 4$), and the pdf data from the fifth patient was used to test the resulting optimized solutions. This is meant to simulate the realistic situation where the uncertainty set may be generated from past data, whereas the pdfs realized during treatment may lie outside this uncertainty set. Each experiment used a different combination of four patients (out of the five total patients) to generate the uncertainty set. Since we did not have a Δ matrix for the patient at hand with excursions beyond its 1 cm amplitude, the RPM data for the patient with a mean amplitude of 1.5 cm (patient 1) was taken into account by increasing the probability of being at the extreme phases (those corresponding to inhale and exhale). Below, we present two cases, which represent the experiments that showed the least and most difference between the

nominal and robust solutions in terms of tumor coverage, respectively.

For the first experiment (the one that showed the least difference), we used data from patients 2 – 5, totaling 250 traces (multiple RPM traces are taken during most fractions), to generate the uncertainty set according to equations (2.20) and (2.21). Then, to measure the quality of the resulting optimized solutions, we used the 78 pdfs from patient 1. Let us call this experiment #1. We compare the performance of the nominal, robust, and margin solutions below. The three formulations were optimized using the barrier algorithm of CPLEX 9 on a 3 GHz computer with 3 GB of memory. The following table shows the number of variables and constraints in each formulation, along with the associated solution times.

Table 2.3: Basic information regarding the three formulations.

	Nominal	Robust	Margin
Number of variables	1625	67,565	1625
Number of constraints	10,990	65,940	54,950
Solution time (in min)	3	13	9

With an optimal solution, $\boldsymbol{w}^*$, of one of these formulations in hand, we can simulate its delivery by computing

$$\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}\tilde{p}(x)w_b^*$$

for each voxel $v \in \mathcal{V}$, where $\tilde{p}$ is a particular realized pdf. Different $\tilde{p}$'s in this equation lead to different delivered dose distributions. In the case of the robust formulation, for example, a $\tilde{p}$ that lies within the uncertainty set will guarantee

$$\sum_{b\in\mathcal{B}}\sum_{x\in X}\Delta_{v,x,b}\tilde{p}(x)w_b^* \geq \theta_v \quad \forall\, v \in \mathcal{T}.$$

The dose distribution corresponding to the robust solution realizing the nominal pdf is shown in Figure 2-10. The tumor receives sufficient coverage as expected. In this 3D case, the tumor is synonymous with the Clinical Target Volume (CTV), which contains the Gross Tumor Volume (GTV).

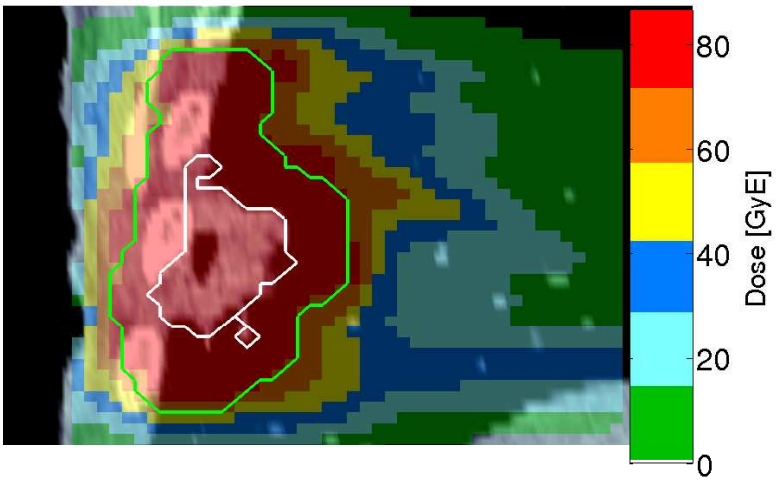


Figure 2-10: Dose distribution corresponding to the robust solution when the realized pdf is the same as the nominal one. The white outline denotes the GTV, while the green outline denotes the CTV.

A more useful way of visualizing three dimensional dose distributions is through a dose-volume histogram (DVH). A DVH illustrates what fraction of a particular organ of interest receives (at least) a certain level of dose. Figures 2-11, 2-12, and 2-13 show the DVHs corresponding to the results of delivering the nominal, robust, and margin solutions, respectively, for 78 different realizations of $\tilde{p}$ in experiment #1. The main organs of interest in this study are the tumor and left lung (in which the tumor resides), while secondary organs-at-risk (due to their relative remoteness, in this case) are the esophagus, spinal cord, and heart. These three figures overlay the 78 DVHs (one for each realized pdf) on the same axes, resulting in blurred lines that highlight the range of possible DVHs achievable with these realizations of the uncertain pdf. The vertical line at 72 Gy in the figures illustrates the location of the minimum dose requirement for the tumor. We will adopt the terminology of [14] and refer to these figures as "DVH clouds." Note that while our DV-clouds are quite narrow, the nominal solution exhibits more "blurriness" in the tumor cloud near 70 Gy and 80 Gy. That is, there is more variability in what fraction of the tumor receives a sufficient dose (equivalently, gets underdosed) and what fraction receives the most

dose. The fact that the tumor DV-clouds for the robust and margin solutions exhibit sharper falloffs is indicative of their increased robustness to the uncertain pdfs. For comparison, Figure 2-14 plots the three solutions for one particular realized pdf on the same axes. It can be seen that the nominal and robust solutions typically deliver less dose than the margin solution to the healthy tissue, especially in the left lung, which is the main organ-at-risk here. However, the nominal solution produces an inferior dose distribution in the tumor, increasing both the fraction of the tumor that gets underdosed and overdosed, as compared to the robust and margin solutions. Also, it is interesting to note that while the margin solution delivers the most homogeneous dose to the tumor (as evidenced by the tumor DV-line that is closest to vertical), it is outperformed by the robust solution near the minimum dose requirement. That is, the robust solution exhibits a steeper initial dropoff in its DV-line. This is due to the fact that the margin solution enforces tumor coverage in every phase of the breathing cycle, whereas the robust solution provides tumor coverage averaged over the phases (based on the family of pdfs in the uncertainty set).

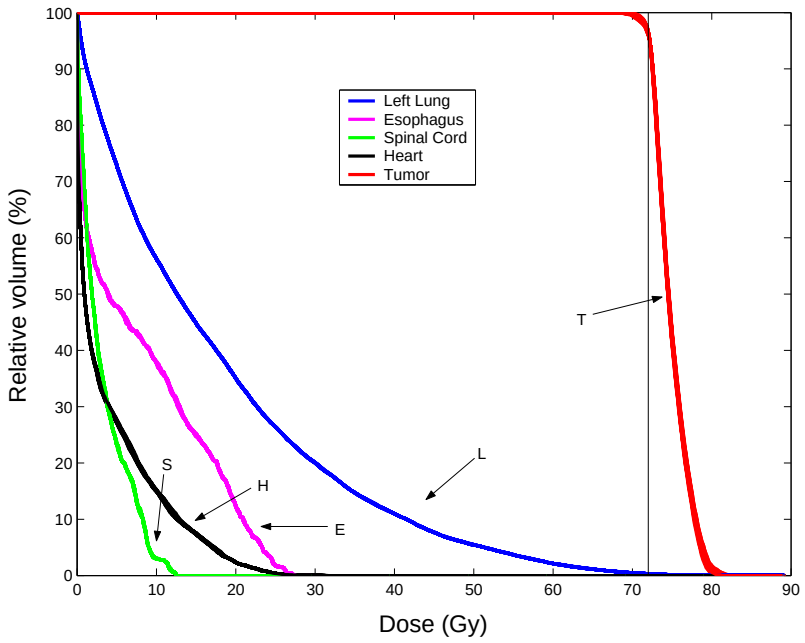


Figure 2-11: Nominal solution DVHs for 78 realized pdfs in experiment #1.

Numerical data from these three solutions are summarized in Table 2.4.

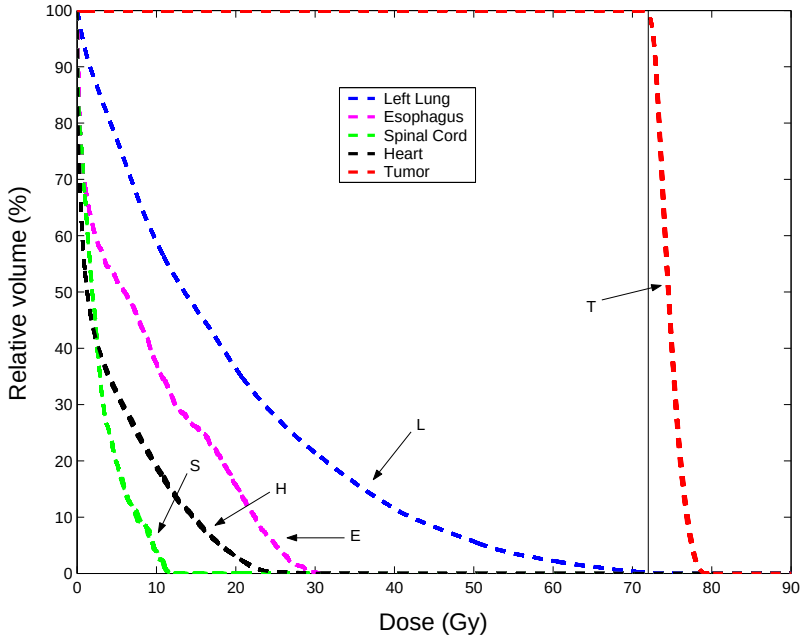


Figure 2-12: Robust solution DVHs for 78 realized pdfs in experiment #1.

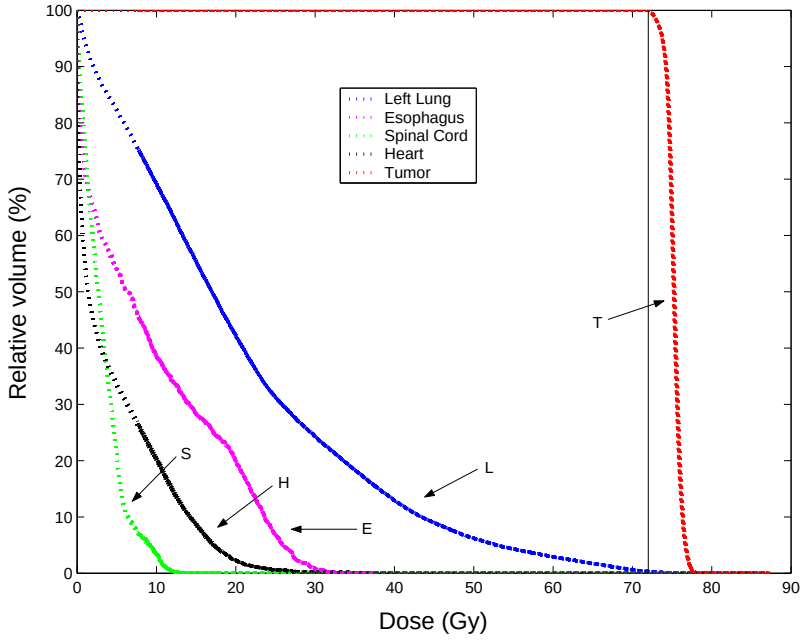


Figure 2-13: Margin solution DVHs for 78 realized pdfs in experiment #1.

Note that the robust solution essentially ensures that the tumor receives the required dose (only an average 1% underdose), while significantly reducing the dose

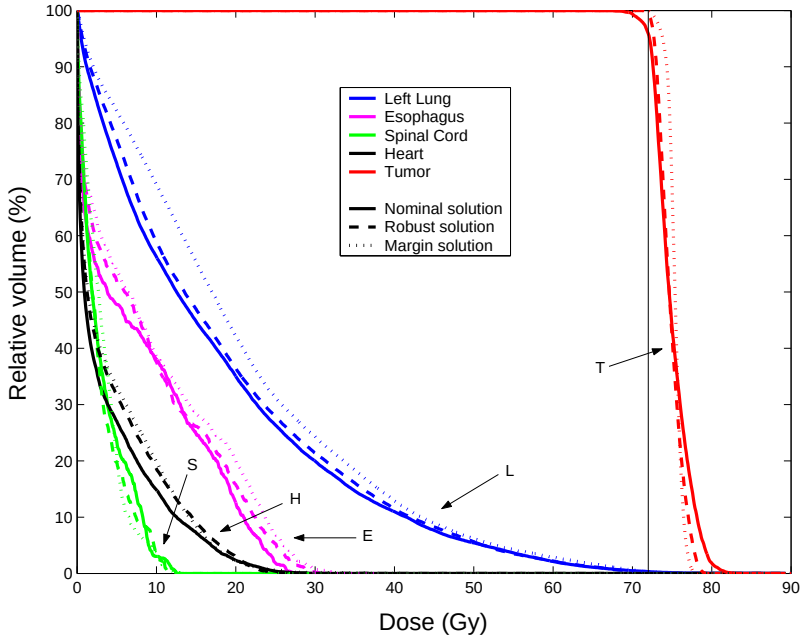


Figure 2-14: The nominal, robust, and margin solution DVHs for one realized pdf in experiment #1.

Table 2.4: Comparing the nominal, robust, and margin formulations using the average of the 78 trials in experiment #1. This experiment showed the least difference between the nominal and robust solutions in terms of tumor coverage. The left subcolumns represent values in Gy, while the right subcolumns are percentages. The first row percentages are relative to 72 Gy, which is the desired dose level, while the next two rows are quoted as a percentage of the dose delivered by the margin solution.

	Nominal		Robust		Margin	
	Gy	%	Gy	%	Gy	%
Minimum dose delivered in tumor	67.72	94.06	71.40	99.17	72.04	100.06
Integral dose to left lung	17.40	85.29	18.23	89.36	20.40	100.00
Integral dose to normal tissue	9.04	88.98	9.44	92.91	10.16	100.00

delivered to the healthy tissue, especially to the left lung where the integral dose delivered is reduced by roughly 11% as compared to the margin solution. In fact, less than 0.6% of the tumor volume, on average, received less than 72 Gy, which makes this robust solution clinically acceptable. While the nominal solution performs the best in terms of dose minimization, it also allows a 6% average underdose to over 4% of the tumor volume. Furthermore, the worst-case realization of these 78 trials for

the nominal solution results in an 8% underdose (66.30 Gy), whereas the worst-case realization for the robust solution resulted in only a 1% underdose (71.23 Gy).

The results for the experiment that showed the greatest improvement in tumor coverage of the robust solution over the nominal solution are summarized in Table 2.5 (in experiment #2). The relevant DVHs are presented in Figures 2-15, 2-16, 2-17, and 2-18. In this experiment, the 276 pdfs from patients 1 – 4 were used to generate the uncertainty set, while the 52 pdfs from patient 5 were used to test the optimized solutions.

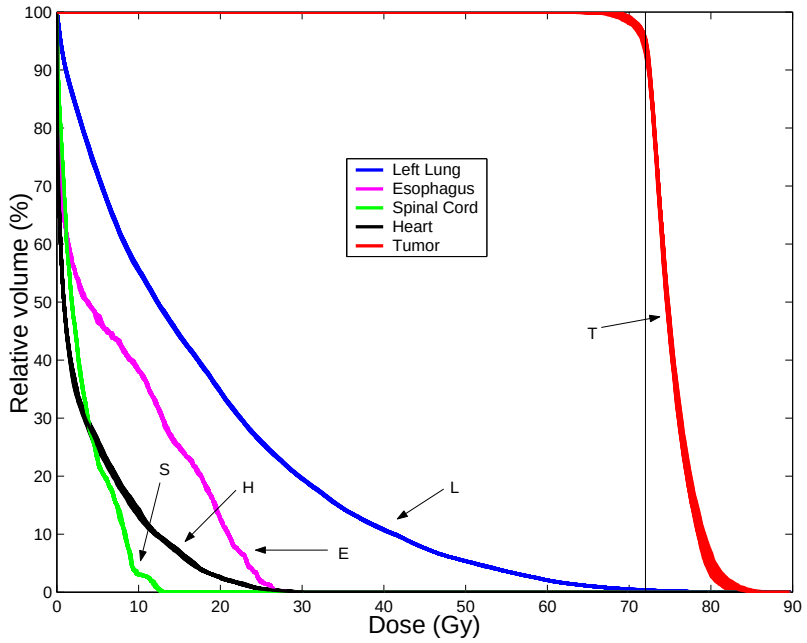


Figure 2-15: Nominal solution DVHs for 52 realized pdfs in experiment #2.

In experiment #2, note that the robust solution is able to maintain almost 100% tumor coverage (0.13% underdose to roughly 0.2% of the tumor) on average when tested with the 52 pdfs, whereas the nominal solution underdoses by almost 11% on average (to roughly 6% of the tumor). Comparing the nominal solution in this experiment (Figure 2-15) to the nominal solution of the previous experiment (Figure 2-11), we can see that there is even more variability in the parts of the tumor that receive the least and most dose. However, the robust solution still manages to retain a steep tumor DV-line (Figure 2-16) like in the previous experiment (Figure 2-12). It

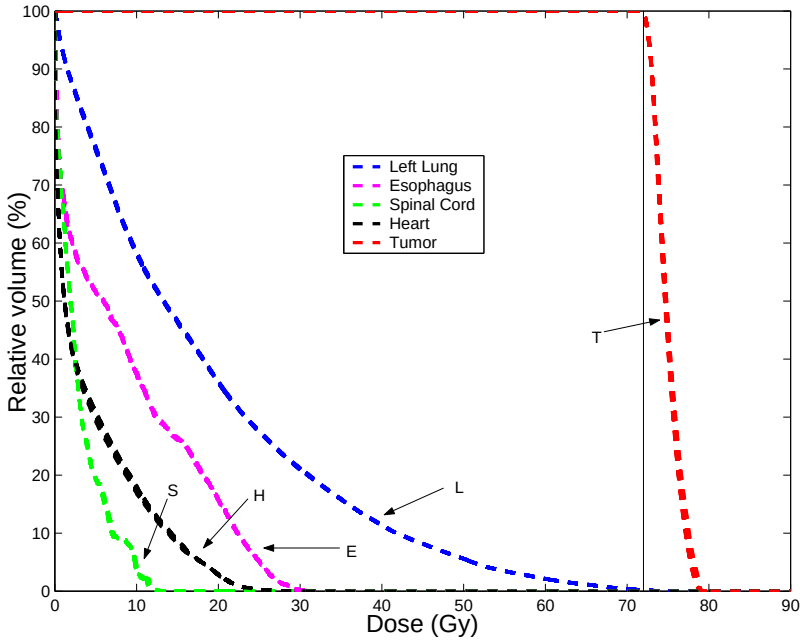


Figure 2-16: Robust solution DVHs for 52 realized pdfs in experiment #2.

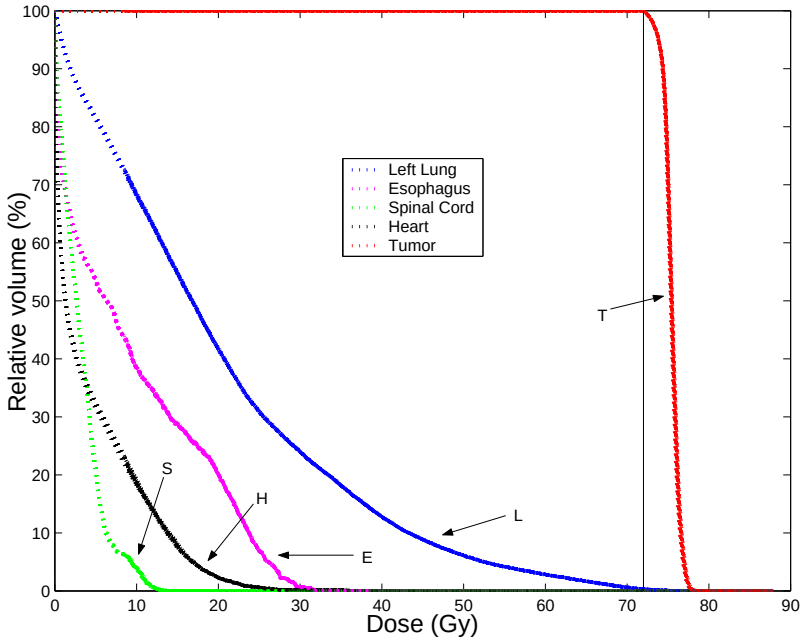


Figure 2-17: Margin solution DVHs for 52 realized pdfs in experiment #2.

is also important to recognize that the statistics related to healthy tissue and left lung dosage are very similar to those presented in Table 2.4 and 2.5. These observations

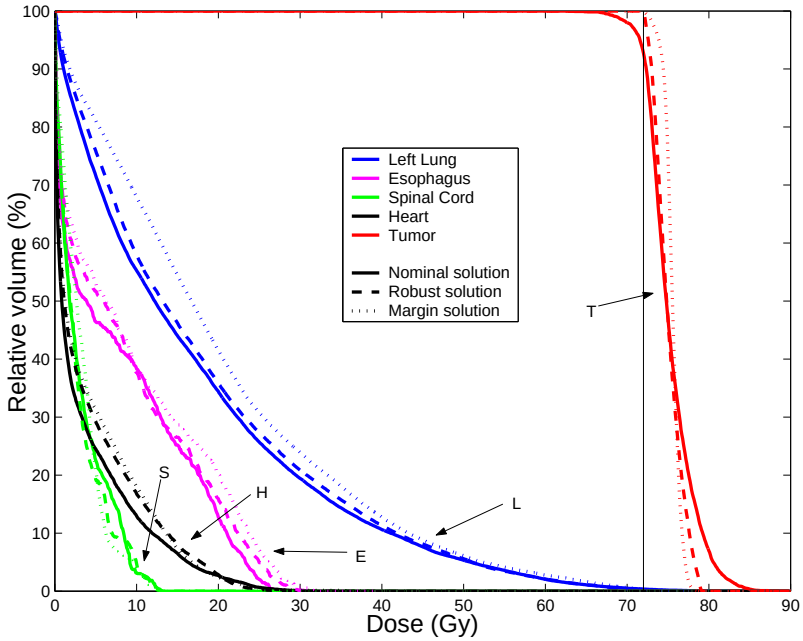


Figure 2-18: The nominal, robust, and margin solution DVHs for one realized pdf in experiment #2.

Table 2.5: Comparing the nominal, robust, and margin formulations using the average of 52 pdfs in experiment #2. This experiment showed the greatest difference between the nominal and robust solutions in terms of tumor coverage. The left subcolumns represent values in Gy, while the right subcolumns are percentages. The first row percentages are relative to 72 Gy, which is the desired dose level, while the next two rows are quoted as a percentage of the dose delivered by the margin solution.

	Nominal		Robust		Margin	
	Gy	%	Gy	%	Gy	%
Minimum dose delivered in tumor	64.26	89.25	71.91	99.87	72.05	100.07
Integral dose to left lung	17.18	85.11	18.02	89.27	20.18	100.00
Integral dose to normal tissue	8.99	88.92	9.39	92.89	10.11	100.00

are true of all the other experiments as well. Overall, these experiments suggest that, unlike nominal solutions, robust solutions can maintain consistently acceptable tumor coverage when tested with different realized pdfs, all other things being equal.

In accordance with the theory we presented, these results suggest that the robust solution combines the dose minimization capability of the nominal solution with the robustness of the margin solution. This suggests another angle from which to view

the robust formulation, namely, that of a multi-objective optimization perspective.

2.8 A Pareto viewpoint

The process of optimizing radiation therapy treatment plans inherently involves trading off the sparing of healthy tissue with the delivery of sufficient dose to the tumor. In fact, the robust formulation is facilitating this trade-off in a rigorous way. The robust formulation produces a solution that resides on the Pareto frontier corresponding to these two competing objectives. With these two objectives in mind, Figure 2-19 shows the relevant data points from Table 2.4 to illustrate this trade-off. By varying the uncertainty set from the singleton $\{p\}$ (the nominal pdf) to the most conservative set (where P_U is the unit simplex in $|X|$-dimensional Euclidean space), the robust formulation maps out the entire frontier of efficient solutions from the nominal solution to the margin solution, respectively. In effect, Figure 2-19 provides a rough bound on this surface using only three data points.

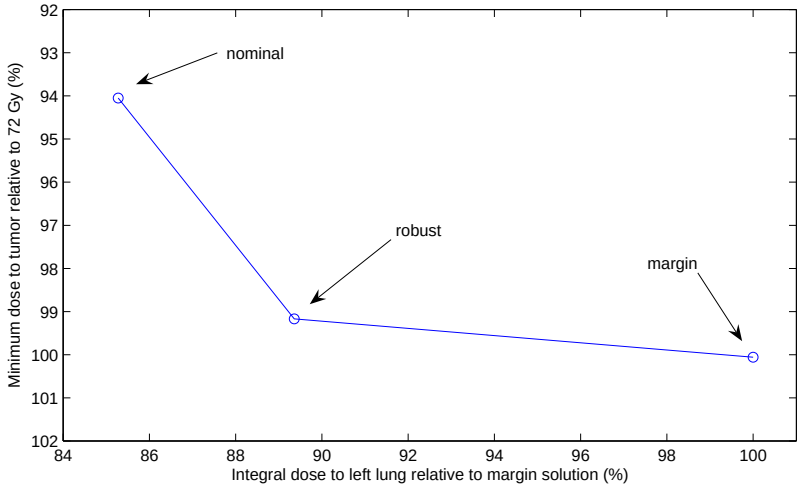


Figure 2-19: An approximation of the Pareto surface generated by the robust formulation.

By creating an appropriate uncertainty set for the patient at hand, solving the robust formulation will identify the point on the Pareto surface that bests trades off dose minimization with uncertainty management. This creates what we term

the "continuum of robustness," where an optimal robust solution corresponds to a particular uncertainty set in this continuum. Intuitively, a robust solution combines aspects of the nominal solution (i.e., intensity-modulation) and the margin solution (i.e., intensity-homogeneity) in order to produce the best possible treatment for the given uncertainty set. The upshot is that a treatment planner should focus on the creation of the uncertainty set, and "let the data do the talking." With a good model of uncertainty, the robust formulation will automatically find the appropriate solution on the Pareto surface for the patient at hand.

2.9 Conclusions

In this chapter, we developed a robust approach to the management of motion uncertainty in the optimization of intensity modulated radiation therapy treatments. We considered uncertainty in the probability density function describing motion and incorporated a suitable model of uncertainty into a linear optimization problem. Since the corresponding uncertainty set was a polyhedral set, the resulting robust counterpart remained a linear program. We showed that the robust formulation generalized current mathematical programming approaches to IMRT optimization, and compared the robust solution's performance to a nominal (aggressive solution based on complete knowledge of the underlying motion) and margin (conservative solution based on minimal knowledge of the underlying motion) solution on both a simplified one-dimensional problem and on a clinical lung case. In the 1D case, we saw that the robust formulation produced intensity maps and dose distributions that were insensitive to the uncertain motion, while minimizing the adverse effects to the healthy tissue. We will explore the structure of these solutions in Chapter 3 more thoroughly. In the more realistic 3D case, our experiments showed that while the nominal solution allowed anywhere between an unacceptable 6% to 11% underdose in the tumor, the robust solution consistently provided nearly the same level of protection as the margin solution, allowing around a 1% underdose on average. In addition, the robust solution also managed to consistently lower the dose delivered to the left lung

(the main organ-at-risk) by almost 11% as compared to the margin solution. Finally, we positioned our robust approach as a method to rigorously trade off competing objectives, and sketched the associated Pareto surface. While our development was motivated by lung cancer, this type of robust approach can be used for other sites and uncertainties as well (cf. Chapter 4).

In our numerical implementation, we assumed that variations of the breathing phases as seen in an external surrogate (the RPM signal) were highly correlated with internal breathing variations. As progress is made in image-guided treatments and more real-time data of internal anatomy movements are available, hopefully reliance on such an assumption will diminish. However, even image-guided treatments will not be error-free, so the need for robust methods to incorporate these measurements into the treatment planning routine will still exist. While we focused only on motion uncertainty in this chapter, there are clearly other sources of uncertainty that can benefit from a similar robust approach. Overall, future robust approaches should move beyond considering one type of uncertainty at a time, and try to integrate many sources of uncertainty (e.g., setup error, organ delineation error, etc.) into one comprehensive framework. This is a critical step in the process of moving towards clinical acceptance of robust methods.

Finally, we note that our approach essentially assumed that the radiation would be delivered in "one shot" as opposed to over many treatment fractions. By considering optimization on a day-to-day basis, it should be clear that our current "open-loop" policy will benefit from feedback acquired over the course of a treatment. Therefore, robustifying adaptive treatments based on data acquired throughout treatment is a relevant extension of the work presented here. In particular, for the motion uncertainty considered here, we could update our model of uncertainty on a daily basis as new data is acquired. This would allow us to fine tune our solution as the treatment progresses and would put more emphasis on the uncertainty that has been exhibited by the particular patient at hand.

Chapter 3

Optimal Parameterized Intensity Maps

In this chapter, we consider motion that is represented by a Gaussian distribution and derive the optimal intensity maps for two parameterized families of solutions. Our basic formulation is similar to the nominal formulation presented in Section 2.4.1. Namely, we require our solution to ensure tumor coverage in the presence of motion (and possibly motion uncertainty), while minimizing the total dose delivered.

The two specific types of intensity maps that we focus on in this chapter are margins and "edge-enhanced" maps. A margin is simply a volumetric expansion around the tumor that is essentially treated like a part of the tumor. That is, the anatomy corresponding to the margin is typically scheduled to receive the same amount of radiation as the tumor itself, to compensate for possible uncertainties in the location of the tumor due to errors in patient positioning or motion uncertainty. An edge-enhanced intensity map combines properties of a margin and intensity-modulation. More intensity is delivered at the edge of the tumor than at the middle of the tumor (which receives a uniform intensity like in a margin intensity map) in order to compensate for uncertainty. By increasing the intensity at the edge of the tumor, we may reduce the size of the margin around the tumor while maintaining tumor coverage.

In the previous chapter, we showed that edge-enhancements appear in robust solutions for intrafraction motion uncertainty management. Furthermore, since the

robust formulation (2.13) generalizes both the nominal formulation (2.3) and margin formulation (2.14), a robust solution can be thought of as specific a combination of intensity-modulation (from the nominal solution) and intensity-homogeneity (from the margin solution). This specific combination turns out to be an edge-enhanced solution in Section 2.6 (cf. Figure 2-9). Motivated by these observations from the previous chapter, we study the structure of margin and edge-enhanced solutions in this chapter.

Optimal margin sizes have been studied in [16, 57], however, these previous analyses focused on deriving margins that would be acceptable for a large fraction of a population of individuals using statistical methods. In this chapter, we focus on individual patients and present an analytical approach to determining optimal margin sizes for any instance of a particular problem. Furthermore, we extend our analysis to include robust formulations where the exact distribution of motion is not known. Edge-enhanced intensity maps have been witnessed in the medical physics literature in the context of interfraction motion compensation [56, 29] and penumbra (the fact that the x-ray beam is not perfectly sharp) compensation [46, 37, 7]. However, edge-enhancements have not been studied analytically, nor are there guidelines that prescribe the "size" of the edge-enhancements for specific tumor sizes and geometries. We aim to accomplish these tasks in this chapter.

Throughout the chapter, in order to facilitate the analysis, we assume infinitely sharp penumbra. That is, we assume that there is no lateral scatter of the radiation, so that if we delivered one unit of intensity to some area there would be no spillage of radiation outside of this area. This assumption can be partially mitigated by treating the Gaussian motion distribution as a distribution that takes both motion and a Gaussian penumbra into account. Since we are removing the effect of the scatter, we can think of intensity and dose as equivalent concepts, and therefore, we will use these terms interchangeably in this chapter.

In this chapter, our goal is to understand when a margin or edge-enhanced solution is optimal (minimizes dose delivered) in the presence of motion and motion uncertainty. Motion is described by a Gaussian distribution, while motion uncertainty is

described by uncertainty in the parameters, μ and σ, of the Gaussian. We derive simple, analytical relationships between the problem parameters that characterize the structure of the resulting optimal solutions. In the sections that focus on margins, our results tell us exactly *when* to use a margin and, by solving a particular equation, exactly *what* the margin size should be. These determinations rely only upon the basic problem data such as the size of the tumor and the standard deviation of the Gaussian distribution. We make similar determinations in the sections focused on edge-enhancements.

The remainder of this chapter is organized as follows. In Sections 3.1 we derive optimal margins for a one-dimensional geometry in the presence of Gaussian motion, and extend the analysis to the case where the parameters of the Gaussian are uncertain in Section 3.2. In Sections 3.3 and 3.4, we extend the analysis of the previous two sections to a three-dimensional tumor. In Sections 3.5 and 3.6 we generalize our analyses for both the non-robust and robust versions, respectively, of a one-dimensional tumor when the optimal intensity maps are restricted to edge-enhanced solutions.

3.1 Optimal margins

In this section, we derive optimal margin sizes for a one-dimensional geometry. In the presence of motion, the dose delivered is blurred. Mathematically, this is equivalent to the convolution of the motion pdf (a Gaussian distribution) with the delivered intensity. Our goal is to find an intensity map that minimizes the total dose delivered while ensuring that sufficient dose is delivered to the tumor even after the convolution with a Gaussian pdf. We will restrict our search for optimal solutions to the class of intensity maps that consists of a uniform (i.e., isotropic) margin around an intensity map that has uniform intensity.

Consider a one-dimensional tumor of length t that requires one unit of dose throughout. In the absence of motion, the optimal intensity map is a "square wave" intensity map with a height of one and a length of t centered on the middle of the tumor. For now, let us focus on intensity maps that are centered on the tumor. We will

refer to this particular intensity map as the static dose distribution or static intensity map. Convolving the static intensity map with a Gaussian pdf results in a function that has a height of strictly less than one. In order to improve tumor coverage, we have two options. First, we can simply increase the intensity of our static intensity map. That is, instead of delivering one unit of intensity over the entire tumor, we deliver strictly more than one unit. Second, we can add a uniform margin around the tumor, which is equivalent to adding a margin to the static intensity map. Then, instead of delivering one unit of dose to an interval of length t, we deliver one unit of dose to an interval of length $t + 2m$, where m is the "size" of the margin. We assume that a margin of length m is added to both the right and left sides of the static intensity map. We may also use a combination of these two strategies. We can add a margin of size m and then strictly increase the delivered intensity over the entire intensity map of length $t + 2m$. Figure 3-1 illustrates these concepts.

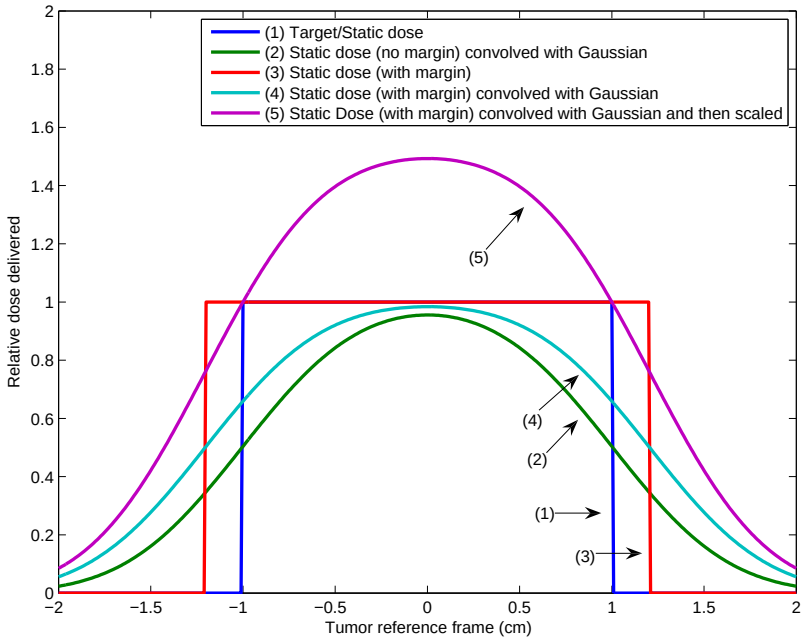


Figure 3-1: Dose delivered using a margin and scaling.

Note that since the Gaussian distribution has an infinite support, the height of the resulting function after convolving the Gaussian with the static intensity map will always be less than the height of the static intensity map. Therefore, regardless of

whether or not a (finite) margin is used, there will always be some positive scaling of the intensity in order to ensure the tumor receives one unit of dose throughout. At the end of this section, we will show that a positive scaling of the intensity may not be necessary if the underlying motion pdf has a finite support.

3.1.1 Summary and Highlights

In this section, we show that the ratio t/σ is necessary and sufficient to determine the structure of the optimal margin intensity map. In particular, when $0 \leq t/\sigma \leq 2.281$, the optimal intensity map involves no margin, and simply a scaling of the static intensity map. On the other hand, when $t/\sigma > 2.281$, the optimal margin size is positive and equal to the unique solution of a particular nonlinear equation. This threshold result is consistent with physical intuition, since a small value of t/σ in practice indicates that the tumor is small, and in the treatment of small tumors it is more common to simply boost the intensity of the static intensity map without incorporating a margin.

3.1.2 Analysis

First, let us establish the intuitive fact that the optimal intensity map will always be centered on the tumor. Let $\mathcal{T}(\mu)$ be the set of voxels of the tumor when it is centered at μ and let $r(x)$ be the dose that point x requires. Then,

$$\mathcal{T}(\mu) = [-t/2 + \mu, t/2 + \mu],$$

and

$$r(x) = \begin{cases} 1, & x \in \mathcal{T}(\mu), \\ 0, & \text{otherwise.} \end{cases}$$

We define the static intensity map with a margin of size m, centered at the origin, to be

$$\delta_m(x) = \begin{cases} 1, & x \in [-t/2 - m, t/2 + m], \\ 0, & \text{otherwise.} \end{cases}$$

Let us define $d_m(y; \mu, \sigma)$ to be the dose delivered to point y, when a static intensity map with margin size m is convolved with a Gaussian distribution (denoted $p(x)$) with mean μ and standard deviation σ. Hence,

$$\begin{aligned} d_m(y; \mu, \sigma) &= \int_{-\infty}^{\infty} \delta_m(y - x) p(x) dx \\ &= \int_{y-t/2-m}^{y+t/2+m} p(x) dx \\ &= \Phi\left(\frac{y + t/2 + m - \mu}{\sigma}\right) - \Phi\left(\frac{y - t/2 - m - \mu}{\sigma}\right), \end{aligned}$$

where Φ is the cumulative distribution function of the standard normal random variable.

Lemma 3.1. *The function $d_m(y; \mu, \sigma)$ is quasiconcave in y for all $y \in \mathbb{R}$.*

Proof: In one-dimension, quasiconcavity is equivalent to unimodality. The Gaussian pdf is strongly unimodal [21]; that is, when it is convolved with a unimodal function (such as $\delta_m(y)$), the result is unimodal. Thus, $d_m(y; \mu, \sigma)$ is quasiconcave. □

Proposition 3.1. *In the presence of motion described by a Gaussian pdf with mean μ and standard deviation σ, the optimal intensity map will be centered at μ.*

Proof: Suppose the tumor is centered at some $\mu \neq 0$, but the dose delivered is centered at the origin and is described by

$$d_m(y; 0, \sigma) = \Phi\left(\frac{y + t/2 + m}{\sigma}\right) - \Phi\left(\frac{y - t/2 - m}{\sigma}\right).$$

Since the function $d_m(y; 0, \sigma)$ is a quasiconcave function of y, the minimum dose delivered to the tumor occurs either at the right edge ($y = \mu + t/2$) or left edge

$(y = \mu - t/2)$ of the tumor. Therefore,

$$\begin{aligned}\min_{y\in\mathcal{T}(\mu)} d(y;0,\sigma) &= \min\{d_m(\mu + t/2;0,\sigma), d_m(\mu - t/2;0,\sigma)\} \\ &= \min\left\{\Phi\left(\frac{\mu+t+m}{\sigma}\right) - \Phi\left(\frac{\mu-m}{\sigma}\right),\right. \\ &\qquad \left.\Phi\left(\frac{\mu+m}{\sigma}\right) - \Phi\left(\frac{\mu-t-m}{\sigma}\right)\right\}.\end{aligned}$$

In order to achieve tumor coverage with the lowest total dosage, we need to scale the intensity map by exactly $1/\min_{y\in\mathcal{T}(\mu)} d(y;0,\sigma)$. This scaling is minimized when the denominator is largest, that is, when

$$\Phi\left(\frac{\mu+t+m}{\sigma}\right) - \Phi\left(\frac{\mu-m}{\sigma}\right) = \Phi\left(\frac{\mu+m}{\sigma}\right) - \Phi\left(\frac{\mu-t-m}{\sigma}\right),$$

which occurs only when $\mu = 0$, due to the symmetry of the Gaussian pdf. This implies that the beam of radiation should always be centered on the tumor. □

From here on, we may assume that $\mu = 0$, since knowing the motion distribution lets us shift the beam of radiation so that it is aligned with the center of the tumor. In the remainder of this section, our goal is to derive the optimal margin size for any problem instance (a given t and σ).

With a margin of size m, the total dose delivered before scaling the intensity map is $t + 2m$. Let γ define the scaling factor variable. Thus, the optimization problem that finds an intensity map which ensures tumor coverage at lowest total dosage can be written as:

$$\begin{aligned}\underset{m,\gamma}{\text{minimize}} \quad & \gamma(t+2m) \\ \text{subject to} \quad & \gamma d_m(y;0,\sigma) \geq 1, \quad \forall\, y \in [-t/2, t/2], \\ & m, \gamma \geq 0.\end{aligned} \tag{3.1}$$

For a given margin size m, let us define γ_m^* to be the optimal (minimal) scaling factor. In the course of proving Proposition 3.1, we have determined a closed form expression

for γ_m^*.

Corollary 3.1. *The quantity γ_m^* is equal to*

$$\gamma_m^* = \frac{1}{\Phi(\frac{t+m}{\sigma}) - \Phi(\frac{-m}{\sigma})} = \frac{1}{\Phi(\frac{m}{\sigma}) - \Phi(\frac{-t-m}{\sigma})}.$$

Using Corollary 3.1, we may simplify formulation (3.1) to the following problem

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & f(m) \\ \text{subject to} \quad & m \geq 0, \end{aligned} \tag{3.2}$$

where

$$f(m) := \gamma_m^*(t + 2m) = \frac{t + 2m}{\Phi(\frac{m+t}{\sigma}) - \Phi(\frac{-m}{\sigma})} \tag{3.3}$$

Without loss of generality, we may also assume $\sigma = 1$, however, we will retain σ explicitly to emphasize its relationship with the tumor and margin sizes. These relationships will be central in the analysis that follows. First, we characterize the structure of the function $f(m)$.

Lemma 3.2. *Let*

$$g(m) := \Phi\left(\frac{m + t}{\sigma}\right) - \Phi\left(\frac{-m}{\sigma}\right).$$

Then, for any $t > 0$ and $\sigma > 0$, $g(m)$ is a positive, increasing, and strictly concave function of m, with $g''(m) < 0$, for $m \geq 0$.

Proof: By definition, $g(m)$ equals the probability of a standard normal random variable lying between $[-m/\sigma, (m + t)/\sigma]$, which is clearly positive for $m \geq 0$. The first two derivatives of $g(m)$ are

$$g'(m) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(m+t)^2}{2\sigma^2}} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(-m)^2}{2\sigma^2}}$$

and

$$g''(m) = -\frac{m + t}{\sigma^3} \frac{1}{\sqrt{2\pi}} e^{-\frac{(m+t)^2}{2\sigma^2}} - \frac{m}{\sigma^3} \frac{1}{\sqrt{2\pi}} e^{-\frac{(-m)^2}{2\sigma^2}},$$

which are strictly positive and negative, respectively, for $m \geq 0$. □

Proposition 3.2. *For any $t > 0$ and $\sigma > 0$, $f(m)$ has a unique local minimum, m^*, for $m \geq 0$.*

Proof: Let $g(m)$ be as defined in Lemma 3.2. The first two derivatives of $f(m)$ are

$$f'(m) = \frac{2g(m) - (t + 2m)g'(m)}{g^2(m)}$$

and

$$f''(m) = \frac{-(t + 2m)g''(m)g^2(m) - (2g(m) - (t + 2m)g'(m))2g(m)g'(m)}{g^4(m)}.$$

Note that $f(m)$ is continuous and differentiable ($f'(m)$ and $f''(m)$ exist and are finite for all $m \geq 0$). Also, $f(m) > 0$ for all $m \geq 0$, and in particular, $\lim_{m\to\infty} f(m) = \infty$. The remainder of the proof will be different depending on whether or not there exists a positive root of $f'(m)$.

Suppose there exists an $m^* > 0$ that satisfies $f'(m^*) = 0$. Then m^* is a strict local minimum since

$$f''(m^*) = \frac{-(t + 2m^*)g''(m^*)}{g^2(m^*)} > 0$$

for $m \geq 0$ (recall $g''(m) < 0$ for all $m \geq 0$ from Lemma 3.2). Furthermore, we claim that if there exists any $m^* > 0$ satisfying $f'(m^*) = 0$, it must be unique. Suppose this is false. Then let $\hat{m} \neq m^*$ be another strict local minimum. Without loss of generality, let $\hat{m} < m^*$. Since both $\hat{m}$ and m^* are strict local minima, for $\epsilon > 0$ sufficiently small, $f'(\hat{m}+\epsilon) > 0$ and $f'(m^*-\epsilon) < 0$. Then, because $f'(m)$ is continuous, the intermediate value theorem implies the existence of a local maximum between $\hat{m}$ and m^*, which contradicts the fact that any stationary point of $f(m)$ is a strict local minimum.

Finally, we show that if there is no $m > 0$ such that $f'(m) = 0$, then $f'(m) \geq 0$ for all $m > 0$. Suppose to the contrary that $f'(\hat{m}) < 0$ for some $\hat{m} > 0$. Since $g(m) \to 1$ and $(t + 2m)g'(m) \to 0$ as $m \to \infty$,

$$\lim_{m\to\infty} f'(m) = 2 > 0.$$

The intermediate value theorem implies the existence of a positive root of $f'(m)$ in the interval $(\hat{m}, \infty)$, which is a contradiction. Thus, in this case, the unique minimum of $f(m)$ occurs at $m^* = 0$. $\square$

The next result builds intuitively from the proof of the previous proposition. The result states that the optimal margin size, m^*, is determined by the sign of $f'(0)$.

Proposition 3.3. *If $f'(0) \geq 0$, then $m^* = 0$. If $f'(0) < 0$, then m^* is the unique solution to the equation*

$$2\left(\int_{\frac{-m}{\sigma}}^{\frac{m+t}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right) - (t+2m)\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(m+t)^2}{2\sigma^2}} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{m^2}{2\sigma^2}}\right) = 0, \qquad (3.4)$$

and $m^* > 0$.

Proof: First, let us consider the case where $f'(0) \geq 0$. Note that if $f'(0) = 0$, then $f''(0) = -tg''(0)/g^2(0) > 0$. Thus, if $f'(0) \geq 0$, $f(\epsilon) > f(0)$ for all $\epsilon > 0$ sufficiently small. To establish the first part of the proposition, suppose that $f'(\hat{m}) < 0$ for some $\hat{m} > 0$. Since $f'(m)$ is continuous, the intermediate value theorem implies the existence of a local maximum between the 0 and $\hat{m}$, which contradicts the fact that all stationary points of $f(m)$ are strict local minima (established in the proof of the previous proposition). Hence, if $f'(0) \geq 0$, then $f'(m) \geq 0$ for all $m \geq 0$, and $f(m)$ is minimized at $m^* = 0$.

On the other hand, if $f'(0) < 0$, then again the intermediate value theorem guarantees the existence of a local minimum at some $m^* > 0$ since $\lim_{m\to\infty} f'(m) > 0$ (established in the proof of the previous proposition). From the previous proposition, this m^* is the unique solution of the equation $f'(m) = 0$, which is equivalent to equation (3.4) in the proposition statement (excluding the denominator of $f'(m)$). $\square$

Finally, we present the main result of this section. This result derives an explicit relationship between the fraction t/σ and a fixed number that characterizes the structure of the optimal intensity map for the instance of the problem with tumor size t

and Gaussian motion standard deviation σ. For convenience, let us define $\phi(x)$ to be the probability density function of a standard normal random variable. That is,

$$\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Theorem 3.1. *Let u^* be the unique positive solution to the equation*

$$\left(\int_{-u}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right) - \frac{u}{\sqrt{2\pi}} \left(1 + e^{-\frac{u^2}{2}} \right) = 0. \tag{3.5}$$

If $t/\sigma \leq u^*$, then the optimal margin size is $m^* = 0$. If $t/\sigma > u^*$, then the optimal margin size, m^*, is the unique solution to (3.4)*, and is positive.*

Proof: Let $u = t/\sigma$ in the expression for $f'(0)$. Then the numerator of $f'(0)$, as a function of u, is equal to a function $h(u)$ defined by

$$h(u) := \left(\int_{-u}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right) - \frac{u}{\sqrt{2\pi}} \left(1 + e^{-\frac{u^2}{2}} \right). \tag{3.6}$$

Note that $h(0) = 0$. First, we need to show that there exists only one positive solution, u^*, to the equation $h(u) = 0$. Then we show that $h(u)$ is positive only on the interval $(0, u^*)$. To show the existence of a solution to the equation $h(u) = 0$, we first note that $h(u)$ is a continuous function of u and $h(u) \to -\infty$ as $u \to \infty$. Then, it suffices to show that the lowest order derivative of $h(u)$ that is not equal to zero at $u = 0$, is positive. Using the relation

$$\phi'(x) = -x\phi(x),$$

we have

$$\lim_{u \to 0} h'(u) = \lim_{u \to 0} (u^2 + 1)\phi(u) - \phi(0) = 0$$

$$\lim_{u \to 0} h''(u) = \lim_{u \to 0} -u(u^2 - 1)\phi(u) = 0$$

$$\lim_{u \to 0} h^{(3)}(u) = \lim_{u \to 0} (u^4 - 4u^2 + 1)\phi(u) > 0.$$

Invoking the intermediate value theorem implies the existence of some $u^* > 0$ such

that $h(u^*) = 0$.

Next, we show that this u^* is unique. The point $u = 1$ is the only positive root of the function $h''(u)$. Furthermore, $h''(u)$ is positive for $0 < u < 1$ and negative for $u > 1$. This implies that $h'(u)$ is increasing for $0 < u \leq 1$ and decreasing for $u > 1$. Hence, $h'(u)$ is unimodal. Since $h'(0) = 0$, $h^{(3)}(0) > 0$ and

$$\lim_{u \to \infty} h'(u) = -\phi(0) < 0,$$

the unimodality of $h'(u)$ implies that $h'(u)$ has a unique root for $u > 0$. Let us call this root $\overline{u}$. Furthermore, it is straightforward to see that $h'(u)$ is positive on the interval $(0, \overline{u})$ and negative for $u > \overline{u}$. By repeating the above argument, we can show that $h(u)$ is unimodal with a unique root for $u > 0$, namely, u^*. Furthermore, $h(u)$ is positive in the interval $(0, u^*)$ and negative for $u > u^*$.

Thus, for $0 \leq u \leq u^*$, $h(u) \geq 0$, which is equivalent to the inequality $f'(0) \geq 0$. Hence, in this case, $m^* = 0$. On the other hand, if $u > u^*$, then $h(u) < 0$ and $f'(0) < 0$. □

The following corollary has been proved in the course of proving Theorem 3.1, but we will state the result here since it will serve as a useful reference later in the chapter. Geometrically, this result relates the slope of the chord between $\Phi(0)$ and $\Phi(u)$ to equally weighted slopes of the tangent lines to $\Phi(x)$ at $x = 0$ and $x = u$.

Corollary 3.2. *Let u^* be defined as in Theorem 3.1. For $0 \leq u \leq u^*$,*

$$\Phi(u) - \Phi(-u) - u(\phi(u) + \phi(0)) \geq 0,$$

with equality at $u = u^*$, and for $u > u^*$,

$$\Phi(u) - \Phi(-u) - u(\phi(u) + \phi(0)) < 0.$$

Unfortunately, there is no closed form expression for u^*, but it can be computed numerically to be approximately $u^* = 2.281$. Theorem 3.1 states that if the ratio

of the tumor size is no more than 2.281 times the standard deviation of the motion, then the optimal margin size is 0. With a margin size of zero, tumor coverage is achieved purely through a scaling of the intensity. It is important to note that this result reaffirms conventional wisdom from the medical physics community. For small tumors, intuition suggests that adding a margin around the tumor is less desirable than simply increasing the intensity of the static intensity map. If $t/\sigma < u^*$, then we can think of the tumor as being roughly the same size as the magnitude of the motion, which for realistic amplitudes, implies that the tumor is small. The motion amplitude of a lung tumor depends on many variables, but for a rough calculation, we can assume the standard deviation of the motion amplitude to be roughly 0.5 cm ([26] summarizes the findings of many studies on organ motion). If $\sigma = 0.5$, then $t/\sigma < u^*$ implies that the tumor is roughly 1 cm or smaller, which is considered to be small. On the other hand, if $t/\sigma > u^*$, then the motion can be thought of as relatively small compared to the size of the tumor, and therefore, adding a small margin around the tumor will only contribute a small increase in the dose delivered to the healthy tissue.

Figure 3-2 depicts the relationship between the optimal margin size and the tumor size, while Table 3.1 gives the values of m^* and the corresponding scaling factor, $\gamma^*_{m^*}$ for particular choices of t.

Table 3.1: The optimal margin size (m^*) for particular values of the tumor size (t), relative to the size of the standard deviation of the Gaussian pdf ($\sigma = 1$ in this case). The corresponding intensity scaling factors ($\gamma^*_{m^*}$) are also shown.

t	1	2	3	4	5	6	7	8	9	10
m^*	0	0	0.539	0.835	1.007	1.129	1.225	1.304	1.371	1.429
$\gamma^*_{m^*}$	2.930	2.095	1.418	1.253	1.186	1.149	1.124	1.106	1.093	1.083

Table 3.1 shows that as the tumor size increases, the optimal margin size increases and the intensity scaling decreases. This makes sense since adding a larger margin reduces the underdose to the tumor, and therefore, reduces the intensity scaling needed to provide tumor coverage. Note that in practical situations where the motion distribution cannot have infinite support, an optimal solution may be just a positive

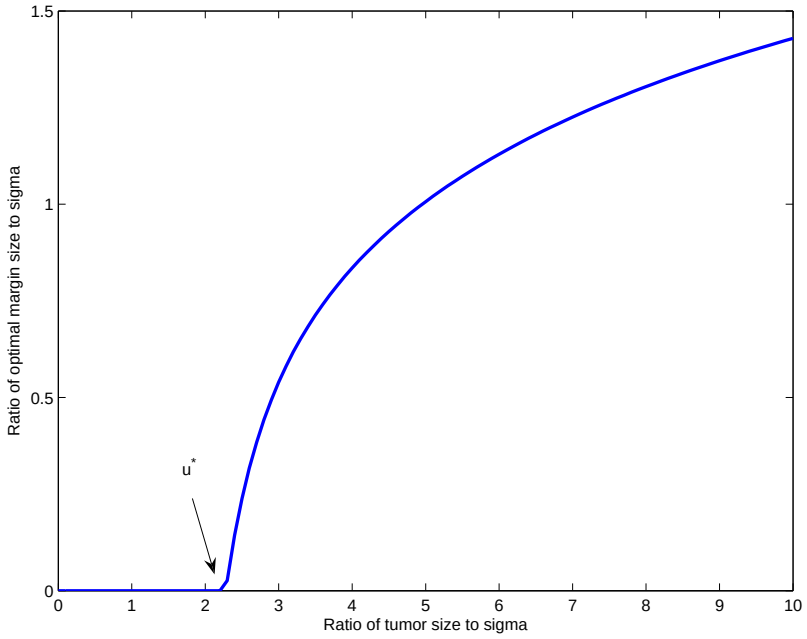


Figure 3-2: The ratio of the optimal margin size to σ as a function of the ratio of the tumor size to σ for a 1D tumor.

margin without any intensity scaling ($\gamma^*_{m^*} = 1$). In our framework, this situation appears (approximately) when the standard deviation is much smaller than the size of the tumor, in which case the intensity scaling is small (e.g., in the case $t = 10$ in Table 3.1). The following is an example that shows that the optimal intensity map may involve a margin with no scaling of the intensity, when the motion pdf has a finite support.

Example 3.1. *Consider a tumor centered at the origin and let* $\mathcal{T}(0) = [-1, 1]$. *If the motion pdf is*

$$p(x) = \begin{cases} 1, & x \in [-1/2, 1/2], \\ 0, & \textit{otherwise}, \end{cases}$$

and we deliver a unit intensity to $\mathcal{T}(0)$ *and the margin, described by the static intensity map*

$$\delta_m(x) = \begin{cases} 1, & x \in [-1-m, 1+m], \\ 0, & \textit{otherwise}, \end{cases}$$

then

$$\begin{aligned} d_m(y) &:= \int_{-\infty}^{\infty} \delta_m(y-x)p(x)dx \\ &= \int_{\max\{y-1-m,-1/2\}}^{\min\{y+1+m,1/2\}} 1dx \\ &= \begin{cases} y+3/2+m, & \text{if } y \le -1/2-m, \\ 1, & \text{if } -1/2-m \le y \le 1/2+m, \\ 3/2+m-y, & \text{if } y \ge 1/2+m. \end{cases} \end{aligned}$$

Note that we may restrict m to be between 0 and 1/2 since we can achieve tumor coverage with any margin of size at least 1/2 (and no intensity scaling). The minimum of $d_m(y)$ over $y \in \mathcal{T}(0)$ occurs at $y = \pm 1$, and at those points, $d_m(-1) = d_m(1) = 1/2 + m$. Hence, the minimal intensity scaling factor is

$$\gamma_m^* = \frac{1}{1/2+m}$$

and the total dose delivered is

$$f(m) := \frac{2+2m}{1/2+m}.$$

The derivative of $f(m)$,

$$f'(m) = \frac{-1}{(1/2+m)^2},$$

is negative for all $m \ge 0$. Thus, the optimal solution is to have a positive margin of size 1/2, and then no intensity scaling is needed.

In the next section, we extend our analysis to the case where there is uncertainty in the parameters of the Gaussian motion distribution.

3.2 Optimal margins with robustness

In this section, we consider uncertainty in the parameters, μ and σ, of the Gaussian pdf and investigate the effects of these uncertainties on the determination of the optimal margin size. Our goal is to derive optimal robust solutions. An optimal robust solution remains feasible (i.e., guarantees tumor coverage) under all realizations of the uncertainty, while minimizing the total dose delivered.

Let us assume that the uncertain parameters, denoted $\tilde{\mu}$ and $\tilde{\sigma}$, lie in the intervals $[\underline{\mu}, \overline{\mu}]$ and $[\underline{\sigma}, \overline{\sigma}]$, respectively. These intervals characterize the range of possible realizations of the uncertain parameters. We will solve a problem similar to the one posed in the nominal formulation (3.1). However, in the robust version, we would like to ensure feasibility for all possible realizations of $\tilde{\mu}$ and $\tilde{\sigma}$.

3.2.1 Summary and Highlights

In this section, we show that the mathematical formulation that incorporates the uncertainty in μ and σ has a deterministic equivalent, and that this deterministic "robust counterpart" has exactly the same structure as the underlying nominal (no uncertainty) formulation. As a result, the threshold that separates the case when the optimal margin size is positive or zero remains the same as in the nominal problem. The threshold is 2.281, as in the previous section, however, we do not use the ratio t/σ anymore. Instead, the ratio we consider is $\bar{t}/\overline{\sigma}$, where $\bar{t} = t + \overline{\mu} - \underline{\mu}$. In the robust problem, we have replaced the nominal values of the tumor size and standard deviation of motion, with appropriate values for the robust case.

From this analysis, we derive the intuition that larger uncertainty in μ will drive the solution towards an intensity map with a positive margin size (since $\bar{t}/\overline{\sigma}$ will be larger), whereas a larger $\overline{\sigma}$ will drive the solution towards a pure intensity scaling. We also provide examples to show how a robust solution generated from the robust counterpart outperforms an intuitive heuristic based on a family of nominal solutions.

3.2.2 Analysis

First, let us analyze the uncertainty in μ. If there is no uncertainty in this parameter, we showed in Proposition 3.1 that the optimal solution would involve a beam that is centered at the middle of the tumor. Under the current setup, $\tilde{\mu} \in [\underline{\mu}, \overline{\mu}]$, so we no longer have one fixed value for μ. However, we can essentially recover the nominal (fixed μ) situation by noting that the requirement of tumor coverage over all $\tilde{\mu} \in [\underline{\mu}, \overline{\mu}]$ is equivalent to requiring tumor coverage for a tumor that is $\overline{\mu} - \underline{\mu}$ units larger than t. Thus, if we center the beam at $(\overline{\mu} + \underline{\mu})/2$ and shift the origin to this location, the uncertainty in μ implies that we need to provide tumor coverage at all points y in the interval $[-t/2 - (\overline{\mu} - \underline{\mu})/2, t/2 + (\overline{\mu} - \underline{\mu})/2]$. Let us define the "effective tumor" centered at the origin to be

$$\overline{\mathcal{T}}(0) := [-t/2 - (\overline{\mu} - \underline{\mu})/2, t/2 + (\overline{\mu} - \underline{\mu})/2].$$

This is the region that needs to be irradiated to the desired dose level.

We now pose our robust optimization problem as

$$\begin{array}{ll} \underset{m,\gamma}{\text{minimize}} & \gamma(t + 2m) \\ \text{subject to} & \gamma d_m(y; 0, \tilde{\sigma}) \geq 1, \quad \forall\, y \in \overline{\mathcal{T}}(0), \forall\, \tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}] \\ & m, \gamma \geq 0. \end{array} \tag{3.7}$$

The following result establishes that the robust counterpart to formulation (3.7) has the same structure as the nominal formulation (3.1). Hence, we may find a robust solution simply by solving a modified version of the nominal formulation.

Theorem 3.2. *Formulation* (3.7) *is equivalent to the following mathematical program*

$$\begin{array}{ll} \underset{m}{\text{minimize}} & \dfrac{\bar{t} + 2m}{\Phi(\frac{m+\bar{t}}{\overline{\sigma}}) - \Phi(\frac{-m}{\overline{\sigma}})} \\ \text{subject to} & m \geq 0, \end{array} \tag{3.8}$$

where $\bar{t} = t + \overline{\mu} - \underline{\mu} > t$.

Proof: Delivering the static intensity map

$$\delta_m(x) = \begin{cases} 1, & x \in [-t/2 - m - (\overline{\mu} - \underline{\mu})/2, t/2 + m + (\overline{\mu} - \underline{\mu})/2], \\ 0, & \text{otherwise}, \end{cases}$$

results in the dose delivered, for a given $\tilde{\sigma}$, to be

$$\begin{aligned} d_m(y; 0, \tilde{\sigma}) &= \int_{-\infty}^{\infty} \delta_m(y - x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\tilde{\sigma}^2}} dx \\ &= \Phi\left(\frac{y + m + t/2 + (\overline{\mu} - \underline{\mu})/2}{\tilde{\sigma}}\right) - \Phi\left(\frac{y - m - t/2 - (\overline{\mu} - \underline{\mu})/2}{\tilde{\sigma}}\right). \end{aligned}$$

Since the dose now depends on $\tilde{\sigma}$, the optimal scaling factor, γ_m^*, is

$$\begin{aligned} \gamma_m^* &= \max_{y \in \overline{\mathcal{T}}(0), \tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}]} \frac{1}{d_m(y; 0, \tilde{\sigma})} \\ &= \frac{1}{\min_{y \in \overline{\mathcal{T}}(0), \tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}]} d_m(y; 0, \tilde{\sigma})} \\ &= \frac{1}{\min_{\tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}]} d_m(t/2 + (\overline{\mu} - \underline{\mu})/2; 0, \tilde{\sigma})} \\ &= \frac{1}{\min_{\tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}]} \left(\Phi\left(\frac{t + m + \overline{\mu} - \underline{\mu}}{\tilde{\sigma}}\right) - \Phi\left(\frac{-m}{\tilde{\sigma}}\right)\right)}, \end{aligned}$$

where the third equality comes from the quasiconcavity of $d_m(y; 0, \tilde{\sigma})$ for any $\tilde{\sigma}$. Finally, $\Phi(x/\sigma)$ as a function of σ is decreasing if $x > 0$ and increasing if $x < 0$. This is straightforward to show since

$$\frac{d}{d\sigma} \Phi\left(\frac{x}{\sigma}\right) = \phi\left(\frac{x}{\sigma}\right)\left(-\frac{x}{\sigma^2}\right).$$

Thus,

$$\begin{aligned} \gamma_m^* &= \frac{1}{\min_{\tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}]} \left(\Phi\left(\frac{t + m + \overline{\mu} - \underline{\mu}}{\tilde{\sigma}}\right) - \Phi\left(\frac{-m}{\tilde{\sigma}}\right)\right)} \\ &= \frac{1}{\Phi\left(\frac{t + m + \overline{\mu} - \underline{\mu}}{\overline{\sigma}}\right) - \Phi\left(\frac{-m}{\overline{\sigma}}\right)}, \end{aligned}$$

and the total dose delivered, as a function of m, is

$$\frac{t + 2m + \overline{\mu} - \underline{\mu}}{\Phi\left(\frac{t+m+\overline{\mu}-\underline{\mu}}{\overline{\sigma}}\right) - \Phi\left(\frac{-m}{\overline{\sigma}}\right)}.$$

□

From this analysis, an underlying relationship between robustness and the type of uncertainty is apparent. Uncertainty in $\tilde{\mu}$ effectively increases the size of the tumor (area to be covered with sufficient dose), while the largest possible value of $\tilde{\sigma}$ affects the spread in the realized dose distribution. Thus, to a first approximation, it seems like a margin is an appropriate compensatory tool for dealing with uncertainty in $\tilde{\mu}$, while increasing the intensity of the beam can offset an increased $\tilde{\sigma}$. Of course, taken in concert, these factors affect both the margin size and intensity scaling, and it is the relative "size" of these factors that determine whether the robust solution will be driven towards a margin solution or a pure intensity scaling. Specifically, when referring to the relative size of these factors, we are interested in the value of the ratio $\overline{t}/\overline{\sigma}$ in relation to u^*. Our intuition is made concrete in Theorem 3.3, which replicates the results of Theorem 3.1 for the robust counterpart (3.8). The proof is omitted since it is essentially the same as the proof of Theorem 3.1, but with t replaced by $\overline{t}$ and σ replaced by $\overline{\sigma}$.

Theorem 3.3. *Let u^* be as defined in Theorem 3.1. If $\overline{t}/\overline{\sigma} \leq u^*$, then the optimal margin size is $m^* = 0$. If $\overline{t}/\overline{\sigma} > u^*$, then m^* is the unique solution to*

$$2\left(\int_{\frac{-m}{\overline{\sigma}}}^{\frac{m+\overline{t}}{\overline{\sigma}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right) - (\overline{t} + 2m)\left(\frac{1}{\overline{\sigma}\sqrt{2\pi}} e^{-\frac{(m+\overline{t})^2}{2\overline{\sigma}^2}} + \frac{1}{\overline{\sigma}\sqrt{2\pi}} e^{-\frac{m^2}{2\overline{\sigma}^2}}\right) = 0$$

and $m^* > 0$.

From Theorem 3.3, it is apparent that $\tilde{\mu}$ and $\tilde{\sigma}$ are competing with each other to affect the robust solution. Consider the ratio $\overline{t}/\overline{\sigma}$. If the uncertainty in $\tilde{\mu}$ is large, then $\overline{t}$ will be large, and then it is more likely that the optimal robust solution will consist of a positive margin. On the other hand, if $\tilde{\sigma}$ is large, then it is more likely that the

optimal robust solution will be a pure intensity scaling with no positive margin. Our intuition here is similar to that from the nominal case, but now we are incorporating uncertainty in $\tilde{\mu}$ and the largest value of $\tilde{\sigma}$ into the analysis.

We close this section by demonstrating that the optimal robust solution from formulation (3.8) provides better dose minimization than a particular, intuitive heuristic procedure. In principle, one way to generate a robust solution would be to solve the nominal problem for every realization of $\tilde{\mu}$ and $\tilde{\sigma}$, and then take the "union" of the resulting solutions. That is, a robust solution generated from this method would be the smallest intensity map that contains the intensity maps corresponding to each solution of the nominal problem for realizations of $\tilde{\mu}$ and $\tilde{\sigma}$. However, this method produces inferior intensity maps in terms of the total dose delivered (although they do provide tumor coverage), as we show in the following examples.

Example 3.2. *Let $t = 3$, $\tilde{\mu} \in [\underline{\mu}, \overline{\mu}] = [-0.5, 0.5]$ and $\tilde{\sigma} = 1$. Note that $\bar{t} = 4 > t > u^*$, so in both the robust and nominal solutions, the optimal margin size is positive.*

Solving the robust problem (3.8) *with $\bar{t} = 4$ yields an optimal solution which we denote as $R1$. This solution has an optimal margin size of $m^*_{R1} = 0.835$ and an intensity scaling factor of* 1.253 *(from Table 3.1). Hence, the total dose delivered is*

$$\frac{\bar{t} + 2m^*_{R1}}{\Phi(m^*_{R1} + \bar{t}) - \Phi(-m^*_{R1})} = \frac{4 + 2 \cdot 0.835}{\Phi(0.835 + 4) - \Phi(-0.835)} = 1.253(4 + 2 \cdot 0.835) = 7.104.$$

We now create a heuristic robust solution, $R2$, from the nominal solution by taking the maximum of all intensity maps (one for each $\tilde{\mu}$). Since the intensity map corresponding to $R2$ is at least as high as the nominal intensity map for any realization of $\tilde{\mu}$, it is feasible for any realization of $\tilde{\mu}$.

The optimal solution, N, for the nominal problem consists of a margin of size $m^*_N = 0.539$ with an intensity scaling factor of 1.419. *Therefore, the robust intensity map, $R2$, is*

$$\delta_{R2}(y) = \begin{cases} 1.418, & y \in [-t/2 + \underline{\mu} - m^*_N, t/2 + \overline{\mu} + m^*_N] = [-2.539, 2.539], \\ 0, & \textit{otherwise}, \end{cases}$$

which leads to a total delivered dose of $2 \cdot 2.539 \cdot 1.418 = 7.204$.

The previous example shows only a small difference in the total dose delivered between the optimal robust solution and the heuristic robust solution. This is due in part to the fact that both intensity maps involve positive margin sizes (since $\bar{t} > t > u^*$). The next example provides a starker contrast between these two solutions since the uncertainty in $\tilde{\mu}$ forces the effective tumor size over the threshold u^* (i.e., $\bar{t} > u^* > t$). Thus, the heuristic robust intensity map (i.e., union of nominal intensity maps) is a union of purely scaled intensity maps, whereas the optimal robust intensity map involves a positive margin with a smaller scaling factor.

Example 3.3. *Consider the same setup as in the previous example, but with* $t = 2$. *With this value of* t *the nominal solution will be a pure intensity scaling, whereas the robust solution will involve a positive margin (since* $\bar{t} = 3 > u^*$*). Solving the robust problem with* $\bar{t} = 3$ *results in a robust solution,* $R1$*, with a margin of size* $m^*_{R1} = 0.539$*, and an intensity scaling factor of* 1.418*. Thus, the total dose delivered by the robust solution,* $R1$*, is*

$$1.418(3 + 2 \cdot 0.539) = 5.785.$$

The optimal solution, N*, for the nominal problem is* $m^*_N = 0$ *with an intensity scaling factor of* 2.095*. The heuristic robust intensity map,* $R2$*, generated from these nominal solutions is*

$$\delta_{R2}(y) = \begin{cases} 2.095, & y \in [-t/2 + \underline{\mu} - m^*_N, t/2 + \overline{\mu} + m^*_N] = [-1.5, 1.5], \\ 0, & \textit{otherwise}, \end{cases}$$

which leads to a total delivered dose of $2 \cdot 1.5 \cdot 2.095 = 6.286$*. This represents an approximately* 9% *increase in total dose delivered as compared to the robust solution* $R1$*.*

Example 3.3 shows that there can be relatively significant gains in terms of dose minimization by using the optimal robust solution from an appropriately formulated mathematical program (formulation (3.8)), as opposed to generating a robust solution

using a heuristic procedure involving only solutions of the nominal problem.

In the next two sections, we extend the analysis presented in this, and the previous section. Up until now, we have studied a one-dimensional geometry, and in the next two sections, we will consider a three-dimensional tumor subject to one-dimensional Gaussian motion.

3.3 Optimal 3D margins

In this section, we consider the case of a three-dimensional tumor subject to one-dimensional motion. Consider a spherical tumor of radius $t/2$ centered at $\boldsymbol{\mu}$, which we write as

$$\mathcal{T}(\boldsymbol{\mu}) := \{\mathbf{x} \in \mathbb{R}^3 \ : \ \|\mathbf{x} - \boldsymbol{\mu}\|_2 \leq t/2\}.$$

We assume the tumor is subject to one-dimensional Gaussian motion along the x_1 axis. As in Section 3.1, we can assume, without loss of generality, that the beam is centered on the tumor, and both are centered at the origin (i.e., $\boldsymbol{\mu} = \mathbf{0}$). We aim, in this section, to design intensity maps that ensure tumor coverage in the presence of motion, while minimizing the total dose delivered. Our goal is to understand when the optimal intensity map will have a positive margin, and to derive simple relationships between the problem parameters that characterize the different cases.

3.3.1 Summary and Highlights

In this section, we show that the optimal margin size will always be positive, regardless of the value of t/σ. Unlike the previous sections, there is no threshold that separates the cases when the optimal margin size is zero or not. Our intuition is built on the fact that a three-dimensional spherical tumor that moves in one dimension has a ring of points along its boundary that will not receive any dose as a result of the blurring effect of the motion. Thus, in order to ensure tumor coverage, a margin *must* be introduced to allow dose from parts of the static intensity map outside of the tumor to blur into this ring of points.

3.3.2 Analysis

First, we must precisely define a margin in this 3D situation. Since we are focused on 1D motion, we consider a margin to be an expansion of the irradiated (3D) volume along the direction of motion. A margin of size m will extend the volume of irradiation in both the positive and negative x_1 direction by m units. Hence, adding a margin of size m to a spherical volume of irradiation results in an overall volume of irradiation that consists of a cylinder of length $2m$ joining two hemispheres with radius $t/2$. See Figure 3-3 for an illustration of the static intensity map.

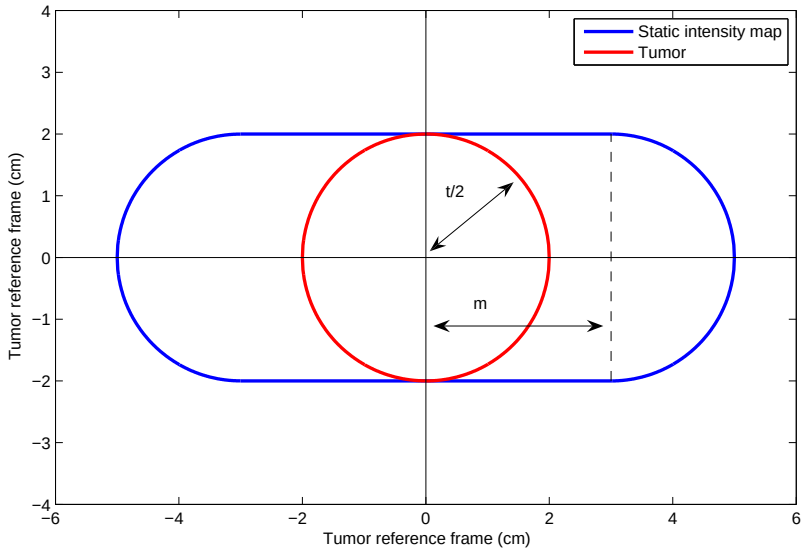


Figure 3-3: A two-dimensional slice of the three-dimensional static intensity map.

Mathematically, we write the static intensity map as

$$\delta_m(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in C(m, t/2) \cup RHS((m,0,0), t/2) \cup LHS((-m,0,0), t/2), \\ 0, & \text{otherwise}, \end{cases}$$

where

$$
\begin{aligned}
C(a,b) &= \{\mathbf{x} \in \mathbb{R}^3 \ : \ -a \leq x_1 \leq a, x_2^2 + x_3^2 \leq b^2\} \\
RHS((a_1,a_2,a_3),b) &= \{\mathbf{x} \in \mathbb{R}^3 \ : \ \|\mathbf{x} - (a_1,a_2,a_3)\|_2 \leq b, x_1 \geq a_1\} \\
LHS((a_1,a_2,a_3),b) &= \{\mathbf{x} \in \mathbb{R}^3 \ : \ \|\mathbf{x} - (a_1,a_2,a_3)\|_2 \leq b, x_1 \leq a_1\}.
\end{aligned}
$$

To account for motion, we convolve $\delta_m(\mathbf{y})$ with a Gaussian pdf, $p(x)$, with standard deviation σ, to arrive at the delivered dose

$$
\begin{aligned}
d_m(\mathbf{y};\mathbf{0},\sigma) &= \int_{y_1-m-\sqrt{(t/2)^2-(y_2^2+y_3^2)}}^{y_1+m+\sqrt{(t/2)^2-(y_2^2+y_3^2)}} p(x_1)dx_1 \\
&= \Phi\left(\frac{y_1+m+\sqrt{(t/2)^2-y_2^2-y_3^2}}{\sigma}\right) - \Phi\left(\frac{y_1-m-\sqrt{(t/2)^2-y_2^2-y_3^2}}{\sigma}\right).
\end{aligned}
$$

In order to ensure sufficient dose to the tumor, we need to scale $d_m(\mathbf{y};\mathbf{0},\sigma)$ by some positive factor, γ. The total dose delivered with a margin of size m before applying the scaling factor γ is equal to the volume of the corresponding static intensity map. Since the static intensity map is the union of two hemispheres and a cylinder, so the static dose delivered before scaling is $(4/3)\pi(t/2)^3 + 2m\pi(t/2)^2$. Thus, we pose our optimization problem as

$$
\begin{aligned}
\underset{m,\gamma}{\text{minimize}} \quad & \gamma((4/3)\pi(t/2)^3 + 2m\pi(t/2)^2) \\
\text{subject to} \quad & \gamma d_m(\mathbf{y};\mathbf{0},\sigma) \geq 1, \quad \forall\, \mathbf{y} \in \mathcal{T}(\mathbf{0}), \\
& m, \gamma \geq 0.
\end{aligned} \tag{3.9}
$$

Since our objective is to minimize the total dose delivered, we would like γ to be as small as possible. Let γ_m^* define the optimal (minimal) intensity scaling factor for a given margin size m. The following result establishes a closed-form expression for γ_m^*.

Lemma 3.3. *The quantity γ_m^* is equal to*

$$\gamma_m^* = \frac{1}{\Phi(\frac{m}{\sigma}) - \Phi(\frac{-m}{\sigma})}.$$

Proof: By definition,

$$\gamma_m^* = \max_{\mathbf{y} \in \mathcal{T}(\mathbf{0})} \frac{1}{d_m(\mathbf{y}; \mathbf{0}, \sigma)} = \frac{1}{\min_{\mathbf{y} \in \mathcal{T}(\mathbf{0})} d_m(\mathbf{y}; \mathbf{0}, \sigma)}.$$

For any fixed y_2 and y_3, $\delta_m(\mathbf{y})$ is a unimodal function of y_1. Specifically, $\delta_m(\mathbf{y})$ as a function of y_1 is an impulse with height 1 and length $2(m + \sqrt{(t/2)^2 - y_2^2 - y_3^2})$ (see Figure 3-3). Hence, by the strong unimodality property of the Gaussian pdf, $d_m(\mathbf{y}; \mathbf{0}, \sigma)$ is unimodal in y_1, for any given y_2 and y_3. This implies that for any y_2, y_3, the minimum of $d_m(\mathbf{y}; \mathbf{0}, \sigma)$ over $\{y_1 \ : \ \mathbf{y} \in \mathcal{T}(\boldsymbol{\mu})\}$ is on the boundary of $\mathcal{T}(\mathbf{0})$; in other words, satisfies $y_1^2 + y_2^2 + y_3^2 = (t/2)^2$. Thus, the minimum of $d_m(\mathbf{y}; \mathbf{0}, \sigma)$ over all $\mathbf{y} \in \mathcal{T}(\mathbf{0})$ lies on the boundary of $\mathcal{T}(\mathbf{0})$. Without loss of generality, let us consider $y_1 \geq 0$. On the boundary of $\mathcal{T}(\mathbf{0})$ and for $y_1 \geq 0$, we may write $d_m(\mathbf{y}; \mathbf{0}, \sigma)$ as

$$d_m(\mathbf{y}; \mathbf{0}, \sigma) = \Phi\left(\frac{2y_1 + m}{\sigma}\right) - \Phi\left(-\frac{m}{\sigma}\right),$$

which is clearly minimized at $y_1 = 0$. □

Lemma 3.3 shows that the set of points $\{\mathbf{y} \in \mathbb{R}^3 \ : \ y_1 = 0, y_2^2 + y_3^2 = (t/2)^2\}$ is the subset of the tumor that receives the least dose. This makes intuitive sense since the Gaussian pdf is spreading the static dose in the x_1 direction, and the dose distribution is "thinnest" along this direction for this set of points. In fact, if no margin is applied, the ring of points defined by $\{\mathbf{y} \in \mathbb{R}^3 \ : \ y_1 = 0, y_2^2 + y_3^2 = (t/2)^2\}$ will get no dose at all, since this set of points will have no neighbors in the x_1 direction that receive any intensity. Using this insight, we can deduce that all optimal intensity maps will involve a positive margin, which is the main result of this section.

Taking into account the closed form expression for γ_m^*, we pose the problem of finding an optimal intensity map that ensures tumor coverage while minimizing total

dose delivered as

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & f(m) \\ \text{subject to} \quad & m \geq 0, \end{aligned} \tag{3.10}$$

where

$$f(m) := \frac{(4/3)\pi(t/2)^3 + 2m\pi(t/2)^2}{\Phi(\frac{m}{\sigma}) - \Phi(\frac{-m}{\sigma})} = 2\pi(t/2)^2 \frac{m + t/3}{\Phi(\frac{m}{\sigma}) - \Phi(\frac{-m}{\sigma})}. \tag{3.11}$$

Our intuition about the optimal margin size being positive is reaffirmed since $f(m)$ is undefined at $m = 0$, which means that we would need an infinitely high (not realizable) intensity scaling to achieve tumor coverage. In order to solve problem (3.10), we first establish that $f(m)$ is a strictly convex function of m.

Proposition 3.4. *The function $f(m)$ is strictly convex for $m > 0$.*

Proof: Without loss of generality, we will prove the result for $\sigma = 1$. Let $g(m) := \Phi(m) - \Phi(-m)$. Note that $g(m)$ is positive for $m > 0$. Furthermore, since

$$\begin{aligned} g'(m) &= \phi(m) + \phi(-m) = 2\phi(m), \\ g''(m) &= -2m\phi(m), \end{aligned}$$

we see that $g(m)$ is strictly concave and strictly increasing for $m > 0$. To show that $f(m)$ is strictly convex, it suffices to show that $1/g(m)$ is strictly convex and $m/g(m)$ is convex, since $f(m)$ can be written as a linear combination of these two terms. Let $h(m) := 1/g(m)$. Then

$$g^4(m)h''(m) = -g''(m)g^2(m) + 2g'(m)g(m)g'(m) > 0$$

for $m > 0$. Hence, $1/g(m)$ is strictly convex.

Let $H(m) := m/g(m)$. Then

$$\begin{aligned} g^4(m)H''(m) &= -mg''(m)g^2(m) - 2(g(m) - mg'(m))g(m)g'(m) \\ &= m^2g'(m)g^2(m) - 2(g(m) - mg'(m))g(m)g'(m) \\ &= g(m)g'(m)(m^2g(m) - 2g(m) + 2mg'(m)), \end{aligned}$$

where the second equality comes from the relation $g''(m) = -mg'(m)$. It suffices to show that $H_1(m) := m^2g(m) - 2g(m) + 2mg'(m) \geq 0$ for $m \geq 0$, or since $H_1(0) = 0$, that $H_1(m)$ is an increasing function of m. The derivative of $H_1(m)$ is

$$H_1'(m) = 2mg(m) + (m^2 - 2)2\phi(m) + 4\phi(m) - 4m^2\phi(m) = 2m(g(m) - m\phi(m)).$$

Repeating the previous argument, let $H_2(m) := g(m) - m\phi(m)$, and since $H_2(0) = 0$, it suffices to show that $H_2'(m) \geq 0$. Taking one more derivative,

$$H_2'(m) = 2\phi(m) - \phi(m) + m^2\phi(m) = (1 + m^2)\phi(m),$$

which is clearly positive for all $m \geq 0$. Hence $H''(m) \geq 0$ for $m \geq 0$, and the result follows. □

Since $f(m)$ is strictly convex for $m > 0$, it must have a unique minimum for $m > 0$.

Theorem 3.4. *The function $f(m)$ has a unique minimum at $m^* > 0$, where m^* is the unique solution to*

$$\Phi\left(\frac{m}{\sigma}\right) - \Phi\left(\frac{-m}{\sigma}\right) - \frac{2(m + t/3)}{\sigma}\phi\left(\frac{m}{\sigma}\right) = 0. \tag{3.12}$$

Proof: Note that

$$\lim_{m\to 0} f(m) = \lim_{m\to\infty} f(m) = \infty,$$

and $f(m) < \infty$ for $m > 0$. Then, since $f(m)$ is differentiable and strictly convex for

$m > 0$, $f(m)$ is minimized at the unique solution to the equation $f'(m) = 0$, which is equivalent to (3.12). □

Theorem 3.4 shows that the optimal margin size will always be positive for the 3D problem presented in this section. This is in contrast to the 1D case, in which it was possible for the optimal margin size to be zero. Thus, for the 3D nominal problem, there is no threshold (cf. Proposition 3.3 and Theorem 3.1) analogous to the one obtained for the 1D nominal problem.

Table 3.2 shows the relationship between tumor size, margin size, and scaling factor in the 3D nominal problem. Figure 3-4 plots m^*/σ versus t/σ.

Table 3.2: The optimal margin size (m^*) for particular values of the tumor size (t), relative to the size of the standard deviation of the Gaussian pdf ($\sigma = 1$ in this case), in the 3D nominal problem. The corresponding intensity scaling factors ($\gamma^*_{m^*}$) are also shown.

t	1	2	3	4	5	6	7	8	9	10
m^*	0.941	1.148	1.282	1.383	1.463	1.530	1.587	1.638	1.682	1.723
$\gamma^*_{m^*}$	1.531	1.335	1.250	1.200	1.168	1.144	1.127	1.113	1.102	1.093

In the next section, we analyze the effect of uncertainty in the parameters of the Gaussian motion distribution.

3.4 Optimal 3D margins with robustness

In this section, we consider uncertainty in $\boldsymbol{\mu}$ and σ. These uncertain quantities will be denoted $\tilde{\boldsymbol{\mu}}$ and $\tilde{\sigma}$, respectively. Our general uncertainty model assumes that $\tilde{\boldsymbol{\mu}}$ lies in a hyperrectangle defined by $[\overline{\boldsymbol{\mu}}, \underline{\boldsymbol{\mu}}]$ ($\mu_i \in [\underline{\mu}_i, \overline{\mu}_i], i = 1, 2, 3$) and $\tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}]$. Since we focus on 1D motion, we will set $\overline{\mu}_i = \underline{\mu}_i = 0$ for $i = 2, 3$. Following a similar argument as in Section 3.2, we center the beam of radiation at $(x_1, x_2, x_3) = ((\overline{\mu}_1 + \underline{\mu}_1)/2, 0, 0)$, and re-center the coordinate axes on this location. For notational convenience, we define $\hat{\mu}_1 := \overline{\mu}_1 - \underline{\mu}_1$. We denote the effective tumor centered at the origin as

$$\overline{\mathcal{T}}(\mathbf{0}) = \{\mathbf{x} \ : \ \mathbf{x} \in C(\hat{\mu}_1/2, t/2) \cup RHS((\hat{\mu}_1/2, 0, 0), t/2) \cup LHS((-\hat{\mu}_1/2, 0, 0), t/2)\}.$$

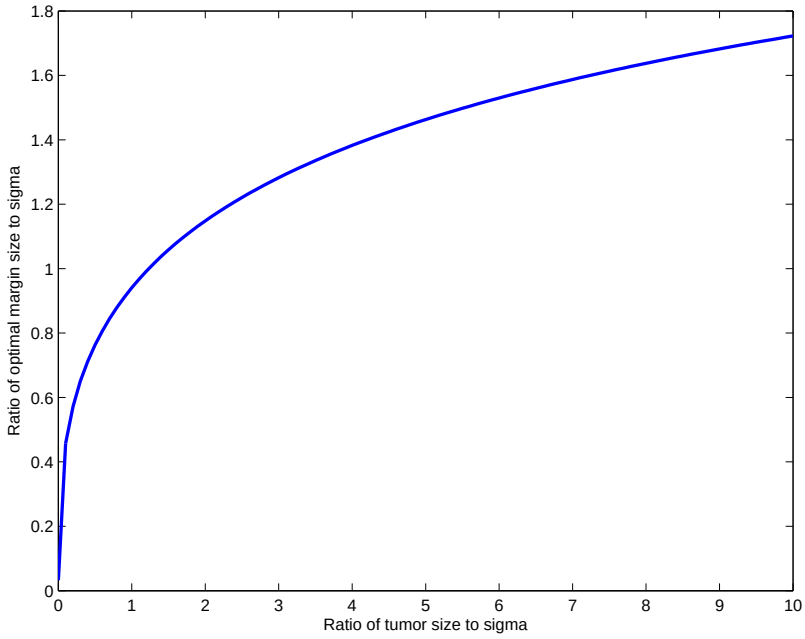


Figure 3-4: The ratio of the optimal margin size to σ as a function of the ratio of tumor size to σ for a 3D tumor.

3.4.1 Summary and Highlights

In this section, we show that the robust problem has a deterministic robust counterpart with a similar structure to both the one-dimensional and three-dimensional nominal margin problems. Furthermore, we show that a series of thresholds determines whether or not the optimal margin size will be positive. If $t/\hat{\mu}_1 > 0.139$, then the optimal margin size is always positive, and is equal to the unique solution of a nonlinear equation. If $t/\hat{\mu}_1 = 0.139$, then the optimal margin size is zero if $\hat{\mu}_1/\overline{\sigma} = t/(0.139\overline{\sigma}) = 1.500$, and positive if $\hat{\mu}_1/\overline{\sigma} \neq 1.500$. Finally, if $t/\hat{\mu}_1 < 0.139$, the optimal margin size is positive or zero, depending on the relationship between $\hat{\mu}_1/\overline{\sigma}$ and two constants determined by the solution to a system of two equations (which depend on the value of $t/\hat{\mu}_1$) in two unknowns.

Intuitively, the results in this section suggest that the optimal robust margin size will always be positive, unless the size of the tumor is much smaller than the uncertainty in μ, which in practice means that the tumor itself is physically small.

3.4.2 Analysis

We pose our robust optimization problem as

$$\begin{aligned} \underset{m,\gamma}{\text{minimize}} \quad & \gamma((4/3)\pi(t/2)^3 + (2m + \hat{\mu}_1)\pi(t/2)^2) \\ \text{subject to} \quad & \gamma d_m(\mathbf{y}; \mathbf{0}, \tilde{\sigma}) \geq 1, \quad \forall\, \mathbf{y} \in \overline{\mathcal{T}}(\mathbf{0}), \forall\, \tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}], \\ & m, \gamma \geq 0. \end{aligned} \tag{3.13}$$

Under the current setup, uncertainty in μ_1 essentially adds to the size of the tumor in the x_1 direction. The static intensity map with a margin of size m looks like the static intensity map shown in Figure 3-3, except with $m + \hat{\mu}_1/2$ in place of m. Our first result derives a closed form expression for γ_m^*, the optimal scaling factor for any given m, and shows that the robust counterpart of formulation (3.13) has a similar (but not exactly the same) structure as the nominal formulation (3.9).

Theorem 3.5. *The robust counterpart of formulation* (3.13) *is*

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & \frac{A + Bm}{\Phi(\frac{m+\hat{\mu}_1}{\overline{\sigma}}) - \Phi(\frac{-m}{\overline{\sigma}})}, \\ \text{subject to} \quad & m \geq 0, \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} A &= \pi(t/2)^2\hat{\mu}_1 + (4/3)\pi(t/2)^3, \\ B &= 2\pi(t/2)^2. \end{aligned}$$

Proof: The proof of this proposition essentially follows the same line of reasoning as the proof of Theorem 3.2 (1D robust margin problem). The main difference is the calculation of the static dose distribution. Since we are focused on the effective

tumor, the static intensity map delivered is,

$$\delta_m(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in (C(m+\hat{\mu}_1/2, t/2) \cup RHS((m+\hat{\mu}_1/2, 0, 0), t/2) \\ & \qquad \cup LHS((-m-\hat{\mu}_1/2, 0, 0), t/2)), \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the delivered dose can be written as

$$d_m(\mathbf{y}; \mathbf{0}, \tilde{\sigma}) = \Phi\left(\frac{y_1 + m + \hat{\mu}_1/2 + \sqrt{(t/2)^2 - y_2^2 - y_3^2}}{\tilde{\sigma}}\right) - \Phi\left(\frac{y_1 - m - \hat{\mu}_1/2 - \sqrt{(t/2)^2 - y_2^2 - y_3^2}}{\tilde{\sigma}}\right).$$

We derive the minimal scale factor, γ_m^*, as in Lemma 3.3 (focus on the edge of the effective tumor, $y_1 = \hat{\mu}_1/2$, $y_2^2 + y_3^2 = (t/2)^2$; replace $\tilde{\sigma}$ with $\overline{\sigma}$) to arrive at

$$\gamma_m^* = \frac{1}{\Phi(\frac{m+\hat{\mu}_1}{\overline{\sigma}}) - \Phi(\frac{-m}{\overline{\sigma}})}.$$

Finally, the volume of the static intensity map is $4/3\pi(t/2)^3 + \pi(t/2)^2(2m + \hat{\mu}_1)$, and the result follows. □

Let us define $f(m)$ to be the objective function in (3.14) as a function of m:

$$f(m) := \frac{A + Bm}{\Phi(\frac{m+\hat{\mu}_1}{\overline{\sigma}}) - \Phi(\frac{-m}{\overline{\sigma}})}. \tag{3.15}$$

Note that the structure of $f(m)$ is very similar to the 3D nominal objective function (3.11) in that it is a linear function of m divided by the difference of two error functions. However, the one major difference that ends up affecting the structural results is the presence of the $\hat{\mu}_1$ in the first error function in the denominator of (3.15). In fact, $f(m)$ shares a more similar structure to the objective function in the 1D nominal problem (3.1) (where $\hat{\mu}_1$ takes the place of the length of the tumor in the denominator). Therefore, unlike the 3D nominal problem, we now expect to derive a threshold that

separates the cases where a positive margin is optimal or not. The reason is that the effective tumor, $\overline{\mathcal{T}}(0)$, is now elongated in the x_1 direction, as compared to the original tumor $\mathcal{T}(0)$. Thus, the points in the tumor that were without neighbors in the x_1 direction in the nominal problem (the set of points $\{\mathbf{y} \in \mathbb{R}^3 : y_1 = 0, y_2^2 + y_3^2 = (t/2)^2\}$) are now surrounded by other parts of the effective tumor along that direction. This allows a nonzero "spreading" of the dose along this direction (due to the Gaussian distribution) that was absent in the nominal case.

Since $f(m)$ has a similar structure to the objective function from the 1D case, it is straightforward to show (proof omitted) that $f(m)$ has a unique minimum for $m \geq 0$ (cf. Proposition 3.2), and that the sign of $f'(0)$ dictates whether or not the optimal margin size is positive (cf. Proposition 3.3).

Proposition 3.5. *For any $t > 0$, $\hat{\mu}_1 > 0$ and $\overline{\sigma} > 0$, $f(m)$ has a unique local minimum, m^*, for $m \geq 0$.*

Proposition 3.6. *If $f'(0) \geq 0$, then $m^* = 0$. If $f'(0) < 0$, then m^* is the unique solution to the equation*

$$B\left(\Phi\left(\frac{m+\hat{\mu}_1}{\overline{\sigma}}\right) - \Phi\left(\frac{-m}{\overline{\sigma}}\right)\right) - \frac{A+Bm}{\overline{\sigma}}\left(\phi\left(\frac{m+\hat{\mu}_1}{\overline{\sigma}}\right) + \phi\left(\frac{-m}{\overline{\sigma}}\right)\right) = 0, \quad (3.16)$$

and $m^* > 0$.

Next, we derive a result that is analogous to Proposition 3.3 and Theorem 3.1 in the 1D case. The derivative of $f(m)$ with respect to m is

$$f'(m) = \frac{B(\Phi(\frac{m+\hat{\mu}_1}{\overline{\sigma}}) - \Phi(\frac{-m}{\overline{\sigma}})) - \frac{A+Bm}{\overline{\sigma}}(\phi(\frac{m+\hat{\mu}_1}{\overline{\sigma}}) + \phi(\frac{-m}{\overline{\sigma}}))}{(\Phi(\frac{m+\hat{\mu}_1}{\overline{\sigma}}) - \Phi(\frac{-m}{\overline{\sigma}}))^2}$$

and the numerator of $f'(m)$ at $m = 0$ is

$$\begin{aligned} & B\left(\Phi\left(\frac{\hat{\mu}_1}{\overline{\sigma}}\right) - \Phi(0)\right) - \frac{A}{\overline{\sigma}}\left(\phi\left(\frac{\hat{\mu}_1}{\overline{\sigma}}\right) + \phi(0)\right) \\ = \; & \pi(t/2)^2\left(2\int_0^{\frac{\hat{\mu}_1}{\overline{\sigma}}} \phi(x)dx - \frac{\hat{\mu}_1 + 2t/3}{\overline{\sigma}}\left(\phi\left(\frac{\hat{\mu}_1}{\overline{\sigma}}\right) + \phi(0)\right)\right). \end{aligned} \quad (3.17)$$

Note the difference between this expression and the corresponding one from the 1D case (equation (3.4) with $m = 0$). While the $\hat{\mu}_1$ appears to be acting somewhat like the t in the 1D case, there is still a t present in equation (3.17), so we cannot directly simplify this expression to a function of one variable as before (cf. equation (3.6), where $u := t/\sigma$). However, to overcome this complication, we parameterize t as a multiple of $\hat{\mu}_1$ through the relationship $t = \alpha\hat{\mu}_1$, for some parameter α, and then define $u := \hat{\mu}_1/\overline{\sigma}$. With this definition of u and the parameterization of t in terms of $\hat{\mu}_1$, we can write the right hand side of equation (3.17) (without the leading constant) as

$$h_\alpha(u) := 2\int_0^u \phi(x)dx - u\left(1 + \frac{2}{3}\alpha\right)(\phi(u) + \phi(0)). \tag{3.18}$$

Proposition 3.7. *There exists a unique solution $(\alpha^*, \overline{u})$ to the system of equations*

$$h_\alpha(u) = 0, \quad h'_\alpha(u) = 0$$

in the two variables α and u, respectively. Furthermore, for $0 < \alpha < \alpha^*$, there exist two positive solutions, $u^*_{1,\alpha}$ and $u^*_{2,\alpha}$ (assume $u^*_{1,\alpha} < u^*_{2,\alpha}$) to the equation $h_\alpha(u) = 0$. For $u \in (u^*_{1,\alpha}, u^*_{2,\alpha})$, $h_\alpha(u) > 0$, and for $u \notin [u^*_{1,\alpha}, u^*_{2,\alpha}]$, $h_\alpha(u) < 0$.

Proof: First, we determine the conditions under which $h_\alpha(u)$ has positive roots, by exploring the structure of $h'_\alpha(u)$. Then, we will show the existence and uniqueness of α^* and $\overline{u}$.

The derivative of $h_\alpha(u)$ is

$$\begin{aligned} h'_\alpha(u) &= 2\phi(u) - \left(1 + \frac{2}{3}\alpha\right)(\phi(u) + \phi(0)) - u\left(1 + \frac{2}{3}\alpha\right)(-u\phi(u)) \\ &= 2\phi(u) + (u^2 - 1)\left(1 + \frac{2}{3}\alpha\right)\phi(u) - \left(1 + \frac{2}{3}\alpha\right)\phi(0), \end{aligned}$$

and

$$h'_\alpha(0) = -\frac{4}{3}\alpha\phi(0).$$

Thus, for any $\alpha > 0$, $h_\alpha(u)$ is strictly decreasing for u sufficiently close to 0. Also, note that $h_\alpha(0) = 0$ and $\lim_{u\to\infty} h_\alpha(u) = -\infty$.

Now, we would like to explore the positive roots of $h'_\alpha(u)$, since this will give us information about $h_\alpha(u)$.

$$\begin{aligned}
& h'_\alpha(u) = 0 \\
\Leftrightarrow \quad & 2\phi(u) + (u^2-1)\left(1+\frac{2}{3}\alpha\right)\phi(u) = \left(1+\frac{2}{3}\alpha\right)\phi(0) \\
\Leftrightarrow \quad & e^{-\frac{u^2}{2}}\left(u^2 - 1 + \frac{2}{1+\frac{2\alpha}{3}}\right) = 1 \\
\Leftrightarrow \quad & 2x - 1 + \frac{2}{1+\frac{2\alpha}{3}} = e^x,
\end{aligned}$$

for $x = u^2/2$. Hence, the roots of $h'_\alpha(u)$ correspond to the intersections between the exponential function e^x and the linear function $l_\alpha(x) = 2x - 1 + 6/(3+2\alpha)$. We now characterize these intersection points.

If $\alpha = 0$, then $l_\alpha(x) = 2x + 1$, which intersects with e^x at the origin and at one positive x. The functions e^x and $l_\alpha(x)$ are tangent at $x = \ln 2$, which corresponds to the case when $\alpha = \overline{\alpha} := 3/(3 - \ln 4) - 3/2 \approx 0.359$ in $l_\alpha(x)$. As α increases, $l_\alpha(x)$ decreases, thus, for $0 < \alpha < \overline{\alpha}$, e^x and $l_\alpha(x)$ have two positive intersection points, which implies that $h'_\alpha(u)$ has two positive roots. Let us denote these roots as $\overline{u}_{1,\alpha}$ and $\overline{u}_{2,\alpha}$, where $\overline{u}_{1,\alpha} < \overline{u}_{2,\alpha}$. Moreover, by comparing e^x and $l_\alpha(x)$ when $0 < \alpha < \overline{\alpha}$, we can see that $l_\alpha(x) < e^x$ for $x < \overline{u}_{1,\alpha}$ and for $x > \overline{u}_{2,\alpha}$, and $l_\alpha(x) > e^x$ for $\overline{u}_{1,\alpha} < x < \overline{u}_{2,\alpha}$. For $\alpha = \overline{\alpha}$, $h'_\alpha(u)$ has one positive root (where e^x and $l_\alpha(x)$ are tangent), and for $\alpha > \overline{\alpha}$, $h'_\alpha(u)$ has no positive roots.

We now use this information on $h'_\alpha(u)$, coupled with the knowledge that $h'_\alpha(0) < 0$ for $\alpha > 0$ and $h'_\alpha(u) \to -\infty$ as $u \to \infty$, to describe $h_\alpha(u)$. If $\alpha \geq \overline{\alpha}$ ($h_\alpha(u)$ has at most one positive stationary point), then $h_\alpha(u)$ must be negative for $u > 0$. If $0 < \alpha < \overline{\alpha}$, then $h_\alpha(u)$ has two stationary points at $\overline{u}_{1,\alpha}$ and $\overline{u}_{2,\alpha}$. Since $l_\alpha(x) < e^x$ for $x < \overline{u}_{1,\alpha}$ and $l_\alpha(x) > e^x$ for $\overline{u}_{1,\alpha} < x < \overline{u}_{2,\alpha}$, then $\overline{u}_{1,\alpha}$ is a strict local minimum of $h_\alpha(u)$. By a similar argument, $\overline{u}_{2,\alpha}$ is a strict local maximum of $h_\alpha(u)$. Figure 3-5 plots the function $h_\alpha(u)$ for different values of α.

By examining $h_\alpha(u)$ we see that it is decreasing in α. Thus, we search for the existence of an α^* in the interval $(0, \overline{\alpha})$ such that $h_{\alpha^*}(u)$ is tangent to the u-axis at

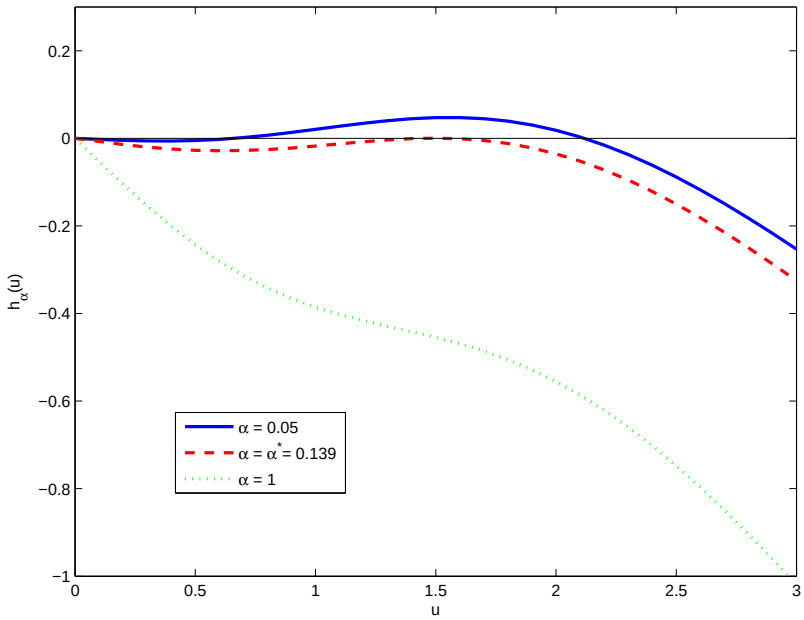


Figure 3-5: The function $h_\alpha(u)$ for different values of α.

some point u. At such an α^*, given the structure of $h_\alpha(u)$ that we have established already, $h_{\alpha^*}(u)$ will be tangent to the u-axis at exactly its local maximum. Since $h_\alpha(u)$ is decreasing in α, this implies that for $0 < \alpha < \alpha^*$, $h_\alpha(u)$ will have two positive roots, and for $\alpha > \alpha^*$, it will have none.

Solving for α in the equation $h'_\alpha(u) = 0$, we get

$$\begin{aligned}\alpha(u) &= \frac{-2\phi(u) - (u^2-1)\phi(u) + \phi(0)}{\frac{2}{3}(u^2-1)\phi(u) - \frac{2}{3}\phi(0)} \\ &= \frac{-(u^2+1)\phi(u) + \phi(0)}{\frac{2}{3}(u^2-1)\phi(u) - \frac{2}{3}\phi(0)},\end{aligned}$$

and substituting this in place of α in the expression for $h_\alpha(u)$ gives us

$$\begin{aligned}h_{\alpha(u)}(u) &= \Phi(u) - \Phi(-u) - u\left(1 + \frac{2}{3}\cdot\frac{-(u^2+1)\phi(u) + \phi(0)}{\frac{2}{3}(u^2-1)\phi(u) - \frac{2}{3}\phi(0)}\right)(\phi(u) + \phi(0)) \\ &= \Phi(u) - \Phi(-u) + \frac{2u\phi(u)(\phi(u) + \phi(0))}{(u^2-1)\phi(u) + \phi(0)}.\end{aligned}$$

Let us define $H(u) := h_{\alpha(u)}(u)$. We now show that $H(u)$ has exactly one positive root ($H(u)$ also has a root at $u = 0$). First, note that since $u\phi(u) \to 0$ as $u \to \infty$,

$H(u) \to 1$ as $u \to \infty$. Also, by taking derivatives (not shown), we determine that the lowest order derivative of $H(u)$ that is not equal to zero at $u = 0$ is the third derivative, and it can be shown to be

$$H^{(3)}(0) = -\sqrt{\frac{2}{\pi}} < 0.$$

Thus, by the intermediate value theorem (since $H(u)$ is continuous for $u \geq 0$), $H(u)$ has at least one positive root. In fact, we show that $H(u)$ has exactly one positive root. By construction, a root of $H(u)$ is a simultaneous root of $h_\alpha(u)$ and $h'_\alpha(u)$. By considering the structure of $h_\alpha(u)$ that we have already established, the only possible u where this can be achieved is at the unique strict local maximum of $h_\alpha(u)$. Hence, $H(u)$ has exactly one positive root, which we denote as $\overline{u}$. Using $\overline{u}$, we can define the threshold value of α, which we denoted as α^* above, to be

$$\alpha^* = \frac{3}{2}\left(\frac{\Phi(\overline{u}) - \Phi(-\overline{u})}{\overline{u}(\phi(\overline{u}) + \phi(0))} - 1\right).$$

Thus, for $\alpha = \alpha^*$, $h_\alpha(u) \leq 0$ and is tangent to the u-axis at $\overline{u}$, which is a strict local maximum of $h_\alpha(u)$. Since $h_\alpha(u)$ is decreasing in α, for $0 < \alpha < \alpha^*$, $h_\alpha(u)$ will be positive at its strict local maximum. As a result of the continuity of $h_\alpha(u)$, for $0 < \alpha < \alpha^*$, $h_\alpha(u)$ will have two positive roots, which we denote as $u^*_{1,\alpha}$ and $u^*_{2,\alpha}$ (assume $u^*_{1,\alpha} < u^*_{2,\alpha}$), and will be positive for $u \in (u^*_{1,\alpha}, u^*_{2,\alpha})$. □

Using this proposition, we now derive the thresholds that separate the cases when the optimal margin size will be positive or not.

Theorem 3.6. *Let $\alpha^*, u^*_{1,\alpha}, u^*_{2,\alpha}$, and $\overline{u}$ be as defined in Proposition 3.7. Then, if $0 < \alpha < \alpha^*$, the optimal margin size, m^*, is positive if $\hat{\mu}_1/\overline{\sigma} < u^*_{1,\alpha}$ or $\hat{\mu}_1/\overline{\sigma} > u^*_{2,\alpha}$, and zero if $u^*_{1,\alpha} \leq \hat{\mu}_1/\overline{\sigma} \leq u^*_{2,\alpha}$. If $\alpha = \alpha^*$, the optimal margin size is positive if $\hat{\mu}_1/\overline{\sigma} \neq \overline{u}$, and zero if $\hat{\mu}_1/\overline{\sigma} = \overline{u}$. If $\alpha > \alpha^*$, the optimal margin size is always positive. In the cases where the optimal margin size is positive, m^* is the unique solution to equation* (3.16).

Proof: The sign of $h_\alpha(u)$ corresponds to the sign of $f'(0)$. Hence, from Proposition 3.7, if $0 < \alpha < \alpha^*$, $f'(0) \geq 0$ for $u \in [u^*_{1,\alpha}, u^*_{2,\alpha}]$. This implies that the optimal margin size is zero. If $0 < \alpha < \alpha^*$ and $u \notin [u^*_{1,\alpha}, u^*_{2,\alpha}]$, then $f'(0) < 0$, and the optimal margin size is positive. If $\alpha = \alpha^*$ and $u \neq \overline{u}$, then $f'(0) < 0$, which implies that the optimal margin size is positive. Finally, if $\alpha > \alpha^*$, then $f'(0) < 0$, which implies the optimal margin size is positive. The optimal margin size is the unique positive root of $f'(m) = 0$, which is equivalent to (3.16). □

Numerically, we find that $\overline{u} = 1.500$ and $\alpha^* = 0.139$. Note that the structure of $h_\alpha(u)$ is very similar to the corresponding function (3.6) in the 1D case, with the only difference being the factor of $2\alpha/3$. This extra term changes the structure of $f'(0)$ significantly, resulting in an additional α-threshold, that is needed to classify the margin thresholds. If $0 < \alpha < \alpha^* = 0.139$, then $t < 0.139\hat{\mu}_1$, which means that the effective tumor is long and skinny (its length is at least ∼8 times its width). In this situation, there are some intermediate values of the ratio t/σ (those between $u^*_{1,\alpha}$ and $u^*_{2,\alpha}$) where the optimal margin size is zero, but otherwise, the optimal margin size is positive.

Table 3.3 summarizes the possible combinations between values of α and values of the optimal margin size, m^*.

Table 3.3: The relationship between α (the multiplicative factor in the relation $t = \alpha\hat{\mu}_1$) and m^* (the optimal margin size). A check (✓) indicates that the outcome is possible, while a cross (×) indicates that the outcome is not possible.

	$0 < \alpha \leq \alpha^*$	$\alpha > \alpha^*$
$m^* > 0$	✓	✓
$m^* = 0$	✓	×

Figure 3-6 illustrates the relationship between the optimal margin size and the tumor size for different values of α.

Next, we will move beyond a simple margin intensity map, and consider intensity modulation in one-dimension. We add edge-enhancements to our framework and derive results similar to those we have found for the 1D margin case.

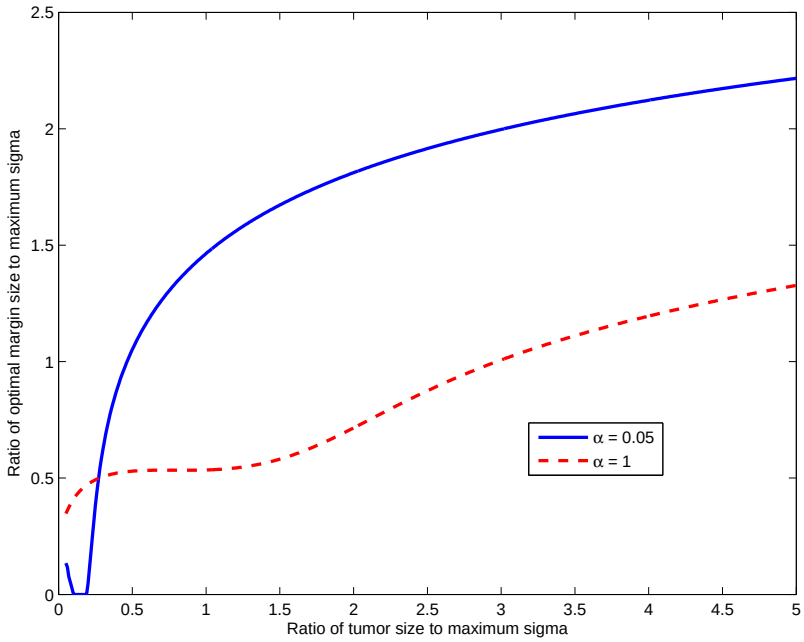


Figure 3-6: The ratio of the optimal margin size to σ as a function of the ratio of tumor size to σ for a 3D tumor, for different values of α.

3.5 Optimal edge-enhancements

In this section, we generalize the analysis from Section 3.1 by considering intensity maps that allow intensity modulation in the edge region of the tumor. These "edge-enhancements" give us the flexibility to scale the intensity delivered around the edge of the tumor without modifying the intensity near the middle of the tumor. Our analysis in this section is motivated by the structure of the robust solution that was observed in Section 2.6 (cf. Figure 2-9). The basic problem setup involves a tumor of length t that is subject to Gaussian motion. However, in addition to the margin variable m from the previous sections, we now introduce a variable l that describes the width of the part of the edge-enhancement that extends into the tumor, as well a variable h that describes the height of the edge-enhancements above the desired dose level. The part of the intensity map that is not edge-enhanced remains at the desired dose level - there is no scaling of the intensity in this region. Due to the symmetry of the Gaussian distribution, we only consider symmetric intensity maps. Figure 3-7 illustrates the relevant variables associated with an edge-enhanced intensity map.

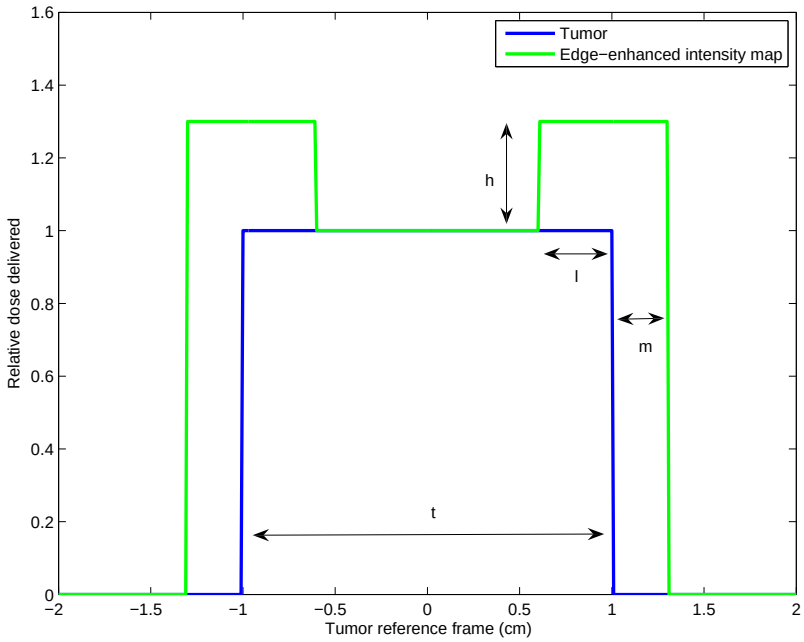


Figure 3-7: An edge-enhanced intensity map.

Note one minor difference between h and the scaling factor γ from the margin problem. The variable h represents the amount of intensity delivered *above* the required dose level, whereas the scaling factor γ is a multiplicative factor that scales the entire intensity map. For example, $h = 1$ means that the maximum height of the edge-enhanced intensity map is 2, whereas $\gamma = 1$ means that the maximum height of the margin intensity map is 1.

We assume that the beam is centered on the tumor, where the origin of our coordinate system is located. We let $\mathcal{T}(0) := [-t/2, t/2]$ denote the tumor centered at the origin. We also define $\mathcal{E}(\mu)$ to be the location of the edge-enhancements when the intensity map is centered at μ. Using the variables m and l, we may write $\mathcal{E}(\mu)$ as

$$\mathcal{E}(\mu) = [-t/2 - m + \mu, -t/2 + l + \mu] \cup [t/2 - l + \mu, t/2 + m + \mu].$$

We constrain l to be at most $t/2$, since the two edge-enhancements would overlap if $l > t/2$. Furthermore, the special case where $l = t/2$ is exactly the margin problem from the previous subsection (except for the slight notational difference between h

and γ mentioned earlier).

The goal of this section is to derive bounds on the ratio t/σ that characterize the structure of the resulting intensity maps.

3.5.1 Summary and Highlights

In this section, we show that if $0 \leq t/\sigma \leq 2.109$, then a locally-optimal solution involves a scaling of the entire static intensity map, with no separate intensity modulation over the edge of the tumor and no positive margin. If $t/\sigma > 2.109$, we show that an edge-enhanced solution (where the optimal l is strictly less than $t/2$) will always outperform a non-edge-enhanced solution. In particular, if $2.109 < t/\sigma \leq 2.281$, then there exists a locally-optimal solution that exhibits edge-enhancements and no positive margin. Finally, we conjecture that the locally-optimal solutions presented in this chapter are globally-optimal solutions, and that there will only be one threshold (at 2.109) that separates the cases when the globally-optimal solution will exhibit intensity modulation or not.

3.5.2 Analysis

We define the static intensity map, $\delta_{m,l,h}(x)$, to be

$$\delta_{m,l,h}(x) = \begin{cases} 1, & x \in \mathcal{T}(0) \backslash \mathcal{E}(0), \\ 1+h, & x \in \mathcal{E}(0), \\ 0, & \text{otherwise.} \end{cases}$$

Convolving $\delta_{m,l,h}(x)$ with a Gaussian distribution with mean μ and standard de-

viation σ, we can write the delivered dose $d_{m,l,h}(y;\mu,\sigma)$, as a function of y, as

$$\begin{aligned}
d_{m,l,h}(y;\mu,\sigma) &= \int_{-\infty}^{\infty} \delta_{m,h,l}(y-x)p(x)dx \\
&= \int_{y-t/2+l}^{y+t/2-l} p(x)dx + \int_{y-t/2-m}^{y-t/2+l}(1+h)p(x)dx + \int_{y+t/2-l}^{y+t/2+m}(1+h)p(x)dx \\
&= \Phi\left(\frac{y+t/2-l-\mu}{\sigma}\right) - \Phi\left(\frac{y-t/2+l-\mu}{\sigma}\right) \\
&\quad + (1+h)\left[\Phi\left(\frac{y-t/2+l-\mu}{\sigma}\right) - \Phi\left(\frac{y-t/2-m-\mu}{\sigma}\right)\right] \\
&\quad + (1+h)\left[\Phi\left(\frac{y+t/2+m-\mu}{\sigma}\right) - \Phi\left(\frac{y+t/2-l-\mu}{\sigma}\right)\right] \\
&= (1+h)\Phi\left(\frac{y+t/2+m-\mu}{\sigma}\right) - h\Phi\left(\frac{y+t/2-l-\mu}{\sigma}\right) \\
&\quad + h\Phi\left(\frac{y-t/2+l-\mu}{\sigma}\right) - (1+h)\Phi\left(\frac{y-t/2-m-\mu}{\sigma}\right).
\end{aligned}$$

The total dose delivered is simply the area under the edge-enhanced intensity map, which is

$$\begin{aligned}
f(m,l,h) &:= 2(m+l)(1+h) + t - 2l \\
&= t + 2m + 2(m+l)h.
\end{aligned} \tag{3.19}$$

Hence, we can pose our problem as the following mathematical program (as in earlier sections, we set $\mu = 0$, without loss of generality)

$$\begin{aligned}
\underset{m,l,h}{\text{minimize}} \quad & 2(m+l)(1+h) + t - 2l \\
\text{subject to} \quad & d_{m,l,h}(y;0,\sigma) \geq 1, \quad \forall\, y \in [-t/2, t/2], \\
& l \leq t/2, \\
& m, l, h \geq 0.
\end{aligned} \tag{3.20}$$

As in the margin problem presented in Section 3.1, we will simplify formulation (3.20) by showing that the set of constraints

$$d_{m,l,h}(y;0,\sigma) \geq 1, \quad \forall\, y \in [-t/2, t/2]$$

is equivalent to a single constraint involving a particular y on the boundary of $\mathcal{T}(0)$. This shows that for any edge-enhanced solution, the only part of the tumor we need to worry about underdosing is the edge of the tumor.

Due to the inherent symmetry in our setup, we focus on just one side of the tumor ($y \geq 0$). Without loss of generality, and for notational simplicity, we assume $\sigma = 1$. We allow a slight abuse of notation by writing $d_{m,l,h}(y)$ in place of $d_{m,l,h}(y; 0, 1)$. First, we establish properties about the shape of the dose distribution $d_{m,l,h}(y)$.

Lemma 3.4. *If $0 \leq a \leq b$, then*

$$a \sinh bx \cosh ax - b \sinh ax \cosh bx \leq 0,$$

for all $x \geq 0$, with equality at $x = 0$.

Proof: Let $g(x) := a \sinh bx \cosh ax - b \sinh ax \cosh bx$. Since $g(0) = 0$, it suffices to show that $g(x)$ is a nonincreasing function of x for $x \geq 0$. The derivative of $g(x)$ is

$$\begin{aligned} g'(x) &= ab \cosh bx \cosh ax + a^2 \sinh ax \sinh bx - ab \cosh ax \cosh bx - b^2 \sinh ax \sinh bx \\ &= -(b^2 - a^2) \sinh ax \sinh bx, \end{aligned}$$

which is clearly nonpositive for $x \geq 0$. □

Lemma 3.5. *If $0 \leq a \leq b$,*

$$b \sinh ax \leq a \sinh bx,$$

for all $x \in \mathbb{R}$, with equality holding at $x = 0$. Furthermore, for $b > 0$, $\sinh ax / \sinh bx$ is a decreasing function of x for $x > 0$ and an increasing function of x for $x < 0$.

Proof: The case where $b = 0$ is trivial, so assume $b > 0$. Let $g(x) := \sinh ax / \sinh bx$. Using L'Hôpital's rule, $\lim_{x \to 0} g(x) = a/b$. Then, since $g(x)$ is an even function of x, it is sufficient to show that $g'(x) \leq 0$ for all $x \geq 0$. The derivative of $g(x)$ is

$$g'(x) = \frac{a \sinh bx \cosh ax - b \sinh ax \cosh bx}{\sinh^2 bx},$$

which is nonpositive, from Lemma 3.4. □

Using the lemmas just established, we can show that $d_{m,l,h}(y)$ is a quasiconcave function of y for $y \geq 0$.

Theorem 3.7. *For $t > 0$, $m \geq 0$ and $l \in [0, t/2]$, $d_{m,l,h}(y)$ is a quasiconcave function of y for $y \geq 0$.*

Proof: Recall the form of $d_{m,l,h}(y)$,

$$\begin{aligned} d_{m,l,h}(y) = (1+h)\left(\Phi(y+t/2+m) - \Phi(y-t/2-m)\right) \\ - h\left(\Phi(y+t/2-l) - \Phi(y-t/2+l)\right). \end{aligned}$$

Let us consider the following two mutually exclusive and collectively exhaustive scenarios. Either

$$e^{-\frac{1}{2}((t/2+m)^2-(t/2-l)^2)} \geq \frac{h}{1+h} \cdot \frac{t/2-l}{t/2+m} \tag{3.21}$$

or

$$e^{-\frac{1}{2}((t/2+m)^2-(t/2-l)^2)} < \frac{h}{1+h} \cdot \frac{t/2-l}{t/2+m}. \tag{3.22}$$

We will show that $d_{m,l,h}(y)$ is quasiconcave for $0 \leq y \leq t/2$ in both cases. In particular, $d_{m,l,h}(y)$ is decreasing when the first inequality holds, and it is increasing and then possibly decreasing when the second inequality holds. For notational convenience, let us define $a := t/2 - l$ and $b := t/2 + m$. Clearly, $b \geq a \geq 0$ and $b > 0$. The first derivative of $d_{m,l,h}(y)$ with respect to y is

$$\begin{aligned} d'_{m,l,h}(y) = (1+h)\phi(y+t/2+m) - h\phi(y+t/2-l) \\ + h\phi(y-t/2+l) - (1+h)\phi(y-t/2-m). \end{aligned}$$

Suppose inequality (3.21) holds. Then

$$\begin{aligned} e^{-\frac{1}{2}(b^2-a^2)} &\geq \frac{h}{1+h} \cdot \frac{a}{b} \\ &\geq \frac{h}{1+h} \cdot \frac{\sinh ay}{\sinh by} \\ &= \frac{h}{1+h} \cdot \frac{e^{-ay} - e^{ay}}{e^{-by} - e^{by}} \\ &= \frac{h}{1+h} \cdot \frac{e^{-ay} - e^{ay}}{e^{-by} - e^{by}} \cdot \frac{e^{-\frac{1}{2}(a^2+y^2+b^2)}}{e^{-\frac{1}{2}(a^2+y^2+b^2)}} \\ &= \frac{h}{1+h} \cdot \frac{e^{-\frac{1}{2}(a+y)^2} - e^{-\frac{1}{2}(a-y)^2}}{e^{-\frac{1}{2}(b+y)^2} - e^{-\frac{1}{2}(b-y)^2}} \cdot \frac{e^{-\frac{1}{2}b^2}}{e^{-\frac{1}{2}a^2}} \\ &= \frac{h}{1+h} \cdot \frac{\phi(y+a) - \phi(y-a)}{\phi(y+b) - \phi(y-b)} \cdot e^{-\frac{1}{2}(b^2-a^2)}, \end{aligned}$$

where the second inequality comes from Lemma 3.5. Since $\phi(y+b) - \phi(y-b)$ is negative for $y \geq 0$,

$$d'_{m,l,h}(y) = (1+h)(\phi(y+b) - \phi(y-b)) - h(\phi(y+a) - \phi(y-a)) \leq 0.$$

In other words, $d_{m,l,h}(y)$ is decreasing for $y \geq 0$.

For the second case, suppose inequality (3.22) holds:

$$e^{-\frac{1}{2}(b^2-a^2)} < \frac{h}{1+h} \cdot \frac{a}{b}.$$

Recall from Lemma 3.5 that $\sinh ay / \sinh by$ is decreasing for $y > 0$ and equals a/b at $y = 0$. Therefore, the function $h \sinh ay / ((1+h) \sinh by)$ must cross $e^{-\frac{1}{2}(b^2-a^2)}$ at most once, which (from the above chain of equalities) implies that $d_{m,l,h}(y)$ is increasing and then possibly decreasing for $y \geq 0$. □

With knowledge of the shape of the dose distribution corresponding to *any* edge-enhanced intensity map, we may now focus on the parts of the tumor that are most susceptible to being underdosed.

Corollary 3.3. *For $0 \leq y \leq t/2$, $d_{m,l,h}(y)$ is minimized at $y = 0$ or $y = t/2$.*

Proof: Since $d_{m,l,h}(y)$ is quasiconcave for $y \geq 0$, the minimum of $d_{m,l,h}(y)$ on $[0, t/2]$ occurs at an extreme point of the interval $[0, t/2]$. Hence, $d_{m,l,h}(y)$ is minimized at 0 or $t/2$. □

Corollary 3.3 implies that it is sufficient to ensure tumor coverage at only three points: $y = -t/2, y =, 0$ and $y = t/2$. If these three points receive enough dose, then the entire tumor will receive enough dose. Of course, due to symmetry, we only need to consider one of $t/2$ and $-t/2$.

Substituting $y = t/2$ and $y = 0$ into the expression for $d_{m,l,h}(y)$ gives

$$\begin{aligned} d_{m,l,h}(t/2) = (1+h)\Phi(t+m) - h\Phi(t-l) \\ + h\Phi(l) - (1+h)\Phi(-m) \end{aligned}$$

and

$$\begin{aligned} d_{m,l,h}(0) = (1+h)\Phi(t/2+m) - h\Phi(t/2-l) \\ + h\Phi(-t/2+l) - (1+h)\Phi(-t/2-m), \end{aligned}$$

respectively. Hence, the mathematical program (3.20) is equivalent to

$$\begin{aligned} \underset{m,l,h}{\text{minimize}} \quad & t + 2m + 2(m+l)h \\ \text{subject to} \quad & d_{m,l,h}(0) \geq 1, \\ & d_{m,l,h}(t/2) \geq 1, \\ & l \leq t/2, \\ & m, l, h \geq 0, \end{aligned} \tag{3.23}$$

or more explicitly

$$\begin{aligned}
\underset{m,l,h}{\text{minimize}} \quad & t + 2m + 2(m+l)h \\
\text{subject to} \quad & (1+h)(\Phi(t/2+m) - \Phi(-t/2-m)) - h(\Phi(t/2-l) - \Phi(-t/2+l)) \geq 1, \\
& (1+h)(\Phi(t+m) - \Phi(-m)) - h(\Phi(t-l) - \Phi(l)) \geq 1, \\
& l \leq t/2, \\
& m, l, h \geq 0.
\end{aligned} \tag{3.24}$$

It turns out that formulation (3.24) can be simplified further. Indeed, we will show that we may drop the constraint $d_{m,l,h}(0) \geq 1$ without affecting the feasible region. In effect, this means that we only need to be concerned with underdosing the tumor at its edge. As long as the edge of the tumor receives enough dose, the middle of the tumor is guaranteed to receive enough as well for *any* edge-enhanced intensity map. Conversely, if the middle of the tumor is underdosed, then the edge of the tumor will also be underdosed.

We first make some definitions for notational convenience.

Definition 3.1. *For any $t > 0, m \geq 0$, and $0 \leq l \leq t/2$,*

1. $a_m := \Phi(m + t/2) - \Phi(-m - t/2)$.
2. $b_l := \Phi(t/2 - l) - \Phi(-t/2 + l)$.
3. $c_m := \Phi(m + t) - \Phi(-m)$.
4. $d_l := \Phi(t - l) - \Phi(l)$.

The following relationships between the functions a_m, b_l, c_m, d_l are straightforward to show, so the proofs are omitted.

Lemma 3.6. *For any $t > 0, m \geq 0$, and $0 \leq l \leq t/2$,*

$$1 \geq a_m \geq b_l \geq d_l,$$

and

$$1 \geq a_m \geq c_m \geq d_l.$$

Using the functions from Definition 3.1, we now show that we may drop the constraint $d_{m,l,h}(0) \geq 1$ in formulation (3.23) or (3.24).

Theorem 3.8. *For any $t > 0, m \geq 0$, and $0 \leq l \leq t/2$, if $d_{m,l,h}(t/2) \geq 1$, then $d_{m,l,h}(0) \geq 1$.*

Proof: Let us first consider the case where $m = l = 0$. In this case,

$$d_{0,0,h}(t/2) = \Phi(t) - \Phi(0),$$
$$d_{0,0,h}(0) = \Phi(t/2) - \Phi(-t/2).$$

Thus, to prove the result, it suffices to show that $\Phi(t/2) - \Phi(-t/2) \geq \Phi(t) - \Phi(0)$, which is straightforward because

$$\begin{aligned}\frac{\Phi(t/2) - \Phi(-t/2)}{2} &= \frac{2\Phi(t/2) - 1}{2} \\ &= \Phi(t/2) - 1/2 \\ &\geq \frac{\Phi(t) + \Phi(0)}{2} - 1/2 \\ &= \frac{\Phi(t) - \Phi(0)}{2},\end{aligned}$$

where the first equality comes from the symmetry of the Gaussian pdf and the inequality comes from the concavity of $\Phi(t)$ for $t \geq 0$.

Now, let us consider the case when not both $m = 0$ and $l = 0$. If $d_{m,l,h}(t/2) \geq 1$, then

$$h \geq \frac{1 - \Phi(m+t) + \Phi(-m)}{\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l)}$$

(note that the denominator is strictly positive when either $m > 0$ or $l > 0$), which

implies

$$d_{m,l,h}(0) \geq \Phi(m+t/2) - \Phi(-m-t/2) + \frac{1-\Phi(m+t)+\Phi(-m)}{\Phi(m+t)-\Phi(-m)-\Phi(t-l)+\Phi(l)} \cdot \left(\Phi(m+t/2) - \Phi(-m-t/2) - \Phi(t/2-l) + \Phi(-t/2+l)\right).$$

Hence, to show $d_{m,l,h}(0) \geq 1$, it is sufficient to show (recall Definition 3.1)

$$F(m,l) := a_m + \frac{1-c_m}{c_m - d_l}(a_m - b_l) \geq 1,$$

for all $m \geq 0$ and $0 \leq l \leq t/2$. To show this inequality, first note that for $t > 0$ and $0 \leq l \leq t/2$,

$$\lim_{m\to\infty} F(m,l) = 1, \tag{3.25}$$

since $a_m, c_m \to 1$ as $m \to \infty$. Thus, it suffices to establish that for any $t > 0$ and $0 \leq l \leq t/2$, $F(m,l)$ is a nonincreasing function of m for $m \geq 0$. To see why this is sufficient, suppose that $F(m,l)$ is nonincreasing in m and that $F(\overline{m},l) < 1$ for some $\overline{m} > 0$. Then, as a consequence of the limit (3.25), there exists an M such that for $m \geq M$, $F(m,l) > F(\overline{m},l)$. This implies that $F(m,l)$ must be strictly increasing over some subset of $(\overline{m}, M)$, which is a contradiction.

To complete the proof, we prove a series of claims that, when taken together, establish that $F(m,l)$ is a nonincreasing function of m for any $t > 0$ and $0 \leq l \leq t/2$. The claims are listed below:

1. For any $m \geq 0$, the inequality $\partial F(m,l)/\partial m \leq 0$ is equivalent to $G(m,l) := (a_m - b_l)c'_m - (c_m - d_l)a'_m \geq 0$.

2. $G(0,l) \geq 0$.

3. $G(m,l) \to 0$ as $m \to \infty$.

4. $G(\overline{m},l) \geq 0$ for all $\overline{m}$ that satisfy $\partial G(\overline{m},l)/\partial m = 0$.

5. To establish the previous claim (Claim 4), it is sufficient to show to $g(m,l) := (c_m - d_l)\phi(m+t/2) - (a_m - b_l)\phi(m+t) \geq 0$, for any $m \geq 0$.

6. $g(0, l) \geq 0$.

7. $g(m, l) \to 0$ as $m \to \infty$.

8. $g(\overline{m}, l) \geq 0$ for all $\overline{m}$ that satisfy $\partial g(\overline{m}, l)/\partial m = 0$.

The first claim establishes a necessary and sufficient condition for $F(m, l)$ to be nonincreasing in m. The next three claims together prove that $G(m, l) \geq 0$ for all $m \geq 0$. The fifth claim establishes a necessary and sufficient condition for the fourth claim. The final three claims are used to prove the fifth claim. Before we proceed with the proof of these eight claims, let us show how Claims 2, 3, and 4 establish that $G(m, l) \geq 0$ for all $m \geq 0$. Suppose there exists some $\hat{m} > 0$ such that $G(\hat{m}, l) < 0$. Note that $G(m, l)$ is a continuous and differentiable function of m. Since $G(0, l) \geq 0$, there exists some $m_1 < \hat{m}$ such that $G(m_1, l) < 0$ and $\partial G(m_1, l)/\partial m < 0$. Also, since $G(m, l) \to 0$ as $m \to \infty$, there exists some $m_2 > \hat{m}$ such that $G(m_2, l) < 0$ and $\partial G(m_2, l)/\partial m > 0$. By the intermediate value theorem, there exists an $\overline{m} \in (m_1, m_2)$ such that $\partial G(\overline{m}, l)/\partial m = 0$ and, by continuity, $G(\overline{m}, l) < 0$, which is a contradiction to Claim 4. A similar argument can be used to show how Claim 5 follows from Claims 6, 7, and 8.

Finally, we prove the eight claims.

Proof of Claim 1

$$\begin{aligned} F(m, l) &= \frac{a_m(c_m - d_l) + a_m - b_l - a_m c_m + b_l c_m}{c_m - d_l} \\ &= \frac{a_m(1 - d_l) - b_l(1 - c_m)}{c_m - d_l}. \end{aligned}$$

Taking the partial derivative of $F(m, l)$ with respect to m,

$$\begin{aligned} \frac{\partial F}{\partial m} &= \frac{a'_m(1 - d_l)(c_m - d_l) + b_l c'_m(c_m - d_l) - a_m(1 - d_l)c'_m + b_l(1 - c_m)c'_m}{(c_m - d_l)^2} \\ &= \frac{a'_m(1 - d_l)(c_m - d_l) - c'_m(-b_l(c_m - d_l) + a_m(1 - d_l) - b_l(1 - c_m))}{(c_m - d_l)^2} \\ &= \frac{a'_m(1 - d_l)(c_m - d_l) - c'_m(a_m - b_l)(1 - d_l)}{(c_m - d_l)^2}, \end{aligned}$$

which is clearly nonpositive if and only if (since $d_l \leq 1$)

$$(a_m - b_l)c'_m \geq (c_m - d_l)a'_m.$$

Proof of Claim 2

For any $0 \leq l \leq t/2$,

$$\begin{aligned} G(0,l) &= (a_0 - b_l)c'_0 - (c_0 - d_l)a'_0 \\ &= (\Phi(t/2) - \Phi(-t/2) - \Phi(t/2 - l) + \Phi(-t/2 + l))(\phi(t) + \phi(0)) \\ &\quad - (\Phi(t) - \Phi(0) - \Phi(t - l) + \Phi(l))2\phi(t/2). \end{aligned}$$

At $l = 0$, $G(0,0) = 0$, so it suffices to show that $G(0,l)$ is a nondecreasing function of l for $0 \leq l \leq t/2$ and $t > 0$. The derivative of $G(0,l)$ with respect to l is

$$\frac{dG(0,l)}{dl} = 2\phi(t/2 - l)(\phi(t) + \phi(0)) - (\phi(t - l) + \phi(l))2\phi(t/2),$$

and the nonnegativity of this expression is equivalent to

$$\frac{\phi(t - l) + \phi(l)}{\phi(\frac{t}{2} - l)} \leq \frac{\phi(t) + \phi(0)}{\phi(\frac{t}{2})}.$$

This inequality holds with equality when $l = 0$, so all we need to show is that the left hand side is a nonincreasing function of l for $0 \leq l \leq t/2$. The left hand side as a function of l can be written as

$$\begin{aligned} \frac{\phi(t - l) + \phi(l)}{\phi(\frac{t}{2} - l)} &= \frac{e^{-\frac{1}{2}(t-l)^2} + e^{-\frac{1}{2}l^2}}{e^{-\frac{1}{2}(\frac{t}{2}-l)^2}} \\ &= \frac{e^{-\frac{1}{2}(-2tl+t^2)} + 1}{e^{-\frac{1}{2}(-tl+\frac{t^2}{4})}} \\ &= e^{\frac{tl}{2} - \frac{3t^2}{8}} + e^{\frac{-tl}{2} + \frac{t^2}{8}}. \end{aligned}$$

Finally, taking the derivative of the left hand side function

$$\begin{aligned}\frac{d}{dl}\left(e^{\frac{tl}{2}-\frac{3t^2}{8}}+e^{\frac{-tl}{2}+\frac{t^2}{8}}\right) &= \frac{t}{2}e^{\frac{tl}{2}-\frac{3t^2}{8}}-\frac{t}{2}e^{\frac{-tl}{2}+\frac{t^2}{8}} \\ &= \frac{t}{2}e^{\frac{-tl}{2}+\frac{-3t^2}{8}}(e^{tl}-e^{\frac{t^2}{2}}) \\ &\leq 0,\end{aligned}$$

for $0 \leq l \leq t/2$.

Proof of Claim 3

This claim is straightforward since

$$\begin{aligned}a'_m &= 2\phi(m+t/2) \\ c'_m &= \phi(m+t)+\phi(m)\end{aligned}$$

and both expressions go to zero as m approaches infinity.

Proof of Claim 4

Using Claims 6, 7, and 8, we can establish the inequality in Claim 5, which, as we show next, is equivalent to Claim 4.

Proof of Claim 5

The partial derivative of $G(m,l)$ with respect to m is

$$\begin{aligned}\frac{\partial G}{\partial m}(m,l) &= (a_m-b_l)c''_m-(c_m-d_l)a''_m \\ &= -(a_m-b_l)((m+t)\phi(m+t)+m\phi(m))+(c_m-d_l)2(m+t/2)\phi(m+t/2) \\ &= -mG(m,l)-(a_m-b_l)t\phi(m+t)+(c_m-d_l))t\phi(m+t/2).\end{aligned}$$

Suppose that $\overline{m}$ satisfies $\partial G(\overline{m},l)/\partial m = 0$. Then,

$$\frac{\overline{m}G(\overline{m},l)}{t} = -(a_{\overline{m}}-b_l)\phi(\overline{m}+t)+(c_{\overline{m}}-d_l)\phi(\overline{m}+t/2).$$

If the condition in Claim 5 holds, then $G(\overline{m},l) \geq 0$, which establishes Claim 5.

Proof of Claim 6

To show $g(0,l) \geq 0$ for any $0 \leq l \leq t/2$, first note that $g(0,0) = 0$. Hence, it suffices to show that $g(0,l)$ is a nondecreasing function of l for $0 \leq l \leq t/2$.

$$\begin{aligned}
\frac{\partial g(0,l)}{\partial l} &= (\phi(t-l) + \phi(l))\phi(t/2) - 2\phi(t/2-l)\phi(t) \\
&= \frac{e^{-\frac{1}{2}(t-l)^2} + e^{-\frac{1}{2}l^2}}{e^{-\frac{1}{2}(\frac{t}{2}-l)^2}} - \frac{2e^{-\frac{1}{2}t^2}}{e^{-\frac{1}{2}(\frac{t}{2})^2}} \\
&= \frac{e^{tl-\frac{t^2}{2}} + 1}{e^{-\frac{t^2}{8}+\frac{tl}{2}}} - 2e^{-\frac{3}{8}t^2} \\
&= e^{-\frac{3}{8}t^2} e^{-\frac{tl}{2}} \left(e^{tl} + e^{\frac{t^2}{2}} - 2e^{\frac{tl}{2}} \right).
\end{aligned}$$

The expression

$$e^{tl} + e^{\frac{t^2}{2}} - 2e^{\frac{tl}{2}}$$

is nonnegative at $l = 0$, and its derivative with respect to l is

$$te^{lt} - te^{\frac{lt}{2}},$$

which is nonnegative for all $l \geq 0$. Hence, $g(0,l)$ is a nondecreasing function of l for $0 \leq l \leq t/2$.

Proof of Claim 7

This is obvious since $\phi(m + t/2)$ and $\phi(m + t)$ converge to zero as m approaches infinity.

Proof of Claim 8

The partial derivative of $g(m,l)$ with respect to m is

$$\begin{aligned}\frac{\partial g(m,l)}{\partial m} &= (\phi(m+t)+\phi(m))\phi(m+t/2) + (c_m - d_l)(-m-t/2)\phi(m+t/2) \\ &\quad - 2\phi(m+t/2)\phi(m+t) - (a_m - b_l)(-m-t)\phi(m+t) \\ &= (-m-t)g(m,l) + \phi(m+t/2)(c_m - d_l)t/2 \\ &\quad + (\phi(m+t)+\phi(m))\phi(m+t/2) - 2\phi(m+t/2)\phi(m+t).\end{aligned}$$

If $\overline{m}$ satisfies $\partial g(\overline{m}, l)\partial m = 0$, then

$$\begin{aligned}(\overline{m}+t)g(\overline{m},l) &= (\phi(\overline{m}+t)+\phi(\overline{m}))\phi(\overline{m}+t/2) \\ &\quad - 2\phi(\overline{m}+t/2)\phi(\overline{m}+t) + \phi(\overline{m}+t/2)(c_{\overline{m}} - d_l)t/2 \\ &= \phi(\overline{m}+t/2)(\phi(\overline{m}) - \phi(\overline{m}+t)) + \phi(\overline{m}+t/2)(c_{\overline{m}} - d_l)t/2,\end{aligned}$$

which is nonnegative since $\phi(m) > \phi(m+t)$ for any $m \geq 0$. □

Using Theorem 3.8, the optimization problem (3.24) can be simplified to

$$\begin{aligned}\underset{m,l,h}{\text{minimize}} \quad & t + 2m + 2(m+l)h \\ \text{subject to} \quad & (1+h)(\Phi(t+m) - \Phi(-m)) - h(\Phi(t-l) - \Phi(l)) \geq 1, \\ & l \leq t/2, \\ & m, l, h \geq 0.\end{aligned} \tag{3.26}$$

We can further simplify formulation (3.26) by noting that the main constraint is tight at an optimal solution.

Lemma 3.7. *There exists an optimal solution to* (3.26) *at which the constraint*

$$(1+h)(\Phi(t+m) - \Phi(-m)) - h(\Phi(t-l) - \Phi(l)) \geq 1 \tag{3.27}$$

is tight.

Proof: The coefficient of h in the objective function of formulation (3.26) is nonneg-

ative. The coefficient of h in (3.27),

$$\Phi(t+m) - \Phi(-m) - \Phi(t-l) + \Phi(l),$$

is positive. Hence, any solution (m, l, h) that satisfies (3.27) with a strict inequality, will have cost no better than a solution $(m, l, h-\epsilon)$, for ϵ positive and sufficiently small. □

Solving for h when equation (3.27) is satisfied with equality,

$$h = \frac{1 - \Phi(t+m) + \Phi(-m)}{\Phi(t+m) - \Phi(-m) - \Phi(t-l) + \Phi(l)}$$

and substituting it back into the objective function, we arrive at an unconstrained problem over the positive quadrant (together with the additional constraint $l \leq t/2$). The new objective function is

$$\begin{aligned} f(m,l) &:= t + 2m + 2(m+l)\frac{1 - \Phi(t+m) + \Phi(-m)}{\Phi(t+m) - \Phi(-m) - \Phi(t-l) + \Phi(l)} \\ &= t - 2l + 2(m+l)\frac{1 - \Phi(t-l) + \Phi(l)}{\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l)} \\ &= t + \frac{2m(1 - \Phi(t-l) + \Phi(l)) + 2l(1 - \Phi(m+t) + \Phi(-m))}{\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l)}. \end{aligned}$$

For future reference, its partial derivatives are

$$\begin{aligned} \frac{\partial f(m,l)}{\partial m} = &\frac{\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l) - (m+l)(\phi(m+t) + \phi(-m))}{(\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l))^2} \\ &\cdot 2(1 - \Phi(t-l) + \Phi(l)) \end{aligned} \tag{3.28}$$

$$\begin{aligned} \frac{\partial f(m,l)}{\partial l} = &\frac{\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l) - (m+l)(\phi(t-l) + \phi(l))}{(\Phi(m+t) - \Phi(-m) - \Phi(t-l) + \Phi(l))^2} \\ &\cdot 2(1 - \Phi(m+t) + \Phi(-m)). \end{aligned} \tag{3.29}$$

For completeness, the optimization problem (3.26) in its simplest form is

$$\begin{aligned} \underset{m,l}{\text{minimize}} \quad & f(m,l) \\ \text{subject to} \quad & l \leq t/2, \\ & m, l \geq 0. \end{aligned} \tag{3.30}$$

As we mentioned in the beginning of the subsection, setting $l = t/2$ recovers the margin problem from Section 3.1. We can check this by noting that

$$\begin{aligned} f(m, t/2) &= t + \frac{2m(1 - \Phi(t/2) + \Phi(t/2)) + t(1 - \Phi(m+t) + \Phi(-m))}{\Phi(m+t) - \Phi(-m) - \Phi(t/2) + \Phi(t/2)} \\ &= t + \frac{2m + t(1 - \Phi(m+t) + \Phi(-m))}{\Phi(m+t) - \Phi(-m)} \\ &= \frac{t + 2m}{\Phi(m+t) - \Phi(-m)}, \end{aligned}$$

which is the same as the objective function (3.3) from the margin problem.

Next, we establish that the ratio of t to σ can be used, as in the margin case, to characterize the optimal edge-enhanced intensity maps for a particular problem instance. Without a global characterization of $f(m, l)$, we will focus on local optimality. Recall that we have assumed $\sigma = 1$ in this section. Hence, references to t should be thought of as t/σ in general.

Theorem 3.9. *If $0 \leq t \leq \hat{u}$, a locally-optimal solution to the optimization problem* (3.30) *is $(m^*, l^*) = (0, t/2)$, where $\hat{u}$ is the unique positive solution to the equation*

$$\Phi(t) - \Phi(0) - t\phi(t/2) = 0. \tag{3.31}$$

Proof: First, let us show that there is exactly one positive root of the function

$$h(t) := \Phi(t) - \Phi(0) - t\phi(t/2).$$

The function $h(t)$ is a continuous function of t, $h(0) = 0$ and $\lim_{t\to\infty} h(t) = 1/2$.

Then, since

$$\lim_{t\to 0} h'(t) = \lim_{t\to 0}\left(\phi(t) + (\frac{t^2}{4} - 1)\phi(t/2)\right) = 0$$
$$\lim_{t\to 0} h''(t) = \lim_{t\to 0}\left(-t\phi(t) + (\frac{3t}{4} - \frac{t^3}{16})\phi(t/2)\right) = 0$$
$$\lim_{t\to 0} h^{(3)}(t) = \lim_{t\to 0}\left((t^2-1)\phi(t) + (\frac{3}{4} - \frac{3t^2}{8} + \frac{t^4}{64})\phi(t/2)\right) = -\frac{1}{4}\phi(0) < 0,$$
$$\lim_{t\to\infty} h(t) > 0,$$

the intermediate value theorem implies that there exists a positive root, $\hat{u}$, of $h(t)$. To show that $\hat{u}$ is the only positive root of $h(t)$, first note that

$$\begin{aligned} h'(t) &= \phi(t) + \left(\frac{t^2}{4} - 1\right)\phi(t/2) \\ &= e^{-\frac{t^2}{8}}\left(e^{-\frac{3t^2}{8}} - \left(1 - \frac{t^2}{4}\right)\right). \end{aligned}$$

Since $h''(0) = 0$ and $h^{(3)}(0) < 0$, for $t > 0$ sufficiently small,

$$e^{-\frac{3t^2}{8}} \leq 1 - \frac{t^2}{4}.$$

However, it is obvious that the reverse inequality holds for t sufficiently large, and that these two functions only cross once on the positive axis at a point we denote as $\overline{u}$. This implies that $h(t)$ is a quasiconvex function of t, and hence, $h(t)$ has a unique positive root at $\hat{u}$. Furthermore, $h(t) \leq 0$ only for $0 \leq t \leq \hat{u}$.

Numerically, $\hat{u} = 2.10905$. Since $t \leq \hat{u} < u^* = 2.28089$, Theorem 3.1 implies that the optimal margin size (in the absence of edge-enhancements) is 0 and moreover, that the partial derivative of $f(m,l)$ with respect to m at $(m,l) = (0, t/2)$ is strictly positive. What remains is to show that at $(m,l) = (0,t/2)$, $\partial f(m,l)/\partial l \leq 0$. In other words, we need to show that adding intensity modulation (decreasing l from $t/2$) will not improve the solution. The sign of

$$\frac{\partial f(0,t/2)}{\partial l} = \frac{2(1 - \Phi(t) + \Phi(0))(\Phi(t) - \Phi(0) - t\phi(t/2))}{(\Phi(t) - \Phi(0))^2}$$

is determined by the sign of $h(t)$. Since $h(t) \leq 0$ for $0 \leq t \leq \hat{u}$, $\partial f(0, t/2)/\partial l \leq 0$ as well.

Finally, consider an arbitrary directional derivative along the line $l = t/2 - \alpha m$ for some $\alpha \in (0, \infty)$ at the point $(0, t/2)$. Since we are moving in the negative l direction, then the directional derivative will be a convex combination of $-\partial f(0, t/2)/\partial l$ and $\partial f(0, t/2)/\partial m$, which is positive since $\partial f(0, t/2)/\partial m$ is positive. Hence, $(m^*, l^*) = (0, t/2)$ is a locally optimal solution to problem (3.30). □

The previous proposition shows that for t/σ sufficiently small, there exists a locally-optimal solution that consists of a margin of size 0, and furthermore, the added flexibility of having intensity modulation offers no advantage in terms of dose reduction. Next, we show that the existence of a locally-optimal solution that exhibits intensity modulation when t is larger than $\hat{u} = 2.109$. Before we do that, we state a result that was proved in the course of proving Theorem 3.9, which will be useful later.

Corollary 3.4. *Let $\hat{u}$ be defined as in Theorem 3.9. For $0 \leq t \leq \hat{u}$,*

$$2(\Phi(t) - \Phi(0)) - t\phi(t/2) \leq 0,$$

with equality at $t = \hat{u}$, and for $t > \hat{u}$,

$$2(\Phi(t) - \Phi(0)) - t\phi(t/2) > 0.$$

Theorem 3.10. *If $\hat{u} < t \leq u^*$, a locally-optimal solution to problem (3.30) is $m^* = 0$ and l^*, where l^* is the unique solution to*

$$\Phi(t) - \Phi(0) - \Phi(t - l) + \Phi(l) - l(\phi(t - l) + \phi(l)) = 0 \tag{3.32}$$

on the interval $(0, t/2)$.

Proof: Since $t \leq u^*$, it is still not desirable to increase m from 0 at the point

$(m, l) = (0, t/2)$. However, since $t > \hat{u}$, Corollary 3.4 implies that

$$\frac{\partial f(0, t/2)}{\partial l} > 0.$$

So by decreasing l from $t/2$, we can lower the objective function, which shows that the solution $(m, l) = (0, t/2)$ is not locally optimal. The optimal l is not 0 either, since

$$\begin{aligned}\frac{\partial f(0,0)}{\partial l} &= \lim_{l \to 0} \frac{2(1 - \Phi(t) + \Phi(0))(\Phi(t) - \Phi(0) - \Phi(t-l) + \Phi(l) - l(\phi(t-l) + \phi(l)))}{(\Phi(t) - \Phi(0) - \Phi(t-l) + \Phi(l))^2} \\ &= \frac{(1 - \Phi(t) + \Phi(0))(-t\phi(t))}{(\phi(t) + \phi(0))^2} < 0\end{aligned}$$

for any $t > 0$, from repeated invocations of L'Hôpital's rule. Thus, by the continuity of $\partial f(0, l)/\partial l$ as a function of l, there exists an $l \in (0, t/2)$ that satisfies

$$\frac{\partial f(0, l)}{\partial l} = 0.$$

We now show that there is a unique solution, l^*, to the equation $\partial f(0, l)/\partial l = 0$, and that it is local minimum of $f(0, l)$. The sign of $\partial f(0, l)/\partial l$ is determined (cf. equation (3.29)) by the sign of the continuous function

$$g(l) := \Phi(t) - \Phi(0) - \Phi(t - l) + \Phi(l) - l(\phi(t - l) + \phi(l)).$$

The derivative of $g(l)$ is

$$g'(l) = l(l\phi(l) - (t - l)\phi(t - l)).$$

It can be shown that $g(0) = g'(0) = 0$, $g''(0) < 0$ and $g(t/2) = \Phi(t) - \Phi(0) - t\phi(t/2) > 0$ (from Corollary 3.4 for $t > \hat{u}$), thus, there must be at least one root of $g(l)$ for $0 < l < t/2$. We wish to show that $g(l)$ has exactly one root in the interval $(0, t/2)$, and therefore it is sufficient to show that $g'(l)$ has a unique root in the same interval (if $g(l)$ has two roots in the interval $(0, t/2)$, then $g(l)$ must "change directions" at

least twice, which means that it would have at least two stationary points). First, we simplify $g'(l)$.

$$\begin{aligned} g'(l) &= l(l\phi(l) - (t-l)\phi(t-l)) \\ &= \frac{l}{\sqrt{2\pi}}(le^{-\frac{1}{2}l^2} - (t-l)e^{-\frac{1}{2}(t-l)^2}) \\ &= \frac{l}{\sqrt{2\pi}}e^{-\frac{1}{2}l^2}(l - (t-l)e^{-\frac{1}{2}t^2}e^{tl}). \end{aligned}$$

The sign of $g'(l)$ is determined by the relationship between the functions

$$\begin{aligned} g_1(l) &= l \\ g_2(l) &= (t-l)e^{-\frac{1}{2}t^2}e^{tl}. \end{aligned}$$

Note that

$$\begin{aligned} g_2'(l) &= e^{-\frac{1}{2}t^2}e^{tl}((t-l)t - 1) \\ g_2''(l) &= e^{-\frac{1}{2}t^2}e^{tl}(t((t-l)t-1) - t) \\ &= te^{-\frac{1}{2}t^2}e^{tl}((t-l)t - 2). \end{aligned}$$

Recall that $t > \hat{u} > 2$ and $l \leq t/2$, hence,

$$(t-l)t - 1 > (t-l)t - 2 \geq \frac{t^2}{2} - 2 > 0.$$

These inequalities imply that $g_2(l)$ is strictly increasing and strictly convex. Furthermore, the inequalities

$$\begin{aligned} g_1(0) = 0 < te^{-\frac{1}{2}t^2} &= g_2(0) \\ g_1(t/2) = t/2 &= g_2(t/2) \\ g_1'(t/2) = 1 < \frac{t^2}{2} - 1 &= g_2'(t/2) \end{aligned}$$

imply that $g_1(l)$ and $g_2(l)$ must cross exactly once in the interval $(0, t/2)$. Since

$g_1(l)$ and $g_2(l)$ cross exactly once in the interval $(0, t/2)$, $g'(l)$ has exactly one root in the same interval, which implies $g(l)$ also has exactly one root, l^*, in the interval. Furthermore, since the lowest order non-zero derivative of $g(l)$ at $l = 0$ is negative, $g(l)$ is negative for $l < l^*$ and positive for $l > l^*$. The same relationships hold for $\partial f(0, l)/\partial l$, and therefore l^* is the unique local minimum of $f(0, l)$ as a function of $l \in [0, t/2]$.

Finally, we need to show that at $(0, l^*)$, there is no improvement in the objective function if we increase m. To do this, it is sufficient to establish that

$$\frac{\partial f(0, l^*)}{\partial m} > 0.$$

The sign of $\partial f(0, l)/\partial m$ is determined (cf. equation (3.28)) by the sign of

$$h(l) := \Phi(t) - \Phi(0) - \Phi(t - l) + \Phi(l) - l(\phi(t) + \phi(0)).$$

The first two derivatives of $h(l)$ are

$$h'(l) = \phi(t - l) + \phi(l) - \phi(t) - \phi(0)$$
$$h''(l) = (t - l)\phi(t - l) - l\phi(l).$$

Note that $h(0) = 0$, $h'(0) = 0$, and $h''(0) > 0$. Since $h(t/2) = \Phi(t) - \Phi(0) - t/2(\phi(t) + \phi(0)) \geq 0$ (from Corollary 3.2 for $t \leq u^*$), there are two possibilities to consider. Either $h(l)$ is nondecreasing on $[0, t/2]$ without any local extrema or $h(l)$ has a local maximum on $(0, t/2]$. In the first case, it is clear that $h(l) > 0$ for $0 \leq l \leq t/2$ (since $h''(0) > 0$), hence $h(l^*) > 0$ and the result is proved. Now suppose that $h(l)$ has at least one local maximum for $0 \leq l \leq t/2$. In fact, we show that in this case, $h(l)$ has exactly one local maximum for $0 \leq l \leq t/2$. To see this, note that $h''(l)$ has exactly one root for $0 \leq l \leq t/2$ (using the same argument above which showed that $g_1(l)$ and $g_2(l)$ cross exactly once on this interval). This implies that $h'(l)$ changes sign exactly once in this interval (from positive to negative, since $h''(0) > 0$), which is equivalent to $h(l)$ having exactly one local maximum. Let $\bar{l}$ denote this local maximum. Finally,

to show that $h(l^*) > 0$, it suffices to show that $l^* \leq \bar{l}$, since $h(l)$ is nondecreasing and positive for $0 < l \leq \bar{l}$.

Consider the function $h(l) - lh'(l)$ as a function of l. For l sufficiently close to 0, $h(l)$ is strictly convex since $h''(0) > 0$. Thus,

$$h(l) - lh'(l) < h(0) = 0,$$

for l sufficiently close to 0. At $\bar{l}$,

$$h(\bar{l}) - \bar{l}h'(\bar{l}) = h(\bar{l}) > 0.$$

The intermediate value theorem guarantees the existence of an $\hat{l} \in (0, \bar{l})$ such that $h(\hat{l}) - \hat{l}h'(\hat{l}) = 0$. Therefore,

$$\begin{aligned}
0 &= h(\hat{l}) - \hat{l}h'(\hat{l}) \\
&= \Phi(t) - \Phi(0) - \Phi(t - \hat{l}) + \Phi(\hat{l}) - \hat{l}(\phi(t) + \phi(0)) \\
&\quad - \hat{l}(\phi(t - \hat{l}) + \phi(\hat{l}) - \phi(t) - \phi(0)) \\
&= \Phi(t) - \Phi(0) - \Phi(t - \hat{l}) + \Phi(\hat{l}) - \hat{l}(\phi(t - \hat{l}) + \phi(\hat{l})).
\end{aligned}$$

However, recall that l^* is the unique solution to the equation

$$\Phi(t) - \Phi(0) - \Phi(t - l) + \Phi(l) - l(\phi(t - l) + \phi(l)) = 0,$$

therefore, $l^* = \hat{l} < \bar{l}$.

Since the partial derivative of $F(m, l)$ with respect to m at $(m, l) = (0, l^*)$ is strictly positive, the directional derivatives along directions other than the m- and l-axes at $(0, l^*)$ will also be positive. □

Table 3.4 displays numerical values for the locally-optimal l^* from Theorem (3.10) as a function of the ratio of the tumor size to σ, for values of this ratio in the interval $(\hat{u}, u^*)$. The optimal l^* seems quite sensitive to this ratio, as it drops quickly for small

increases in t. The corresponding edge-enhancement height is given by h^*. This h^* was computed to be the height that brought the dose at the edge of the tumor up to the requirement dose.

Table 3.4: The optimal l^* and h^* for particular values of the tumor size, relative to the size of the standard deviation of the Gaussian pdf.

Tumor size	2.12	2.14	2.16	2.18	2.20	2.22	2.24	2.26	2.28
l^*	0.886	0.774	0.698	0.638	0.588	0.545	0.507	0.473	0.443
h^*	1.280	1.474	1.642	1.804	1.966	2.130	2.299	2.473	2.653

Finally, we show that for $t > u^*$, any non-edge-enhanced intensity map can always be outperformed by an intensity map that does have edge-enhancements.

Theorem 3.11. *Suppose $t > u^*$. For any solution $(m, t/2)$, there exists a solution (m^*, l^*) such that $f(m^*, l^*) < f(m, t/2)$, $l^* < t/2$, and m^* is the unique solution to the equation $f(m, t/2) = 0$.*

Proof: Consider the function $f(m, t/2)$ as a function of m. By Theorem 3.1 for $t > u^*$, there exists a unique minimum of this function, at some $m^* > 0$. Hence, $f(m^*, t/2) \leq f(m, t/2)$. To establish that there exists an l^* such that $f(m^*, l^*) < f(m, t/2)$ for any $m \geq 0$, we will show that $f(m^*, l^*) < f(m^*, t/2)$. To accomplish this, we must show that $f(m^*, l)$ is an increasing function of l at $l = t/2$.

Consider $\partial f(m, l)/\partial l$ at $l = t/2$, as a function of m, and note from (3.29), that its sign will be the same as the sign of the function

$$g(m) := \Phi(m + t) - \Phi(-m) - (t + 2m)\phi(t/2).$$

We wish to show that at m^*, $g(m)$ is positive, which is equivalent to showing that $f(m^*, l)$ is increasing in l at $l = t/2$.

Note that $g(0) > 0$ since $t > \hat{u}$ (Corollary 3.4), and as $m \to \infty$, $g(m) \to -\infty$.

Furthermore, $g(m)$ is concave since

$$g'(m) = \phi(m+t) + \phi(-m) - 2\phi(t/2)$$
$$g''(m) = -(m+t)\phi(m+t) - m\phi(-m) < 0,$$

for any $m \geq 0$. Hence, $g(m)$ has exactly one root at some $\hat{m} > 0$. For $0 \leq m < \hat{m}$, $g(m) > 0$, which implies that $\partial f(m, t/2)/\partial l > 0$ over the same interval of m. What remains is to show that $m^* < \hat{m}$. Note that

$$\begin{aligned} g'(0) &= \phi(t) + \phi(0) - 2\phi(t/2) \\ &> 2\frac{\Phi(t) - \Phi(0)}{t} - 2\phi(t/2) \\ &> 0, \end{aligned}$$

where the first inequality comes from Corollary 3.2 for $t > u^*$, and the second inequality comes from Corollary 3.4 for $t > \hat{u}$. Now, let us define

$$h(m) := \Phi(m+t) - \Phi(-m) - \frac{t+2m}{2}(\phi(m+t) + \phi(m))$$

and note that Proposition 3.3 implies m^* satisfies $h(m^*) = 0$. Since

$$h'(m) = \frac{t+2m}{2}((m+t)\phi(m+t) + m\phi(m)) > 0,$$

for all $m \geq 0$, $h(m)$ is increasing in m. Furthermore, $h(0) < 0$ from Corollary 3.2 for $t > u^*$. Consider the function defined by the difference of $g(m)$ and $h(m)$

$$g(m) - h(m) = (t+2m)\left(\frac{\phi(m+t) + \phi(m)}{2} - \phi(t/2)\right).$$

It can be shown that $g(0) - h(0) > 0$ and $g(m) - h(m)$ is decreasing in m, therefore, $g(m) - h(m)$ has a unique root denoted $\overline{m}$. Finally, we show that $m^* < \overline{m} < \hat{m}$. Note that

$$g'(\overline{m}) = \frac{2}{t+2\overline{m}}(g(\overline{m}) - h(\overline{m})) = 0.$$

Since $g(m)$ is concave, $g(m)$ is maximized at $\overline{m}$. So $g(\overline{m}) > 0$, and thus, $\overline{m} < \hat{m}$. Furthermore, $h(\overline{m}) = g(\overline{m}) > 0$, which implies that $m^* < \overline{m}$ since $h(m)$ is increasing in m and has a root at m^*. Since $m^* < \hat{m}$, $g(m^*) > 0$, as desired. □

While we have shown that a margin intensity map can always be outperformed by an edge-enhanced intensity map for $t > u^*$, the exact characterization of a locally-optimal intensity map in this case remains open. We can take our analysis one step further by finding a local minimum in the l-direction, at the m^* defined in the statement of Theorem 3.11.

Proposition 3.8. *Suppose $t > u^*$ and let m^* be the unique minimum of the function $f(m, t/2)$. Then, the value of l that minimizes $f(m^*, l)$ is the unique root of*

$$\Phi(m^* + t) - \Phi(-m^*) - \Phi(t - l) + \Phi(l) - (m^* + l)(\phi(t - l) + \phi(l)).$$

Proof: To find the optimal l, we need to solve $\partial f(m^*, l)/\partial l = 0$, or equivalently

$$h(l) := \Phi(m^* + t) - \Phi(-m^*) - \Phi(t - l) + \Phi(l) - (m^* + l)(\phi(t - l) + \phi(l)) = 0. \quad (3.33)$$

This is sufficient since we will show that there is a unique solution, l^*, to equation (3.33), which is a local minimum of $h(l)$. First, note that $h(t/2) > 0$, which was proved in the previous proposition (equivalent to $g(m^*) > 0$ from the previous proposition). Next, we claim that $h(0) < 0$. Indeed, the condition $h(0) < 0$ is equivalent to

$$\frac{\Phi(m^* + t) - \Phi(t)}{m^*} + \frac{\Phi(0) - \Phi(-m^*)}{m^*} < \phi(t) + \phi(0),$$

which is true since

$$\frac{\Phi(m + t) - \Phi(t)}{m} < \phi(t)$$

and

$$\frac{\Phi(0) - \Phi(-m)}{m} < \phi(0)$$

for any $m > 0$, due to the concavity of $\Phi(m+t) - \Phi(t)$ and convexity of $\Phi(0) - \Phi(-m)$,

respectively. Finally, we show that $h(l)$ crosses zero exactly once on $(0, t/2)$. We recognize part of the derivative of $h(l)$,

$$h'(l) = -(m^* + l)((t - l)\phi(t - l) - l\phi(l)),$$

from the proof of Theorem 3.10. There, we proved the existence of a unique solution to the equation $(t-l)\phi(t-l) - l\phi(l) = 0$ for $0 \leq l \leq t/2$. Furthermore, $h'(l)$ changes sign once, from negative to positive, in $(0, t/2)$. Then, since $h(0) < 0 < h(t/2)$, there is exactly one solution, l^*, to the equation $h(l) = 0$ for $0 < l < t/2$. In fact, l^* is a local minimum of $f(m^*, l)$ as a function of l since $h(l)$ is negative for $l < l^*$ and positive for $l > l^*$. □

We computed locally-optimal intensity maps for different values of t, using a standard function minimization routine in Matlab. Figures 3-8, 3-9, and 3-10 summarize the optimal m^*, l^*, and h^* as a function of t, respectively.

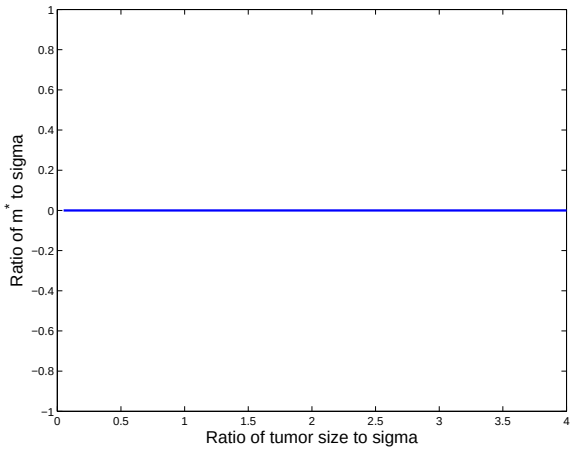


Figure 3-8: The ratio of the optimal margin size to σ in an edge-enhanced intensity map as a function of the ratio of the tumor size to σ for a 1D tumor.

While we do not have a full, analytical characterization of the function $f(m, l)$, we conjecture that Figures 3-8, 3-9, and 3-10 illustrate the structure of globally-optimal edge-enhanced intensity maps. In other words, we believe that for $0 \leq t \leq \hat{u}$, the globally-optimal solution to problem (3.30) is $(m, l) = (0, t/2)$, and for $t > \hat{u}$, the

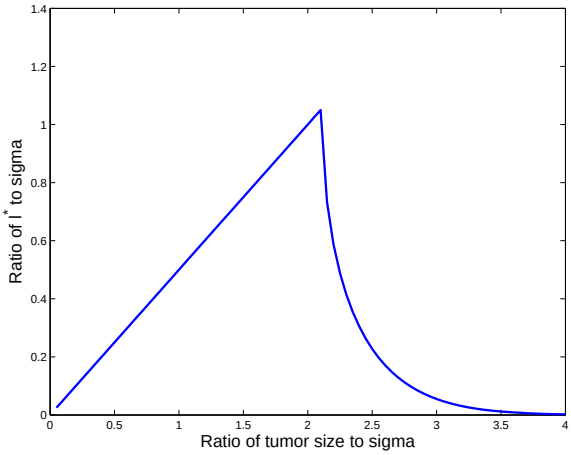


Figure 3-9: The ratio of the optimal l^* to σ in an edge-enhanced intensity map as a function of the ratio of the tumor size to σ for a 1D tumor.

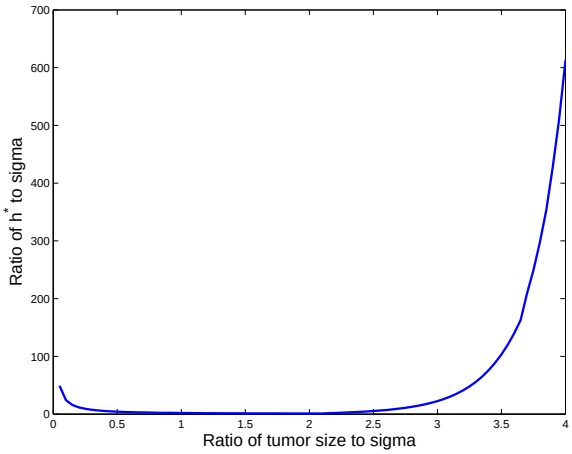


Figure 3-10: The ratio of the optimal h^* to σ in an edge-enhanced intensity map as a function of the ratio of the tumor size to σ for a 1D tumor.

globally-optimal solution is $(m, l) = (0, l^*)$, where l^* is the unique solution of the equation

$$\frac{\partial f(0,l)}{\partial l} = 0.$$

3.6 Optimal edge-enhancements with robustness

We now consider uncertainty in the parameters of the Gaussian motion distribution, while focused on edge-enhanced intensity maps. The uncertain parameters, denoted $\tilde{\mu}$

and $\tilde{\sigma}$, lie in the intervals $[\underline{\mu}, \overline{\mu}]$ and $[\underline{\sigma}, \overline{\sigma}]$, respectively. Our problem is essentially the same as the one posed in the nominal edge-enhancement formulation (3.20), except for the extra requirement that the solution remain feasible for all realizations of $\tilde{\mu}$ and $\tilde{\sigma}$.

3.6.1 Summary and Highlights

In this section, we show how to deterministically incorporate uncertainty in μ into the robust problem. However, it remains open whether or not the full, deterministic robust counterpart of the robust problem has the same (or similar) structure as the underlying nominal problem. The difficulty comes from the uncertainty in σ, which we show affects the dose delivered in non-trivial ways. We point to evidence that the robust counterpart has a similar structure as the nominal problem, but a formal proof to confirm or reject this hypothesis is left to future research.

3.6.2 Analysis

As in the other sections when uncertainty in μ is introduced, we re-center the coordinate axes on the middle of the "effective tumor," which in this case (as in the 1D robust margin case) is written as

$$\overline{\mathcal{T}}(0) := [-t/2 - (\overline{\mu} - \underline{\mu})/2, t/2 + (\overline{\mu} - \underline{\mu})/2].$$

We also need to make a slight modification to the constraint $l \leq t/2$, which is meant to ensure that the edge-enhancements do not physically overlap. Now that $\tilde{\mu}$ is uncertain, it is possible to have l greater than $t/2$ since we should be considering the effective tumor when designing the static intensity map. Therefore, the upper bound on l is

$$l \leq t/2 + (\overline{\mu} - \underline{\mu})/2.$$

Let us define $\hat{\mu} := \overline{\mu} - \underline{\mu}$. Taking into account the effective tumor, the location of

the "effective edge-enhancements" centered at the origin is

$$\overline{\mathcal{E}}(0) = [-t/2 - m + \hat{\mu}/2, -t/2 + l + \hat{\mu}/2] \cup [t/2 - l + \hat{\mu}/2, t/2 + m + \hat{\mu}/2].$$

We write the static intensity map as

$$\delta_{m,l,h}(x) = \begin{cases} 1, & x \in \overline{\mathcal{T}}(0), \\ 1+h, & x \in \overline{\mathcal{T}}(0) \backslash \overline{\mathcal{E}}(0), \\ 0, & \text{otherwise}, \end{cases}$$

and the dose delivered after convolution with a Gaussian pdf with standard deviation $\tilde{\sigma}$ as

$$\begin{aligned} d_{m,l,h}(y;0,\tilde{\sigma}) &= \int_{-\infty}^{\infty} \delta_{m,l,h}(y-x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\tilde{\sigma}^2}} dx \\ &= (1+h)\Phi\left(\frac{y + t/2 + m + \hat{\mu}/2}{\tilde{\sigma}}\right) - h\Phi\left(\frac{y + t/2 - l + \hat{\mu}/2}{\tilde{\sigma}}\right) \\ &\quad + h\Phi\left(\frac{y - t/2 + l - \hat{\mu}/2}{\tilde{\sigma}}\right) - (1+h)\Phi\left(\frac{y - t/2 - m - \hat{\mu}/2}{\tilde{\sigma}}\right). \end{aligned}$$

We now formulate our robust problem as the following mathematical program.

$$\begin{aligned} \underset{m,l,h}{\text{minimize}} \quad & t + 2m + 2(m+l)h \\ \text{subject to} \quad & d_{m,l,h}(y;0,\tilde{\sigma}) \geq 1, \quad \forall\, y \in \overline{\mathcal{T}}(0), \forall\, \tilde{\sigma} \in [\underline{\sigma}, \overline{\sigma}], \\ & l \leq t/2 + (\overline{\mu} - \underline{\mu})/2, \\ & m, l, h \geq 0. \end{aligned} \tag{3.34}$$

Recall that in the robust margin problem (in both 1D and 3D), we showed that the uncertain optimization problem involving $\tilde{\sigma}$ could be transformed back into a deterministic problem using $\overline{\sigma}$ instead (cf. Theorem 3.2 and Theorem 3.5). In the margin problems, these transformations relied on the fact that the dose function $d_m(y;0,\tilde{\sigma})$ at the edge of the effective tumor ($y_R := t/2 + \hat{\mu}/2$) is a decreasing function of $\tilde{\sigma}$ for any $m, l, h \geq 0$. The fact that $d_m(y_R;0,\tilde{\sigma})$ is decreasing in $\tilde{\sigma}$, allowed us to

determine that the minimum dose delivered at the edge of the tumor would occur when $\tilde{\sigma}$ is largest. Hence, in the margin problem, we produce a robust solution by simply protecting against the largest value of $\tilde{\sigma}$, namely $\overline{\sigma}$.

Unfortunately, in the edge-enhancement case, $d_{m,l,h}(y; 0, \tilde{\sigma})$ is not decreasing in $\tilde{\sigma}$ or even quasiconcave in $\tilde{\sigma}$ at the edge of the effective tumor. Note that if $d_{m,l,h}(y_R; 0, \tilde{\sigma})$ is quasiconcave in $\tilde{\sigma}$, it would suffice to protect against only the smallest and largest values of $\tilde{\sigma}$. However, as Figures 3-11 and 3-12 show, $d_{m,l,h}(y_R; 0, \tilde{\sigma})$ can be a different function of $\tilde{\sigma}$ for different values of m, l, and h.

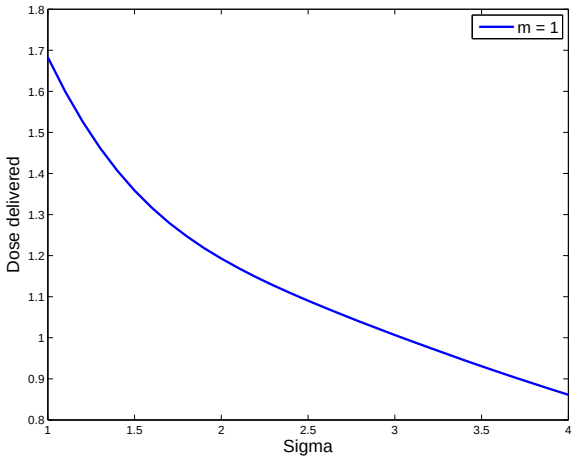


Figure 3-11: Dose delivered as a function of $\tilde{\sigma}$ for $t = 4, m = 1, l = 0.2, h = 2$.

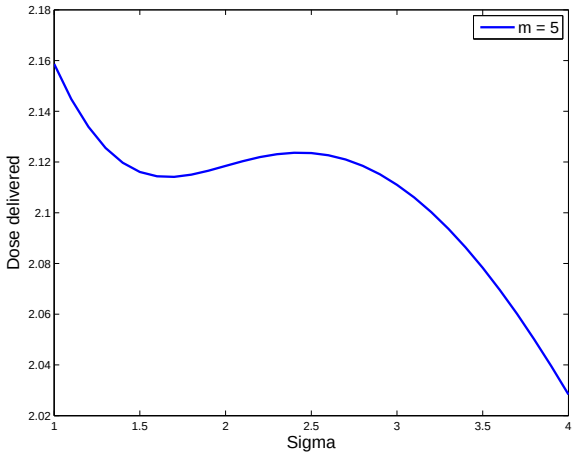


Figure 3-12: Dose delivered as a function of $\tilde{\sigma}$ for $t = 4, m = 5, l = 0.2, h = 2$.

We find that if m is sufficiently small (depending on the values of t, l, h), then $d_{m,l,h}(y_R; 0, \tilde{\sigma})$ is decreasing in $\tilde{\sigma}$. Therefore, it seems possible to establish that the robust counterpart of formulation (3.34) has the same structure as the nominal problem (3.30), if we can show that we do not need to consider m too large. That is, if we only focus on values of m that satisfy

$$-\frac{1+h}{\tilde{\sigma}}\left(\frac{t+m}{\tilde{\sigma}}\phi\left(\frac{t+m}{\tilde{\sigma}}\right)+\frac{m}{\tilde{\sigma}}\phi\left(\frac{m}{\tilde{\sigma}}\right)\right)-\frac{h}{\tilde{\sigma}}\left(-\frac{t-l}{\tilde{\sigma}}\phi\left(\frac{t-l}{\tilde{\sigma}}\right)+\frac{l}{\tilde{\sigma}}\phi\left(\frac{l}{\tilde{\sigma}}\right)\right)\leq 0,$$

which is equivalent to

$$\frac{\partial d_{m,l,h}(y_R; 0, \tilde{\sigma})}{\partial \tilde{\sigma}} \leq 0,$$

then we can show that the robust counterpart of formulation (3.34) has the same structure as the nominal problem (3.30), with $\overline{\sigma}$ in place of σ. While Figure 3-8 seems to suggest that we may restrict our attention to such small values of m (namely, $m = 0$), this conjecture remains open.

3.7 Conclusions

In this chapter, we constructed globally-optimal margin intensity maps and locally-optimal edge-enhanced intensity maps that minimize the total dose delivered while ensuring the tumor receives enough dose in the face of motion that follows a Gaussian distribution. We then extended the analysis to include uncertainty in the parameters of the Gaussian distribution, and established deterministic robust counterparts for both the one-dimensional and three-dimensional margin problems.

For a one-dimensional tumor and focusing only on margin intensity maps, we showed that if the ratio of the tumor size to the standard deviation of the Gaussian is less than 2.281, then the optimal solution involves no margin and uses a uniform scaling of the static intensity map in order to compensate for the motion. If the ratio of the tumor size to the standard deviation of motion exceeds 2.281, then the optimal solution has a positive margin size, which can be determined numerically as the unique solution to a nonlinear equation. Furthermore, we showed that the robust

counterpart of this problem has exactly the same structure as the nominal problem. Hence, the same margin threshold is present in the robust problem.

Next, we considered a three-dimensional tumor, while still focusing on margin intensity maps. We showed that in the nominal problem, the optimal margin size is always positive, and can be determined as the unique solution to a nonlinear equation. However, in the robust problem, a series of thresholds determines whether or not the optimal margin size would be positive. If the ratio of the tumor size to the length of the interval containing all realizable means of the Gaussian is more than 0.139, then the optimal margin size is positive. If this ratio is at most 0.139, then the optimal margin size may be positive or zero, depending on the ratio of the tumor size to the largest value of the Gaussian standard deviation.

Finally, for a one-dimensional tumor and focusing on edge-enhanced intensity maps, we showed that if the ratio of the tumor size to the standard deviation of motion is less than 2.109, then a locally-optimal solution involves a pure scaling of the intensity with no intensity modulation. If this ratio is between 2.109 and 2.281, then a locally-optimal solution exhibits edge-enhancements, where the width of the edge-enhancement is equal to the unique solution to a nonlinear equation. Finally, if this ratio is larger than 2.281, then an edge-enhanced solution will always outperform one that does not exhibit edge-enhancements.

The work presented in this chapter can be extended in multiple directions. First, for the edge-enhancements problem with a one-dimensional tumor, we conjecture that the locally-optimal solutions we established are, in fact, globally-optimal solutions. Furthermore, we believe that the edge-enhanced intensity maps will always have a zero margin size, and instead, rely on intensity modulation within the tumor to achieve tumor coverage. Therefore, we believe there will only be one threshold, at 2.109, that will separate the different kinds of optimal intensity maps.

The characterization of a deterministic robust counterpart for the one-dimensional edge-enhancement problem also remains open. Our belief is that the robust counterpart will have a similar (if not identical) structure to the nominal problem. Another extension is to consider edge-enhanced intensity maps for three-dimensional tumors.

In this case, even defining the edge-enhancement variables is not a straightforward exercise.

Throughout this chapter, we focused solely on a one-dimensional Gaussian as our motion distribution. Future research should consider three-dimensional motion distributions so as to generalize the results from this chapter. In addition to examining other motion distributions, it would be interesting to separate the effect of motion blurring from penumbra blurring, which we have implicitly combined by using a single Gaussian distribution and assuming infinitely sharp beamlets of radiation. This can be accomplished by having separate standard deviation parameters in the analysis.

Overall, the work in this chapter provides new opportunities for analytical studies of optimal intensity maps under motion and motion uncertainty. Insights gleaned from such analyses will lead to a better understand of topological characteristics of desirable intensity maps, as well as rules of thumb to guide treatments in practice. While the work presented in this chapter is mainly mathematical, it is important to not lose sight of the underlying medical motivation, and whenever possible, to try and get as much intuition as possible about practical implications of our analytical studies.

Chapter 4

Robust Intensity-Modulated Proton Therapy

In this chapter, we present a robust approach to the optimization of intensity-modulated proton therapy treatments under uncertainty. The physics of the delivery of radiation via a stream of protons differs markedly from that of a stream of photons.

Unlike the photons used in conventional radiation therapy, protons have the ability to deliver their maximum energy at a well-defined depth within the body. As a result, proton therapy treatment plans are typically better able to target the tumor and spare healthy tissue than their photonic counterparts. This increased precision also necessitates sound management of the uncertainty in the treatment delivery. For example, if the maximum energy of a stream of protons is realized at a depth that is farther than expected, a significant amount of dose could be delivered to a nearby critical structure. In this chapter, we study uncertainty in the "range" of the protons, and present a robust formulation for this problem. Our goal is to show that a robust approach can produce intuitive and clinically meaningful results in this uncertain optimization problem.

The remainder of this chapter is organized as follows. In Section 4.1, we provide some background on intensity-modulated proton therapy. In Sections 4.2 and 4.3, we discuss the sources of proton range uncertainty, and describe how we model this uncertainty, respectively. In Section 4.4, we present the geometry and basic charac-

teristics of the specific problem under consideration. In Section 4.5, we present the robust formulation, along with a nominal (no uncertainty) formulation. We compare the performance of these formulations in Section 4.6, and then summarize our findings in Section 4.7.

4.1 IMPT background

Proton therapy is one of the most advanced methods to treat cancer using radiation from an external beam. While most radiation oncology centers in the United States can deliver photon therapy treatments, there are only five proton therapy centers currently in operation in the US:

- Loma Linda University Medical Center, Loma Linda, CA,
- Midwest Proton Radiotherapy Institute, Bloomington, IN,
- M. D. Anderson Cancer Center, Houston, TX,
- University of Florida Proton Therapy Institute, Jacksonville, FL,
- Francis H. Burr Proton Therapy Center, Boston, MA,

and another five either under construction or in the planning stages (parentheses indicate expected year of first treatment):

- Hampton University Proton Beam Therapy Center, Hampton, VA, (2009),
- Roberts Proton Therapy Center, Philadelphia, PA, (2009),
- ProCure Proton Therapy Center, Oklahoma City, OK, (2009),
- Seattle Cancer Care Alliance Proton Therapy Center, Seattle, WA, (2010),
- Northern Illinois Proton Therapy Center, DeKalb, IL, (2011).

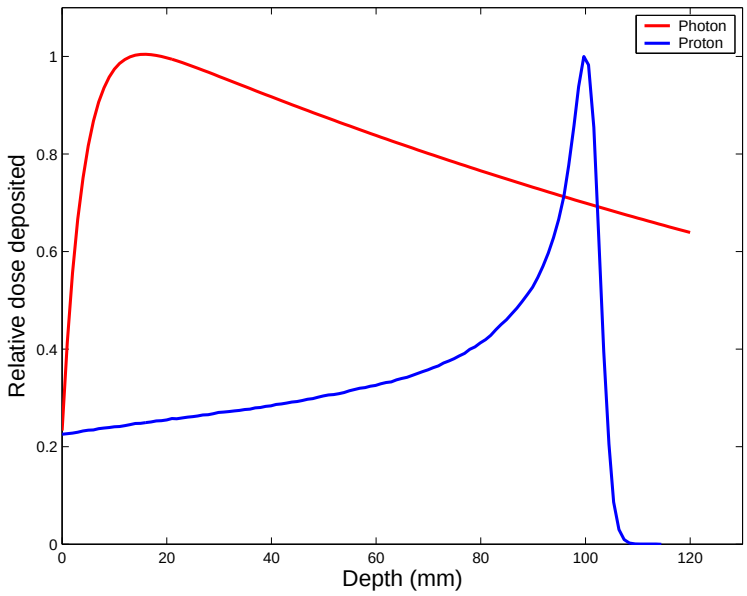


Figure 4-1: The depth-dose curve for a photon and proton.

The main difference between proton therapy and photon therapy that we will focus on is the relationship between dose deposited and distance traveled in the body by the respective particles. Figure 4-1 illustrates this difference.

Figure 4-1 shows that while photons deposit their maximum energy near the body surface, protons can deposit their maximum energy deep inside the body. The peak in the proton curve is known as the *Bragg peak*. By energizing photons to different velocities, the location of the Bragg peak or the "range" of the proton can be adjusted.

Intensity-modulated proton therapy (IMPT) is similar to intensity-modulated photon therapy in that the beam of radiation is divided up into many little beamlets of individually adjustable intensity. However, IMPT affords even more flexibility by further breaking up a traditional photon therapy "beamlet" into many depths at which a Bragg peak can occur, known as "beam spots." In this chapter, we will use the terms "beam spot" and "beamlet" interchangeably. From the optimization point of view, both terms simply refer to the variable of the underlying optimization problem. In particular, we will optimize the intensity or weight of each of these beamlets/beam spots, located throughout the region of interest, so as to produce the desired dose distributions.

Lomax [30] describes different kinds of proton therapy treatments, which are all a form of IMPT. Mathematically, these methods differ only in the presence/absence of constraints that relate the intensity of neighboring beam spots. For example, the "2D IMPT" approach requires that all delivered beam spots emanating from a particular lateral position of the beam (a traditional beamlet in photon therapy) have the same intensity, and that the number of delivered beam spots in each lateral position is the same. Our approach is known as the full "3D IMPT" approach in [30]. This is the most general approach since there are no constraints placed on the intensity of neighboring beam spots. The first clinical paper on the use of 3D IMPT was published in 2001 [31].

4.2 Range uncertainty

Uncertainty in the exact range of the proton can arise from many sources. Changes in the patient geometry constitute a fairly broad class of uncertainties. For example, changes in the location of skin folds will change the amount of tissue a proton must pass through to reach the target. Similarly, the thickness of fat layers under a patient's skin can change if a patient loses weight over the course of treatment. Other types of uncertainty involve the physics of the delivery and the associated calculations. For example, the range of a proton is optimized with a particular anatomical picture in hand, but imaging artifacts in the CT scan may give a false impression of the composition of matter between the source and target. Also, different materials have different "stopping powers" (how much a proton is slowed by traveling through it), but calculations based on CT scans may be imprecise since these different materials may have the same electron density (appearance on the CT scan). Finally, uncertainties in the delivery system may produce implementation errors in the form of protons of different energy than expected. The use of positron emission tomography is being explored as a way to measure range uncertainties empirically [41, 42].

In this chapter, we do not model the individual sources of uncertainty, but instead, we model the result of these uncertainties on the realized beamlet ranges. That is, we

do not address the underlying cause of the range variations directly, but we present an abstract approach to mitigate their final effect on the range of the protons.

4.3 Model of Uncertainty

In this section, we describe our model of range uncertainty. Let $\mathcal{B}$ denote the set of beams, and let $b(i)$ denote the set of beamlets associated with each beam $i \in \mathcal{B}$. Unlike in Chapter 2, we provide an explicit index for each beam direction because this gives us the flexibility to model the situation where the range of all beamlets associated with one beam are related. Let $n = \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} 1$ (total number of beamlets over all beams).

For some beam $i \in \mathcal{B}$ and any beamlet $j \in b(i)$, we denote the (uncertain) range of this particular beamlet as $\tilde{\rho}_j^i$. Let ρ_j^i be the nominal range of this beamlet, and let $\overline{\rho}_j^i$ ($\underline{\rho}_j^i$) denote the maximum amount above (below) the nominal range that this beamlet may achieve. In other words, $\tilde{\rho}_j^i \in [\rho_j^i - \underline{\rho}_j^i, \rho_j^i + \overline{\rho}_j^i]$ for any beamlet j from beam i.

The relationship between dose delivered and range of the proton is a function that is measured empirically, and does not have a closed form expression. As a result, we model the range of a proton as taking on one of a finite number of values, finely sampled over the allowable range interval, and compute the depth-dose curve for each of those ranges. We let R_j^i denote the index set of all possible ranges for beamlet $j \in b(i)$, for each beam $i \in \mathcal{B}$. There is a one-to-one correspondence between the elements of R_j^i and the set of realizable ranges for beamlet j from beam i, so any range $\tilde{\rho}_j^i$ is uniquely identified with some index $\tilde{r}_j^i \in R_j^i$. With a slight abuse of notation, we will let $\tilde{\mathbf{r}}$ refer to the "vector" of ranges with elements $\tilde{r}_j^i$, for each beamlet $j \in b(i)$. We denote r_j^i (no tilde) to be the range index associated with the nominal range, ρ_j^i, of beamlet j from beam i.

We define the uncertainty set U to represent the set of all possible range combinations that the beamlets can realize. Therefore,

$$U = \{\tilde{\mathbf{r}} \in \mathbb{R}^n \ : \ \tilde{r}^i_j \in R^i_j, \ \forall\, j \in b(i), \forall\, i \in \mathcal{B}\}. \tag{4.1}$$

While we develop the subsequent mathematical formulations in full generality, we will make the following simplifying assumption when implementing our solutions on a computational case.

Assumption 4.1. *For each beam $i \in \mathcal{B}$ and any two beamlets $j_1, j_2 \in b(i)$, $R^i_{j_1} = R^i_{j_2}$ and $\tilde{r}^i_{j_1} = \tilde{r}^i_{j_2}$.*

This assumption basically says that all beamlets from the same beam have the same number of possible ranges, and will be shifted in the same relative way. It is important to note that this assumption does not force different beamlets from the same beam to take on the same range, only the same range index. For example, imagine that there are three range indices that are ordered in terms of the resulting range realized: "short", "nominal" and "long." Beamlets from the same beam might all realize the "short" range index (e.g., some range index r'), but for beamlet j_1 that might correspond to 8 cm (e.g., ρ'_{j_1}) while for beamlet j_2 from the same beam that might correspond to 9 cm (e.g., ρ'_{j_2}). If this assumption is viewed as too restrictive, one can imagine partitioning a beam into multiple "sub-beams," each composed of beamlets that are shifted in the same way. Of course, the most general case can be recovered by treating each beamlet as an individual beam.

4.4 Problem setup

The geometry of the problem we study is representative of a tumor wrapped around the spinal cord, which is the major organ-at-risk. This is commonly known as the Radiation Therapy Oncology Group (RTOG) benchmark phantom geometry, which is a standard geometry used to test the outcome of different treatment strategies. Note that the main indicator of the goodness of treatment to the spinal cord is the maximum dose delivered there (the spinal cord is a so-called "serial organ" [38, 60, 61]).

The phantom can be thought of as a two-dimensional, circular slice of a three-dimensional, cylindrical patient. A circular organ-at-risk (OAR) of radius 1.5 cm is centered in the phantom, which has a radius of 10 cm. The tumor is shaped like a "horseshoe" and wraps around half of the OAR with an inner radius of 1.8 cm and an outer radius of 4 cm. Everything outside the tumor and OAR is considered to be healthy, unclassified tissue (UT). Voxels are 1 mm $\times$ 1 mm, resulting in a total of 2186 tumor voxels, 441 OAR voxels, and 28,790 UT voxels. The entire phantom is assumed to be homogeneous and water equivalent. Figure 4-2 illustrates the phantom geometry.

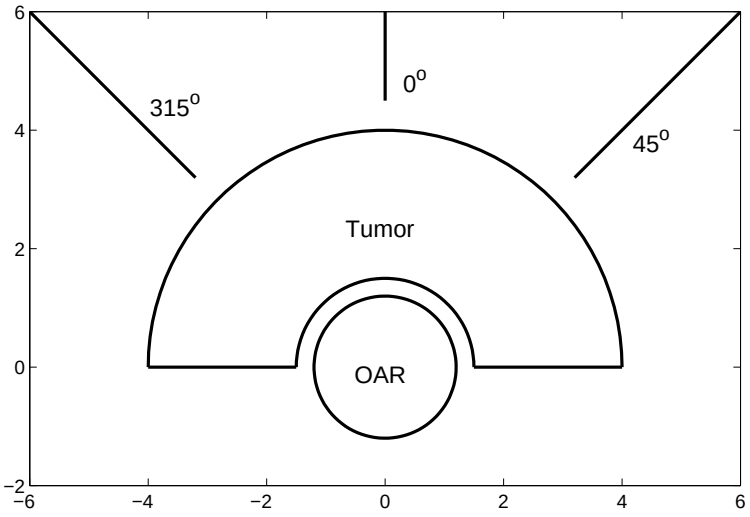


Figure 4-2: The RTOG benchmark phantom and the beam directions used.

Radiation is delivered from three directions (cf. Figure 4-2), replicating the setup presented in [30, 31]. These beam directions are identified by their angle relative to the y-axis of the figure (315°, 0°, and 45° degrees, moving from left to right in the figure). For each beam, beam spots are placed on a rectangular grid, which has grid lines that are oriented parallel and perpendicular to the direction of the incident beam. These spots are separated by 4 mm in both the axial (parallel) and lateral (perpendicular) directions, and are placed throughout the tumor, including a 1 cm isotropic margin around the tumor. Between the three beams, there are 877 beam spots in total.

To calculate the dose deposited from a given beamlet to a given voxel, we use a simple pencil beam algorithm [53, 52], which assumes a lateral dose falloff that is Gaussian with a 5 mm standard deviation at the surface of the patient. The standard deviation varies only slightly as a function of depth, reaching a maximum value of approximately 5.6 mm at a depth just before the location of Bragg peak. It is important to note that the lateral dose falloff is slower than the distal falloff (at the trailing edge of the Bragg peak along the direction that the proton travels, cf. Figure 4-1). This fact will be useful when analyzing the results in Section 4.6.

4.5 Formulations

In this section, we formulate the problem of IMPT optimization under range uncertainty. Let $\mathcal{T}, \mathcal{O}$, and $\mathcal{N}$ denote the set of tumor, organ-at-risk, and normal unclassified tissue voxels, respectively. Let w_j be the intensity of beamlet j and θ_v be the desired dose to voxel v. Also, let $D_{v,j}(\tilde{\rho}_j^i)$ be the dose delivered from beamlet $j \in b(i)$ to voxel v when range $\tilde{\rho}_j^i$ is realized. The dose delivered to a particular voxel v, over all beamlets and beams is written as

$$d_v(\mathbf{w}, \tilde{\boldsymbol{\rho}}) = \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} D_{v,j}(\tilde{\rho}_j^i) w_j.$$

Equivalently, we can write this as

$$d_v(\mathbf{w}, \tilde{\mathbf{r}}) = \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j,$$

where $\Delta_{v,j,\tilde{r}_j^i} := D_{v,j}(\tilde{\rho}_j^i)$. This brings the notation closer to that of Chapter 2. Finally, we will define $\alpha_I, I \in \{\mathcal{T}, \mathcal{O}, \mathcal{N}\}$, to be the "importance factor" or weight of the various organs with respect to one another in the objective function.

Our overall objective is to generate a dose distribution that is as conformal as possible to the tumor, while discouraging dose from being delivered to the spinal cord. In the presence of uncertainty, the trade-off will be between ensuring tumor

coverage while minimizing the dose to the healthy tissue (the OAR in particular). The formulations we present in this section aim to replicate a clinical formulation for the phantom geometry presented. As a result, we will focus on unconstrained formulations which penalize dose distributions that deviate from the ideal dose distribution.

4.5.1 Nominal Formulation

If we assume that there is no uncertainty in the range of each beamlet, then we formulate the *nominal problem* (using the nominal range, $\mathbf{r}$, of all beamlets) as follows:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{I \in \{\mathcal{T},\mathcal{O},\mathcal{N}\}} \alpha_I \sum_{v \in I} |d_v(\mathbf{w}, \mathbf{r}) - \theta_v| \\ \text{subject to} \quad & w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}. \end{aligned} \tag{4.2}$$

This formulation penalizes deviations of the delivered dose from the desired dose (represented by θ_v) for each organ separately, and then combines these individual objectives into one overall objective function, weighted by the importance factor, α_I for each organ I. By adjusting the weights α_I, one can produce dose distributions that favor the tumor, OAR, or UT in terms of their individual objectives. It is important to point out that the individual objective functions for different organs are difficult to compare in a meaningful manner, and their combination into a single function is subjective. Hence, this is only one possibility for a "nominal formulation."

Since the desired dose to the OAR and UT voxels is zero, we can specialize formulation (4.2) to

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v \in \mathcal{T}} |d_v(\mathbf{w}, \mathbf{r}) - \theta_v| + \alpha_{\mathcal{O}} \sum_{v \in \mathcal{O}} d_v(\mathbf{w}, \mathbf{r}) + \alpha_{\mathcal{N}} \sum_{v \in \mathcal{N}} d_v(\mathbf{w}, \mathbf{r}) \\ \text{subject to} \quad & w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \end{aligned} \tag{4.3}$$

which can be written as the equivalent linear program

$$
\begin{aligned}
\underset{\mathbf{w},\mathbf{z}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v \in \mathcal{T}} z_v + \alpha_{\mathcal{O}} \sum_{v \in \mathcal{O}} \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r_j^i} w_j + \alpha_{\mathcal{N}} \sum_{v \in \mathcal{N}} \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r_j^i} w_j \\
\text{subject to} \quad & z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r_j^i} w_j - \theta_v, \quad \forall\, v \in \mathcal{T}, \\
& z_v \geq \theta_v - \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r_j^i} w_j, \quad \forall\, v \in \mathcal{T}, \\
& z_v \text{ free}, \quad \forall\, v \in \mathcal{T}, \\
& w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned}
\tag{4.4}
$$

If there is no uncertainty in the range of the beamlets, then (4.4) is the appropriate formulation to solve. However, implementing the optimized intensity map, $\mathbf{w}^*$, of formulation (4.4) with ranges that differ from the nominal range of each beamlet may lead to a dose distribution with undesirable characteristics (e.g., hot spots in the OAR, cold spots in the tumor, etc.). Next we introduce range uncertainty in the range of the beamlets, and show how to incorporate this uncertainty into a tractable, robust formulation.

4.5.2 Robust formulation

By introducing uncertainty into the range of each beamlet, the objective and constraints must now depend on the allowable ranges in the uncertainty set. We propose the following robust formulation, which is a straightforward extension of the nominal formulation (4.4).

$$
\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v \in \mathcal{T}} \max_{\tilde{\mathbf{r}} \in U} |d_v(\mathbf{w}, \tilde{\mathbf{r}}) - \theta_v| + \alpha_{\mathcal{O}} \sum_{v \in \mathcal{O}} \max_{\tilde{\mathbf{r}} \in U} d_v(\mathbf{w}, \tilde{\mathbf{r}}) + \alpha_{\mathcal{N}} \sum_{v \in \mathcal{N}} d_v(\mathbf{w}, \mathbf{r}) \\
\text{subject to} \quad & w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned}
\tag{4.5}
$$

We believe that the objective function in (4.5) is a meaningful measure of the quality of a treatment plan under uncertainty. We show later that formulation (4.5) effectively incorporates the key aspects of tumor coverage and healthy tissue sparing, while

maintaining tractability. Note that we do not robustify the part of the objective function corresponding to the unclassified tissue. We are less concerned about the unclassified tissue on a voxel by voxel basis, so it is sufficient to measure the dose delivered there on an aggregate level using the nominal range $\mathbf{r}$. This formulation can be rewritten as the following equivalent, linear formulation.

$$
\begin{aligned}
\underset{\mathbf{w},\mathbf{z}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v \in \mathcal{T}} z_v + \alpha_{\mathcal{O}} \sum_{v \in \mathcal{O}} z_v + \alpha_{\mathcal{N}} \sum_{v \in \mathcal{N}} \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r_j^i} w_j \\
\text{subject to} \quad & z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j - \theta_v, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} R_j^i, \\
& z_v \geq \theta_v - \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} R_j^i, \\
& z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j, \quad \forall\, v \in \mathcal{O},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} R_j^i, \\
& z_v \text{ free}, \quad \forall\, v \in \mathcal{T} \cup \mathcal{O}, \\
& w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned}
\tag{4.6}
$$

With a discrete set of range indices, formulation (4.6) is a standard linear program. However, it has a very large number of constraints per tumor voxel and OAR voxel (on the order of the number of range indices to the power of the total number of beamlets). Therefore, we present an alternative, yet equivalent, formulation that is more tractable, and is the one that is used in the results section. This equivalent formulation is constructed in the spirit of the robust formulation from Section 2.4.2.

For a given beamlet j from beam i, its range $\tilde{r}_j^i$ can be any element of R_j^i. Suppose, now, that instead of picking one element from each R_j^i, we pick a probability distribution over the elements of R_j^i. That is, instead of a beamlet realizing one particular range, it somehow manages to realize a distribution of ranges. Of course, this is not physically possible, but it turns out that we can use this probabilistic interpretation to simplify our problem. For each R_j^i, we associate with it a bounded polyhedron whose extreme points are in 1-1 correspondence with the elements of R_j^i. In particular, the unit simplex in $\mathbb{R}^{|R_j^i|}$ is such a polyhedron. Finally, the unit simplex

is equivalent to the set of all discrete probability distributions with support of the same size as the underlying dimension of the simplex itself.

Let $\tilde{p}_j^i(x)$ be a probability mass function (pmf) defined over the elements $x \in R_j^i$, P_j^i be the unit simplex in $\mathbb{R}^{|R_j^i|}$, and $\tilde{\mathbf{p}}$ be a "vector" of pmfs with length $\sum_{i\in\mathcal{B}}\sum_{j\in b(i)}|R_j^i|$. First, we present an equivalent representation of any constraint from formulation (4.5) using $\tilde{p}_j^i$.

Lemma 4.1. *For a given v, any z_v that satisfies*

$$z_v + \theta_v \geq \sum_{i\in\mathcal{B}}\sum_{j\in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j, \quad \forall\, \tilde{\mathbf{r}} \in \prod_{i\in\mathcal{B}}\prod_{j\in b(i)} R_j^i, \tag{4.7}$$

also satisfies

$$z_v + \theta_v \geq \sum_{i\in\mathcal{B}}\sum_{j\in b(i)}\sum_{x\in R_j^i} \Delta_{v,j,x}\tilde{p}_j^i(x) w_j, \quad \forall\, \tilde{\mathbf{p}} \in \prod_{i\in\mathcal{B}}\prod_{j\in b(i)} P_j^i, \tag{4.8}$$

and vice versa.

Proof: If z_v satisfies (4.7), then

$$z_v + \theta_v \geq \sum_{i\in\mathcal{B}}\sum_{j\in b(i)} \left(\max_{x\in R_j^i} \Delta_{v,j,x}\right) w_j,$$

and the largest $\Delta_{v,j,x}$ over $x \in R_j^i$ is clearly larger than any convex combination of the $\Delta_{v,j,x}$ over $x \in R_j^i$. Hence, the forward implication follows. The reverse direction is obvious. □

Using Lemma 4.1, we can write the following equivalent formulation of Formulation (4.6).

$$
\begin{aligned}
\underset{\mathbf{w},\mathbf{z}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v \in \mathcal{T}} z_v + \alpha_{\mathcal{O}} \sum_{v \in \mathcal{O}} z_v + \alpha_{\mathcal{N}} \sum_{v \in \mathcal{N}} \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r_j^i} w_j \\
\text{subject to} \quad & z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \sum_{x \in R_j^i} \Delta_{v,j,x} \tilde{p}_j^i(x) w_j - \theta_v, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{p}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} P_j^i, \\
& z_v \geq \theta_v - \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \sum_{x \in R_j^i} \Delta_{v,j,x} \tilde{p}_j^i(x) w_j, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{p}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} P_j^i, \\
& z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \sum_{x \in R_j^i} \Delta_{v,j,x} \tilde{p}_j^i(x) w_j, \quad \forall\, v \in \mathcal{O},\ \forall\, \tilde{\mathbf{p}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} P_j^i, \\
& z_v \text{ free}, \quad \forall\, v \in \mathcal{T} \cup \mathcal{O}, \\
& w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned}
\tag{4.9}
$$

In formulation (4.6), all extreme points of the overall "range polyhedron" (the polyhedron that considers the ranges of all beamlets together) are included, which can be a large number. Instead, by optimizing formulation (4.9), the optimization process will pick out the "binding" extreme probability distributions (point masses) automatically. Even though formulation (4.9) seems to be more constrained than formulation (4.6), the linearity of the formulations ensures that only the extreme points of the probability distribution polyhedron (which are in 1-1 correspondence with the set of realizable ranges for each beamlet) will be realized at an optimal solution.

Examining formulation (4.9), it is clear that we no longer have a linear program. However, in a manner very similar to the derivation of the linear robust formulation from Chapter 2, we can use duality to generate an equivalent linear formulation in this situation.

Proposition 4.1. *For a given v and y, we have*

$$
y \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \sum_{x \in R_j^i} \Delta_{v,j,x} \tilde{p}_j^i(x) w_j, \quad \forall\, \tilde{\mathbf{p}} \in \prod_{i \in \mathcal{B}} \prod_{j \in b(i)} P_j^i, \tag{4.10}
$$

if and only if there exist $q_{v,i,j}$ *and* $s_{v,x,i,j}$ *such that*

$$\begin{aligned}
& y \geq \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} q_{v,i,j} + \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \sum_{x\in R_j^i} s_{v,x,i,j}, \\
& q_{v,i,j} + s_{v,x,i,j} \geq \Delta_{v,j,x} w_j, \quad \forall\, x \in R_j^i,\ \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \\
& q_{v,i,j} \text{ free}, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \\
& s_{v,x,i,j} \geq 0, \quad \forall\, x \in R_j^i,\ \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned} \tag{4.11}$$

Proof: Equation (4.10) is equivalent to the inequality $y \geq Z$, where Z is the optimal value of the following problem

$$\begin{aligned}
\underset{\tilde{\mathbf{p}}}{\text{maximize}} \quad & \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \sum_{x\in R_j^i} \Delta_{v,j,x} \tilde{p}_j^i(x) w_j \\
\text{subject to} \quad & \sum_{x\in R_j^i} \tilde{p}_j^i(x) = 1, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \\
& \tilde{p}_j^i(x) \leq 1, \quad \forall\, x \in R_j^i,\ \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \\
& \tilde{p}_j^i(x) \geq 0, \quad \forall\, x \in R_j^i,\ \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned} \tag{4.12}$$

The dual of formulation (4.12) is

$$\begin{aligned}
\underset{\mathbf{q}_v, \mathbf{s}_v}{\text{minimize}} \quad & \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} q_{v,i,j} + \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \sum_{x\in R_j^i} s_{v,x,i,j} \\
\text{subject to} \quad & q_{v,i,j} + s_{v,x,i,j} \geq \Delta_{v,j,x} w_j, \quad \forall\, x \in R_j^i,\ \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \\
& q_{v,i,j} \text{ free}, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}, \\
& s_{v,x,i,j} \geq 0, \quad \forall\, x \in R_j^i,\ \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned} \tag{4.13}$$

Since formulation (4.12) is feasible, the result follows from strong duality of linear programming. □

Using this result, suitably applied to all constraints in formulation (4.9), we arrive at a linear program that is equivalent to (4.9).

Theorem 4.1. *Formulation* (4.9) *is equivalent to the following linear program:*

$$
\begin{aligned}
\underset{\substack{\mathbf{w},\mathbf{z},\mathbf{q},\mathbf{s}\\ \mathbf{Q},\mathbf{S},\boldsymbol{\psi},\boldsymbol{\sigma}}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v\in\mathcal{T}} z_v + \alpha_{\mathcal{O}} \sum_{v\in\mathcal{O}} z_v + \alpha_{\mathcal{N}} \sum_{v\in\mathcal{N}} \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \Delta_{v,j,r_j^i} w_j \\
\text{subject to} \quad & z_v + \theta_v \geq \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} q_{v,i,j} + \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \sum_{x\in R_j^i} s_{v,x,i,j}, \quad \forall\, v\in\mathcal{T}, \\
& \theta_v - z_v \leq \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} Q_{v,i,j} - \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \sum_{x\in R_j^i} S_{v,x,i,j}, \quad \forall\, v\in\mathcal{T}, \\
& z_v \geq \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \psi_{v,i,j} + \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \sum_{x\in R_j^i} \sigma_{v,x,i,j}, \quad \forall\, v\in\mathcal{O}, \\
& q_{v,i,j} + s_{v,x,i,j} \geq \Delta_{v,j,x} w_j, \quad \forall\, v\in\mathcal{T},\ \forall\, x\in R_j^i,\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& Q_{v,i,j} - S_{v,x,i,j} \leq \Delta_{v,j,x} w_j, \quad \forall\, v\in\mathcal{T},\ \forall\, x\in R_j^i,\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& \psi_{v,i,j} + \sigma_{v,x,i,j} \geq \Delta_{v,j,x} w_j, \quad \forall\, v\in\mathcal{O},\ \forall\, x\in R_j^i,\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& q_{v,i,j}, Q_{v,i,j} \ \text{ free}, \quad \forall\, v\in\mathcal{T},\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& \psi_{v,i,j} \ \text{ free}, \quad \forall\, v\in\mathcal{O},\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& s_{v,x,i,j}, S_{v,x,i,j} \geq 0, \quad \forall\, v\in\mathcal{T},\ \forall\, x\in R_j^i,\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& \sigma_{v,x,i,j} \geq 0, \quad \forall\, v\in\mathcal{O},\ \forall\, x\in R_j^i,\ \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}, \\
& z_v \text{ free}, \quad \forall\, v\in\mathcal{T}\cup\mathcal{O}, \\
& w_j \geq 0, \quad \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}.
\end{aligned}
\tag{4.14}
$$

Formulation (4.14) is a more tractable version of formulation (4.6). The following tables show the number of constraints and variables of these formulations. Let $m_{\mathcal{T}} = |\mathcal{T}|$ and $m_{\mathcal{O}} = |\mathcal{O}|$. Recall that n denotes the total number of beamlets over all beams.

Table 4.1: Formulation (4.6).

Constraints	$2m_{\mathcal{T}} \prod_{i\in\mathcal{B}} \prod_{j\in b(i)} \lvert R_j^i\rvert + m_{\mathcal{O}} \prod_{i\in\mathcal{B}} \prod_{j\in b(i)} \lvert R_j^i\rvert$
Variables	$m_{\mathcal{T}} + m_{\mathcal{O}} + n$

We see that while the number of variables has increased by a polynomial (in the

Table 4.2: Formulation (4.14).

Constraints	$2m_{\mathcal{T}}(\sum_{i\in\mathcal{B}}\sum_{j\in b(i)}\|R_j^i\|+1)+m_{\mathcal{O}}(\sum_{i\in\mathcal{B}}\sum_{j\in b(i)}\|R_j^i\|+1)$
Variables	$m_{\mathcal{T}}(2n+2\sum_{i\in\mathcal{B}}\sum_{j\in b(i)}\|R_j^i\|+1)+m_{\mathcal{O}}(n+\sum_{i\in\mathcal{B}}\sum_{j\in b(i)}\|R_j^i\|+1)+n$

input size of the problem) amount, from formulation (4.6) to formulation (4.14), the number of constraints has decreased from an exponential number to a polynomial number. Therefore, we expect formulation (4.14) to be much more tractable. Before we proceed to the results section, we present a brief discussion on an alternative robust formulation.

An alternative robust formulation

Since there is no generally agreed upon objective function in practice in the nominal case, there will not be one "correct" robust formulation. Therefore, it is worthwhile to discuss an alternative robust formulation. The following formulation may be viewed as a very natural min-max robust formulation.

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \max_{\tilde{\mathbf{r}}\in U}\left(\alpha_{\mathcal{T}}\sum_{v\in\mathcal{T}}|d_v(\mathbf{w},\tilde{\mathbf{r}})-\theta_v|+\alpha_{\mathcal{O}}\sum_{v\in\mathcal{O}}d_v(\mathbf{w},\tilde{\mathbf{r}})\right)+\alpha_{\mathcal{N}}\sum_{v\in\mathcal{N}}d_v(\mathbf{w},\mathbf{r}) \\ \text{subject to} \quad & w_j\geq 0, \quad \forall\, j\in b(i),\ \forall\, i\in\mathcal{B}. \end{aligned} \tag{4.15}$$

We may write formulation (4.15) as the following linear program (assuming R_j^i is finite as before).

$$
\begin{aligned}
\underset{\mathbf{w},y,\mathbf{z}}{\text{minimize}} \quad & y + \alpha_{\mathcal{N}} \sum_{v\in\mathcal{N}} \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \Delta_{v,j,r_j^i} w_j \\
\text{subject to} \quad & y \geq \alpha_{\mathcal{T}} \sum_{v\in\mathcal{T}} z_{v,\tilde{\mathbf{r}}} + \alpha_{\mathcal{O}} \sum_{v\in\mathcal{T}} \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i\in\mathcal{B}} \prod_{j\in b(i)} R_j^i, \\
& z_{v,\tilde{\mathbf{r}}} \geq \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j - \theta_v, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i\in\mathcal{B}} \prod_{j\in b(i)} R_j^i, \\
& z_{v,\tilde{\mathbf{r}}} \geq \theta_v - \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \Delta_{v,j,\tilde{r}_j^i} w_j, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i\in\mathcal{B}} \prod_{j\in b(i)} R_j^i, \\
& y \text{ free}, \\
& z_{v,\tilde{\mathbf{r}}} \text{ free}, \quad \forall\, v \in \mathcal{T},\ \forall\, \tilde{\mathbf{r}} \in \prod_{i\in\mathcal{B}} \prod_{j\in b(i)} R_j^i, \\
& w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned}
\tag{4.16}
$$

Comparing this formulation with formulation (4.14), the main difference to note is the addition of the index $\tilde{\mathbf{r}}$ to the variable z_v, for all $v \in \mathcal{T}$. Since the variables $z_{v,\tilde{\mathbf{r}}}$ are not decoupled (in terms of the index $\tilde{\mathbf{r}}$) from the right hand side of the corresponding constraints, we can no longer view those constraints as implicit optimization problems as we did for formulation (4.9). Hence, we can no longer use the duality argument from Proposition 4.1 to reduce the number of constraints of formulation (4.16) to a polynomial number.

By comparing the objective functions, it should be clear that formulation (4.14) is more conservative than formulation (4.16) (compare (4.5), which is equivalent to (4.14), with (4.15), which is equivalent to (4.16)). However, because the overall objective functions are a combination of individual objectives for different organs, it is not necessarily true that a higher objective value translates into a worse solution from a clinical perspective. Thus, in the absence of a "gold standard" objective, our goal is to use an objective function that incorporates aspects of dose minimization and tumor coverage, and is tractable. We believe that the objective function in (4.14) achieves this goal. However, a computational study of formulation (4.16) using constraint generation, for example, would be an interesting future exercise.

4.6 Results from a phantom study

In this section, we compare the performance of the robust solution from formulation (4.14) to the nominal solution from formulation (4.4) in the face of range uncertainty.

For the computational results we present in this section, we assume that different beamlets from the same beam will have the same number of possible range indices, and will shift together (recall Assumption 4.1). This effectively means that the range uncertainty can be associated with the beams, as opposed to the individual beamlets. Furthermore, for simplicity, we assume that the set of possible ranges is the same for all three beams. Notation-wise, we simply use R to denote the set of possible ranges that any beamlet from any beam may realize, as opposed to a specific set R_j^i for each beamlet j from beam i. However, we do allow independence of range shifts between the different beams. We let $\tilde{\mathbf{r}}^i$ refer to the vector of ranges associated with only the beamlets of beam i.

More concretely, we focus on range shifts that are between -5 mm and $+5$ mm of the nominal range of each beamlet, in 1 mm increments. This results in a total of 11 possible ranges for each beamlet. The nominal range corresponds to a shift of 0 mm. We set $\theta_v = 1$ for all $v \in \mathcal{T}$ to denote a relative level of desired dose to the tumor. We set $\alpha_{\mathcal{T}} = 1, \alpha_{\mathcal{O}} = 0.5$, and $\alpha_{\mathcal{N}} = 0.01$. It is worth noting that, in practice, there is no universally agreed upon weighting scheme for different objectives combined into one objective function. Some studies prefer using a weighting scheme that also takes into account the number of voxels in each organ, so that an organ with a lot of voxels does not dominate a more important organ with many fewer voxels. In those cases, the α importance factors would be divided by the number of voxels in each organ, thereby weighing the importance of the "average" tumor voxel relative to an OAR voxel and an UT voxel. In our case, we did not see evidence that one organ was dominating the others in the objective based purely on the number of voxels in that organ. Plus, since the importance factors are somewhat arbitrary, we preferred a weighting scheme that was organ-based, as opposed to voxel-based. Since we want

to obtain tumor coverage, while sparing the OAR as much as possible, we choose our importance factors to emphasize that these are our main objectives. In physical terms, we are most concerned with getting a steep dose gradient around the tumor, and especially between the tumor and the OAR.

For completeness, we show the robust formulation that results after incorporating Assumption 4.1, as well as the one that is solved in the optimization. The uncertainty set corresponding to beam i can now be written as:

$$U_i = \{\tilde{\mathbf{r}}^i \in \mathbb{R}^{|b(i)|} \ : \ \tilde{r}^i_j \in R, \ \forall j \in b(i); \ \tilde{r}^i_j = \tilde{r}^i_k, \ \forall j,k \in b(i)\}. \tag{4.17}$$

Alternatively, we can simplify this expression by considering only range uncertainty in the beams, as opposed to beamlets, to arrive at the following uncertainty set (where $\tilde{\mathbf{r}}$ is now a vector of dimension $|\mathcal{B}|$):

$$U_{\mathcal{B}} = \{\tilde{\mathbf{r}} \in \mathbb{R}^{|\mathcal{B}|} \ : \ \tilde{r}^i \in R, \ \forall i \in \mathcal{B}\}. \tag{4.18}$$

Applying this model of uncertainty (4.18) to formulation (4.5), and writing the equivalent linear program, we arrive at

$$\begin{aligned}
\underset{\mathbf{w},\mathbf{z}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v \in \mathcal{T}} z_v + \alpha_{\mathcal{O}} \sum_{v \in \mathcal{O}} z_v + \alpha_{\mathcal{N}} \sum_{v \in \mathcal{N}} \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,r^i_j} w_j \\
\text{subject to} \quad & z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}^i} w_j - \theta_v, \quad \forall\, v \in \mathcal{T}, \ \forall\, \tilde{r}^i \in R, \ \forall\, i \in \mathcal{B}, \\
& z_v \geq \theta_v - \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}^i} w_j, \quad \forall\, v \in \mathcal{T}, \ \forall\, \tilde{r}^i \in R, \ \forall\, i \in \mathcal{B}, \\
& z_v \geq \sum_{i \in \mathcal{B}} \sum_{j \in b(i)} \Delta_{v,j,\tilde{r}^i} w_j, \quad \forall\, v \in \mathcal{O}, \ \forall\, \tilde{r}^i \in R, \ \forall\, i \in \mathcal{B}, \\
& z_v \text{ free}, \quad \forall\, v \in \mathcal{T} \cup \mathcal{O}, \\
& w_j \geq 0, \quad \forall\, j \in b(i), \ \forall\, i \in \mathcal{B}.
\end{aligned} \tag{4.19}$$

Replacing the probability mass functions $p^i_j(x)$ with $p^i(x)$ in formulation (4.9), we derive the final robust counterpart that is solved in this section.

$$
\begin{aligned}
\underset{\substack{\mathbf{w},\mathbf{z},\mathbf{q},\mathbf{s}\\ \mathbf{Q},\mathbf{S},\boldsymbol{\psi},\boldsymbol{\sigma}}}{\text{minimize}} \quad & \alpha_{\mathcal{T}} \sum_{v\in\mathcal{T}} z_v + \alpha_{\mathcal{O}} \sum_{v\in\mathcal{O}} z_v + \alpha_{\mathcal{N}} \sum_{v\in\mathcal{N}} \sum_{i\in\mathcal{B}} \sum_{j\in b(i)} \Delta_{v,j,r^i} w_j \\
\text{subject to} \quad & z_v + \theta_v \geq \sum_{i\in\mathcal{B}} q_{v,i} + \sum_{i\in\mathcal{B}} \sum_{x\in R} s_{v,x,i}, \quad \forall\, v \in \mathcal{T}, \\
& \theta_v - z_v \leq \sum_{i\in\mathcal{B}} Q_{v,i} - \sum_{i\in\mathcal{B}} \sum_{x\in R} S_{v,x,i}, \quad \forall\, v \in \mathcal{T}, \\
& z_v \geq \sum_{i\in\mathcal{B}} \psi_{v,i} + \sum_{i\in\mathcal{B}} \sum_{x\in R} \sigma_{v,x,i}, \quad \forall\, v \in \mathcal{O}, \\
& q_{v,i} + s_{v,x,i} \geq \sum_{j\in b(i)} \Delta_{v,j,x} w_j, \quad \forall\, v \in \mathcal{T},\ \forall\, x \in R,\ \forall\, i \in \mathcal{B}, \\
& Q_{v,i} - S_{v,x,i} \leq \sum_{j\in b(i)} \Delta_{v,j,x} w_j, \quad \forall\, v \in \mathcal{T},\ \forall\, x \in R,\ \forall\, i \in \mathcal{B}, \\
& \psi_{v,i} + \sigma_{v,x,i} \geq \sum_{j\in b(i)} \Delta_{v,j,x} w_j, \quad \forall\, v \in \mathcal{O},\ \forall\, x \in R,\ \forall\, i \in \mathcal{B}, \\
& q_{v,i}, Q_{v,i} \ \text{ free}, \quad \forall\, v \in \mathcal{T},\ \forall\, i \in \mathcal{B}, \\
& \psi_{v,i} \ \text{ free}, \quad \forall\, v \in \mathcal{O},\ \forall\, i \in \mathcal{B}, \\
& s_{v,x,i}, S_{v,x,i} \geq 0, \quad \forall\, v \in \mathcal{T},\ \forall\, x \in R,\ \forall\, i \in \mathcal{B}, \\
& \sigma_{v,x,i} \geq 0, \quad \forall\, v \in \mathcal{O},\ \forall\, x \in R,\ \forall\, i \in \mathcal{B}, \\
& z_v \text{ free}, \quad \forall\, v \in \mathcal{T} \cup \mathcal{O}, \\
& w_j \geq 0, \quad \forall\, j \in b(i),\ \forall\, i \in \mathcal{B}.
\end{aligned}
\tag{4.20}
$$

Note that the derivation of formulation (4.20) is slightly different from that of formulation (4.14), since we eliminated the dependence of the range uncertainty on the individual beamlets j. Accordingly, the dual variables no longer have a subscript j, and a summation over $j \in b(i)$ is deleted from or added to the constraints where appropriate.

All formulations were optimized using the barrier algorithm of CPLEX 9 on a 3 GHz computer with 3 GB of memory. Before we proceed with a comparison between the nominal and robust solutions, we briefly discuss the improved tractability of robust formulation (4.20) over the equivalent instance of robust formulation (4.6).

The appropriate instantiation of formulation (4.6) under the current setup (11 ranges, beamlets of the same beam are coupled, voxels are 1 mm^2, etc.) could not be solved due to insufficient memory. In order to solve this formulation, we had to reduce the problem size by reducing $|R|$ and decreasing the voxel resolution. With the voxels at 4 mm^2 (a $\sim$16 times reduction in the number of voxels) and with only 5 range indices in R (a $\sim$2 times reduction), formulation (4.6) solved in roughly 3 minutes. For comparison, using this setup, formulation (4.20) solved in about 10 seconds (a $\sim$20 times speedup). For the desired setup, solving formulation (4.6) is not practical. It is possible that delayed constraint generation would improve the tractability of formulation (4.6), but this was not explored since we were able to solve formulation (4.20) at the desired granularity using the standard barrier algorithm. A more thorough comparison using constraint generation would be an interesting topic of future research.

Returning to the comparison between the nominal formulation (4.4) and the robust formulation (4.20), Table 4.3 shows the number of variables and constraints in each formulation, along with the associated solution times.

Table 4.3: Basic information regarding the nominal and robust formulations.

	Nominal	Robust
Number of variables	3063	176,772
Number of constraints	4813	163,642
Solution time (in min)	0.5	9

First, we consider the nominal solution and examine what can go wrong if we do not protect against range uncertainty. The dosimetric consequences of using the nominal solution under various realizations of the beamlet ranges are presented in Figure 4-3. In this figure and in the corresponding one for the robust solution, when we refer to a "longer range" or a "shorter range" being realized by the beamlets, this refers to a $+5$ mm and -5 mm shift, respectively. While we assume that the individual beams are independent in the optimization, the figures that refer to "longer range" or "shorter range" represent outcomes where all beams realize the longer or

shorter range, respectively.

In Figure 4-3(a), when all beamlets realize their nominal ranges, we see that the dose distribution is highly conformal around the tumor, except for a slight hot spot in front of the tumor. The nominal solution chooses to deliver more dose to the unclassified tissue in front of the tumor since this allows improved sparing of the more important OAR behind the tumor. Moreover, the steep distal dose falloff that is characteristic of protons (cf. Figure 4-1) is used at the edge of the tumor to produce the sharpest gradients between the tumor and surrounding healthy tissue. This is evident in Figure 4-3(b) and Figure 4-3(d). In both figures, the most intense Bragg peaks are placed right in front of the OAR to produce the steepest dose gradient between the tumor and OAR. Furthermore, from the 45° direction, Bragg peaks are placed at the interface of the tumor and UT. Note that the beam from 315° is a mirror image of the one from 45°, hence it is not shown. While this placement of Bragg peaks produces a dose distribution that effectively covers the tumor and spares the OAR, it is also very sensitive to uncertainty in the range of the beamlets.

The inadequacy of the nominal solution is most apparent in Figure 4-3(e), where the beamlets overshoot their nominal range by 5 mm. Since the dose gradient between the tumor and the OAR is crafted by Bragg peaks placed right in front of the OAR, beamlets that overshoot will end up placing these intense peaks right *into* the OAR. This results in potentially dangerous amounts of dose being delivered to the critical structure. Furthermore, we see that large parts of the tumor also get underdosed by roughly 20% in this scenario. When the beamlets all undershoot (Figure 4-3(c)), there is no danger of placing Bragg peaks into the OAR. However, we do see large hot spots in the tumor, as well as parts of the tumor that receive virtually no dose (the part right in front of the OAR). These problems stem from the fact that the nominal solution is specialized to use the nominal range information of each beamlet to the fullest extent possible. This results in dose distributions delivered by individual beams that are highly inhomogeneous (in particular, the beams at 45° and 315°), and that need to overlap just right in order to produce the desired dose distribution. As a result, any deviations from the nominal range may lead to hot and cold spots in the anatomy.

One can imagine an analogy to interference in waves. Properly tuned overlapping waves can produce a flat surface, but perturbations in the individual waves may cause interference that amplifies their oscillatory nature, which is analogous to the resulting hot and cold spots witnessed in Figure 4-3(c) and Figure 4-3(e).

Finally, Figure 4-3(f) offers another look at the vulnerability of the nominal solution to range uncertainty. This figure illustrates the maximum deviation, over the 11 possible ranges for each beamlet, of the dose delivered from the prescription dose by the nominal solution. We see that the nominal solution is most susceptible to uncertainty at the interface between the tumor and the OAR. This is a result of the nominal solution placing many intense Bragg peaks in this region. Shifts in the range of the beamlets will move these peaks between the OAR and the tumor. This figure succinctly shows that the nominal solution is very sensitive to uncertainty in exactly the locations where we would like more robustness.

Next, we look at the results of using the robust solution in the face of uncertainty. Figure 4-4 shows the relevant dose distributions.

In Figure 4-4(a), we see that when all beamlets realize their nominal range, the robust solution delivers a highly conformal dose distribution. On the side of the tumor furthest from the entering dose, the robust dose distribution appears to be a little less conformal than the nominal one. However, this is partially balanced by the missing hot spot in front of the tumor that the nominal solution exhibited. Overall, it appears that the robust solution is roughly comparable to the nominal solution in terms of conformity. It is in the remaining figures that the robust solution distinguishes itself from the nominal solution.

In Figure 4-4(c) and Figure 4-4(e), we see that the dose distributions corresponding to undershooting or overshooting beamlets, respectively, still manage to provide good tumor coverage while sparing the OAR. The reason that the robust solution accomplishes these tasks becomes apparent in Figure 4-4(b) and Figure 4-4(d). Unlike the nominal solution, the robust solution completely avoids putting intense Bragg peaks directly in front of the OAR (Figure 4-4(d) illustrates this phenomenon most clearly). Instead, it uses the *lateral* falloff of the proton dose profile to shape the

dose gradient between the tumor and OAR. By using the lateral dose falloff, the robust solution is trading off tumor coverage with OAR sparing in the face of range uncertainty. Recall that the proton dose falloff in the lateral direction is shallower than at the distal edge of the Bragg peak. As a result, the robust solution produces a shallower dose gradient at the interface of the tumor and OAR. This implies that the robust solution will be less conformal than the nominal solution in this area. However, this property of the robust solution is balanced by the increased protection from uncertainty in the range of the beamlets, since the robust solution avoids placing intense Bragg peaks directly into the OAR if the beamlets overshoot and avoids significant underdosing in the tumor if the beamlets undershoot. Range uncertainty affects the distal dose falloff much more than the lateral falloff, which can be considered largely constant as a function of depth. Hence, by using the lateral falloff to shape the dose gradient at the interface between the tumor and OAR, variations in the range of the beamlets will have a minimal effect on the quality of the resulting treatment. Furthermore, note that the dose distributions delivered by the individual beams are widely homogeneous in beam direction – another indication that the robust solution is less sensitive to beamlet range shifts. Going back to the wave analogy, the robust solution can be thought of as overlapping waves with a lower amplitude and longer period, so any perturbations of the individual waves will result in only mild constructive interference.

Finally, Figure 4-4(f) shows that the variability in the robust dose distribution is greatly reduced as compared to the nominal dose distribution. The deviations from the prescription dose in the tumor are very slight, and the deviations in the OAR, while spread over a larger volume of the structure, are lower at the maximum, as compared to the nominal solution. This phenomenon is more easily visualized with a DVH. The DVHs for the nominal and robust solutions under the three range scenarios (all beamlets realize their nominal range, longest range, or shortest range) are shown in Figure 4-5.

Figure 4-5(a) shows that the nominal solution produces a dose distribution in the tumor that is very sensitive to range uncertainty in the beamlets. Both undershooting

and overshooting result in extreme hot and cold spots in the tumor. For example, if the beamlet ranges are all shortened by 5 mm, then about 10% of the tumor is underdosed by roughly 70%, and 10% of the tumor is overdosed by 20%. These are completely unacceptable statistics. The effect of the uncertainty on the dose delivered to the unclassified tissue is negligible, as the DVHs are very similar for the different range scenarios. Finally, we see that the dose to the OAR can be very sensitive to range uncertainty, especially if the beamlets overshoot. If the beamlet ranges are all lengthened by 5 mm, then there is an extreme hot spot in the OAR of 120% of the *tumor* prescription dose.

Figure 4-5(b) shows that the robust dose distribution is better protected from the effects of range uncertainty in the beamlets, as compared to the nominal solution. The tumor DVH, in particular, displays far less sensitivity to range variations of the beamlets. When the beamlets undershoot, the resulting hot and cold spots in the tumor are much less significant than in the nominal solution. These hot spots amount to a 4% overdose to 2% of the tumor, while the cold spots correspond to a 5% underdose to 7% of the tumor – these are much more acceptable statistics. The robust solution also does a better job of protecting the OAR from the negative effects of range uncertainty. Recall that we are considering the OAR to be a serial organ, so we are mainly interested in the maximum dose delivered to it. The maximum hot spot in the OAR is just over 80% of the tumor prescription dose, which is almost a one-third reduction as compared to the nominal solution. Even though the maximum dose to the OAR is still quite high, we should focus on the relative improvement over the nominal solution, since different importance factors in the objective function could reduce both of these hot spot values. Finally, as is the case in the nominal solution, the dose distribution over the unclassified tissue is nearly unchanged in the various range scenarios.

To more directly compare the nominal and robust solutions in these three scenarios (beamlets realize nominal range, beamlets are 5 mm shorter, beamlets are 5 mm longer), Figure 4-6 plots the nominal and robust DVHs on the same axes. We focus only on the tumor and OAR, and omit the less important UT DVHs from these

figures.

We see that the robust solution is generally much better than the nominal solution at protecting the tumor from underdosing or overdosing in the various scenarios. While the robust solution is better than the nominal solution in terms of maximum dose delivered to the OAR under the scenario where beamlet ranges are all lengthened, the price to pay for the robustness is a relative increase in the dose delivered to the OAR for the other scenarios.

4.7 Conclusions

In this chapter, we considered the optimization of intensity-modulated proton therapy treatments under uncertainty in the range of the protons. We formulated the problem of optimizing the treatment under no uncertainty (the nominal formulation), and then incorporated a model of proton range uncertainty to arrive at a robust formulation. Using duality, we constructed an equivalent linear, robust formulation that was tractable, and produced relevant and intuitive solutions.

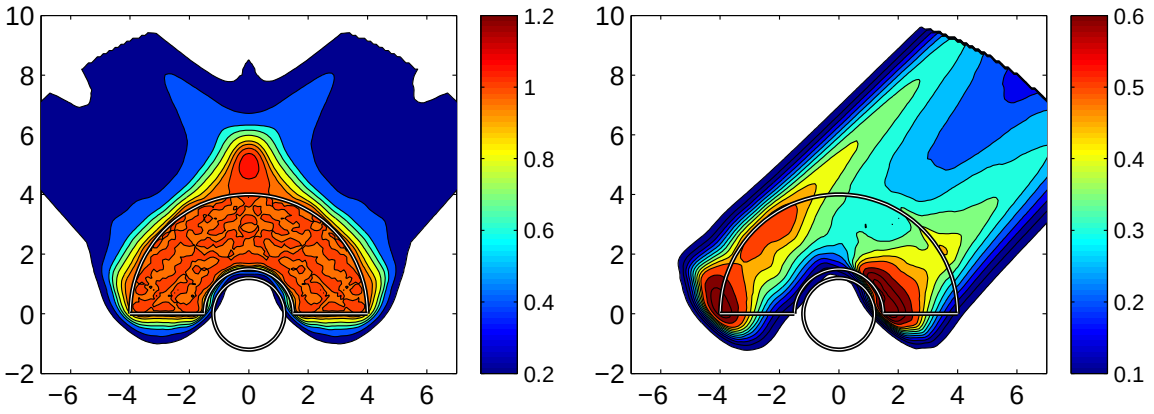
We compared the performance of the nominal treatment plan with the robust treatment plan on the standard RTOG phantom geometry. The nominal solution provided better tumor coverage and OAR sparing over the robust solution, if we assumed that all beamlets realized some pre-defined, nominal range. However, if the beamlets realized a different range than expected, the nominal solution resulted in severe hot spots (20% overdose to 10% of the tumor) and cold spots (70% underdose to 10% of the tumor) in the tumor, and it delivered 120% of the tumor prescription dose in the OAR. The robust solution fared much better in the face of range uncertainty, virtually ensuring the same tumor coverage as if the beamlets had all realized their nominal range, and reducing the worst-case OAR hot spot by 30% as compared to the nominal solution.

The differences between the nominal approach and robust approach can be attributed to the construction of their respective optimized intensity maps. The nominal intensity map placed intense Bragg peaks directly in front of the OAR. While this

configuration creates a sharp dose gradient between the tumor and OAR, the quality of the plan is susceptible to range uncertainty in those beamlets, especially if they overshoot their nominal range. The robust solution utilized an approach that was orthogonal (both figuratively and literally) to the nominal solution. It chose to avoid using the distal falloff of the Bragg peak to shape the dose gradient, and instead, it used the lateral falloff. This type of intensity map results in a shallower dose gradient, but also manages to sidestep the most significant negative effect of the range uncertainty (namely, placing Bragg peaks into the OAR). Since the lateral falloff changes minimally with the range of the proton, the OAR is protected from the overshooting or undershooting of the beamlets closest to it.

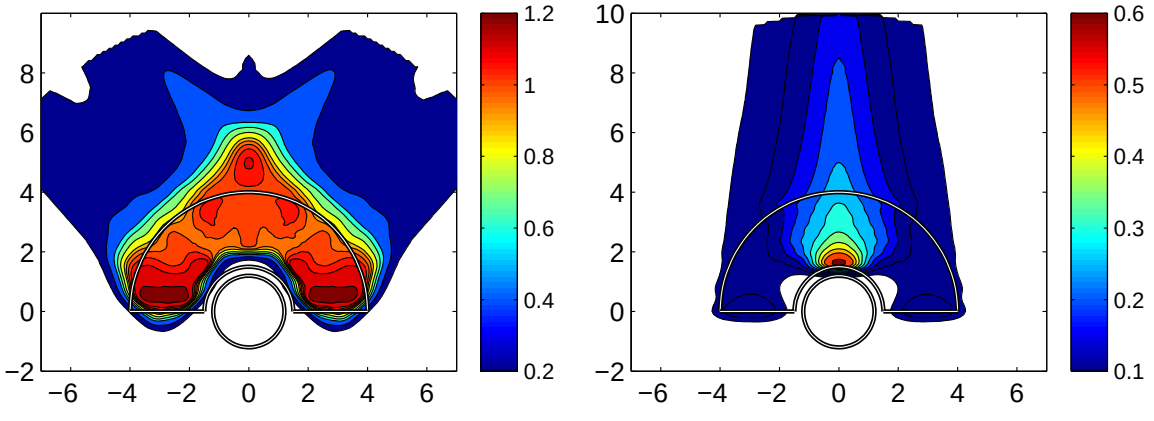
The work presented in this chapter focuses only on range uncertainty. However, a comprehensive robust framework should include other types of uncertainty such as setup error and, of particular importance in proton therapy, implementation errors. Implementation errors could be related to uncertainties in the delivery system that cause variations in the energy of the protons, which would affect the location of their Bragg peaks.

A computational extension of this work would be to study the solution of the various robust formulations presented in this chapter under different algorithms. This research would also be useful to guide the implementation of other robust and/or clinical formulations in both proton and photon therapy. Clinical problems are very large-scale optimization problems, which typically become even larger once robustness is included. The adoption of such techniques in practice, therefore, hinges on the efficient solution of these formulations.



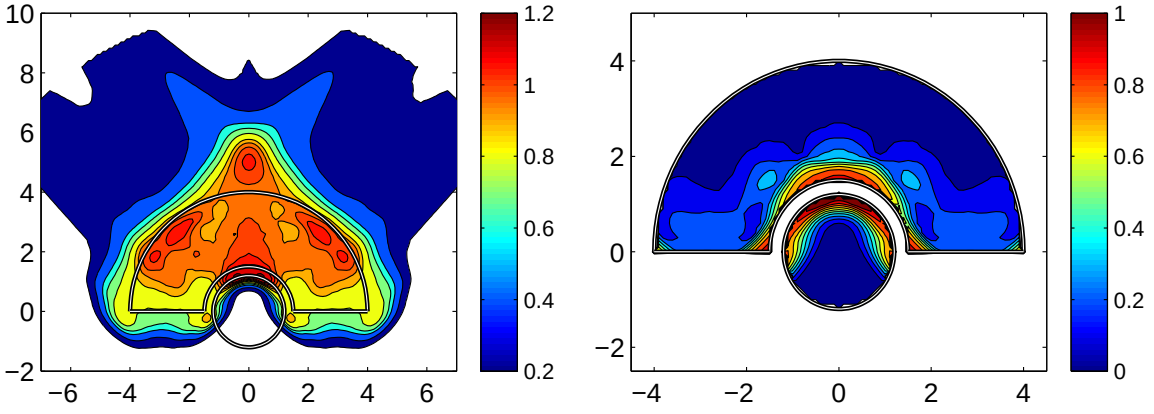
(a) Nominal range.

(b) Beam at 45 degrees.



(c) Shorter range.

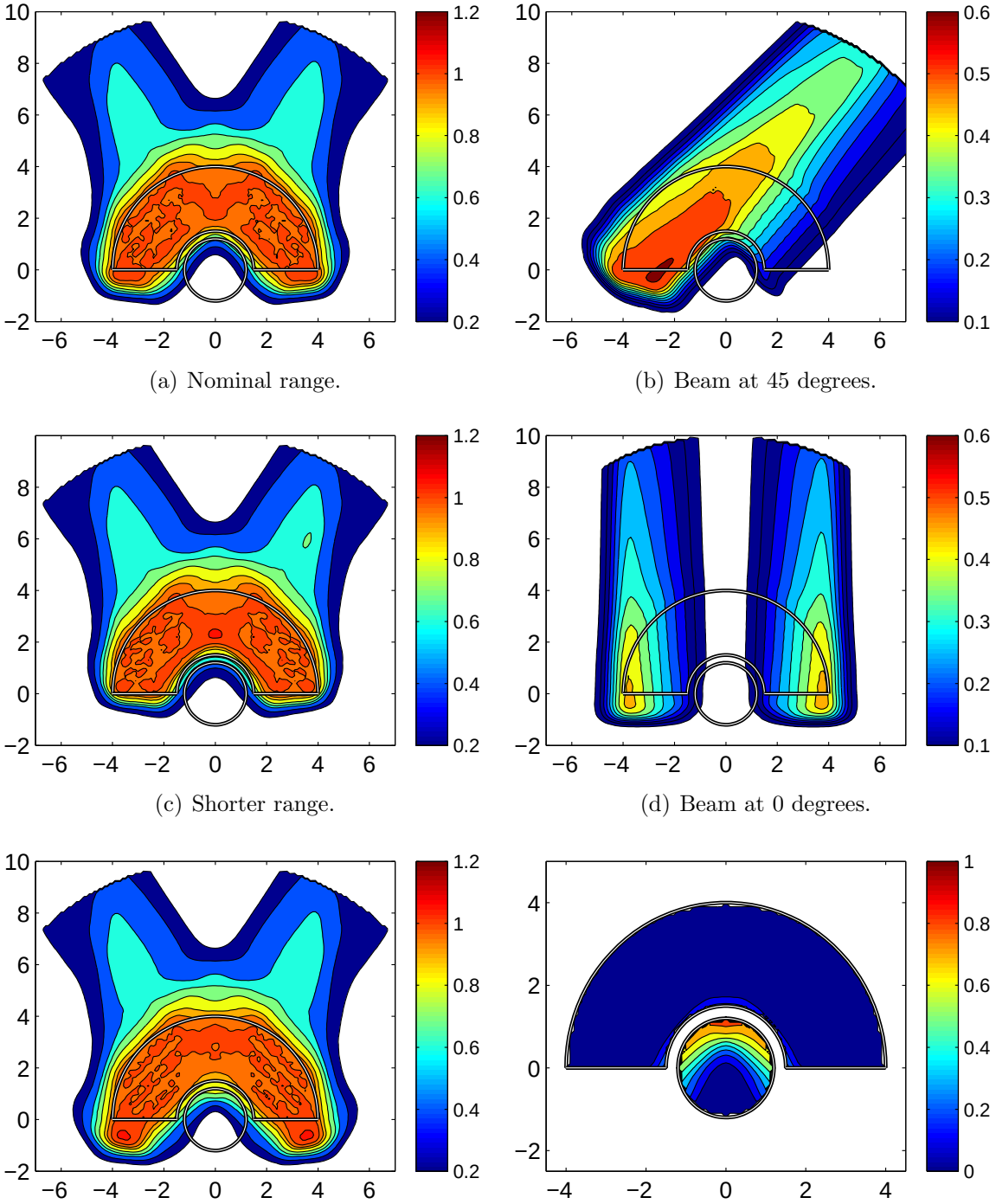
(d) Beam at 0 degrees.



(e) Longer range.

(f) Maximum deviation.

Figure 4-3: The nominal treatment plan. (a) Dose distribution assuming each beamlet realizes its nominal range; (c/e) Dose distribution assuming each beamlet's range is shortened/lengthened by 5 mm; (b/d) Dose distribution of the individual beam at 45°/0° assuming each beamlet realizes its nominal range; (f) The maximum deviation of the delivered dose from the prescribed dose over all 11 ranges.



(a) Nominal range.

(b) Beam at 45 degrees.

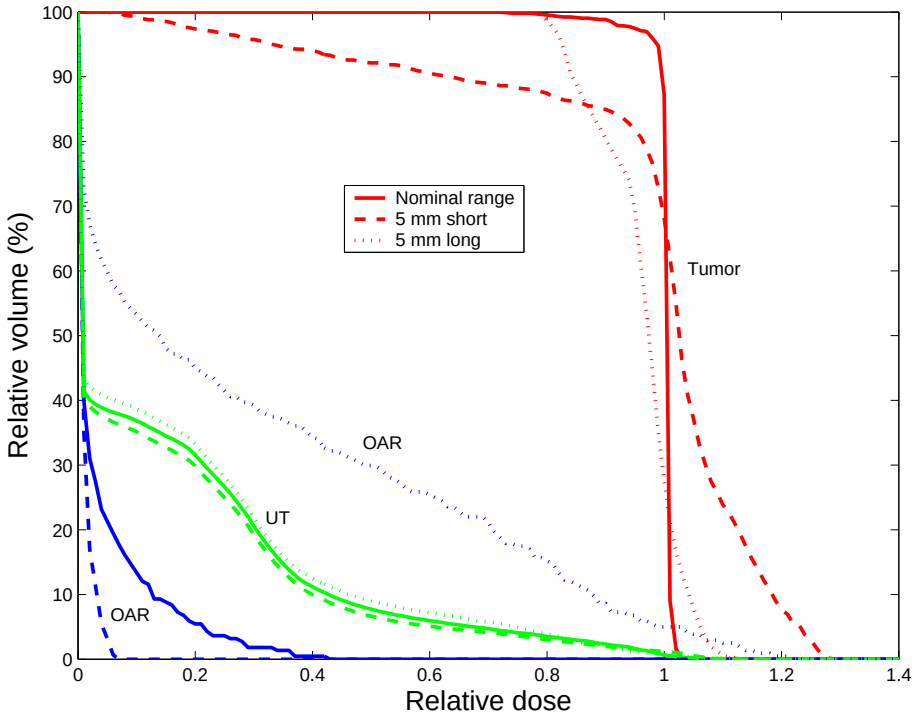
(c) Shorter range.

(d) Beam at 0 degrees.

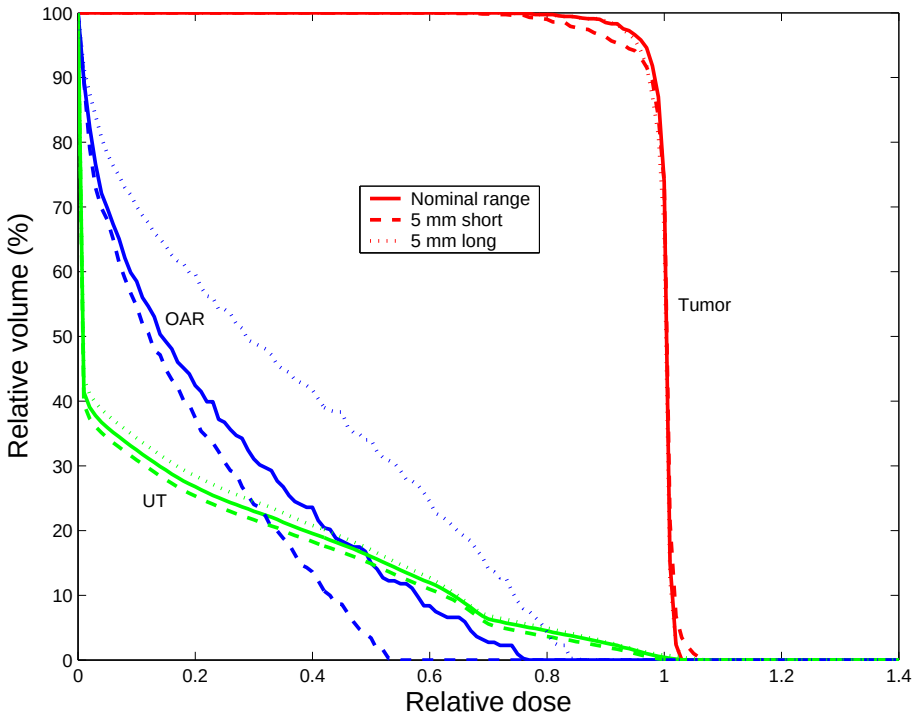
(e) Longer range.

(f) Maximum deviation.

Figure 4-4: The robust treatment plan. (a) Dose distribution assuming each beamlet realizes its nominal range; (c/e) Dose distribution assuming each beamlet's range is shortened/lengthened by 5 mm; (b/d) Dose distribution of the individual beam at 45°/0° assuming each beamlet realizes its nominal range; (f) The maximum deviation of the delivered dose from the prescribed dose over all 11 ranges.

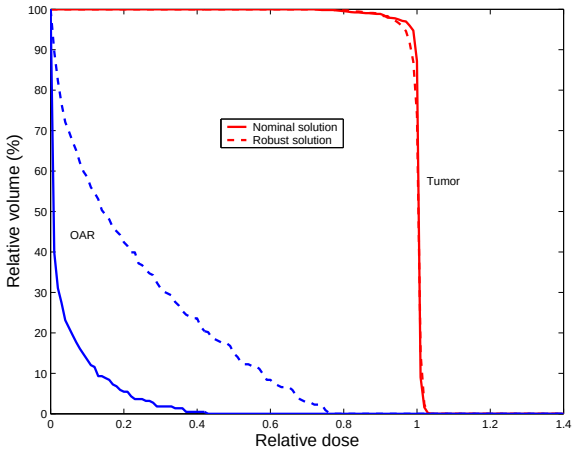


(a) Nominal solution.

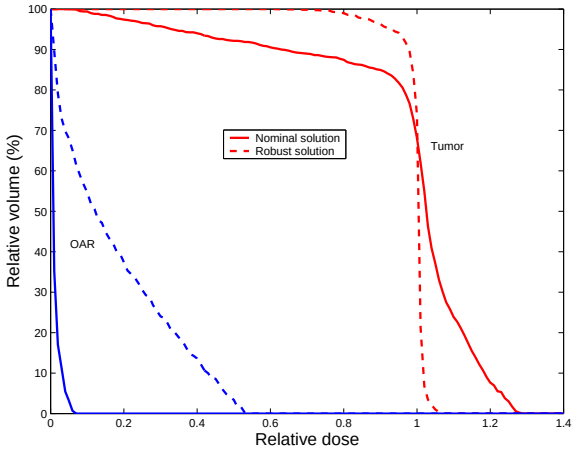


(b) Robust solution.

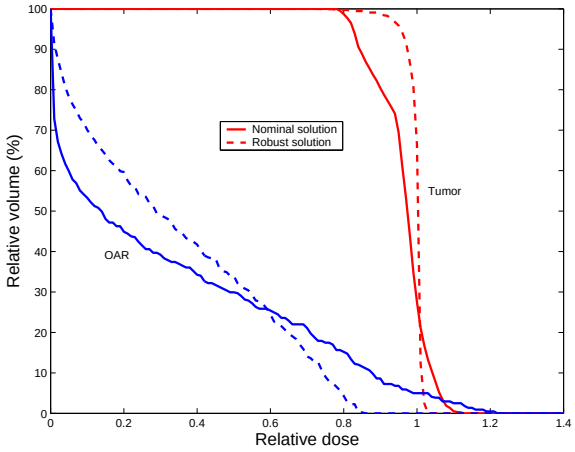
Figure 4-5: DVHs of the nominal (a) and robust (b) solution under three scenarios: all beamlets realize their nominal range, shortest range, or longest range.



(a) Nominal range.



(b) Shorter range.



(c) Longer range.

Figure 4-6: DVHs of the nominal and robust solutions under three scenarios: all beamlets realize their nominal range (a), shortest range (b), or longest range (c).

Chapter 5

Conclusions

Radiation therapy is subject to uncertainties that need to be accounted for when determining a suitable treatment plan for a cancer patient. Traditional optimization methods that are used to produce treatments for a particular patient may suffer from two distinct problems. First, they may not protect against any uncertainty at all, which could result in a tumor that does not receive sufficient dose. Or, they may be far too conservative and sacrifice unnecessary amounts of healthy tissue during the treatment. This thesis addresses these trade-offs using robust optimization. By modeling the various uncertainties in the underlying problem and incorporating them into the optimization, we provide a tailored solution for the particular uncertainty at hand, while balancing the sparing of healthy tissue with ensuring sufficient coverage of the tumor volume.

Currently, there are few, if any, arbitrage opportunities in terms of improving an optimized treatment in all dimensions. It is unlikely that local changes in optimization parameters will result in improved solutions, with respect to all of the different objectives that one can argue are important. Hence, focus should be directed towards uncovering the relevant trade-offs in each case, and equipping treatment planners with the right tools and understanding to make the best possible decisions under uncertainty. A suitable robust optimization framework will allow these decision makers to understand the effects of uncertainty on the treatment outcome, and aid in the creation of comprehensive clinical solutions.

One subtle but important aspect of the work presented in this thesis is that these improvements in robustness are software (i.e., mathematical) fixes. Theoretically, by updating the optimization processes that are employed in current treatment planning systems, our robust approaches could be used in the clinic today. Therefore, the potential exists to realize these improvements with minimal additional cost and changes to current treatment protocols. On the other hand, hardware solutions for addressing uncertainty typically prolong the treatment time, as well as require additional human effort to implement. Also, these solutions may themselves introduce additional operational uncertainty into the problem. Therefore, we must be mindful that the solutions we create to address a particular problem do not introduce other challenges elsewhere in the same problem, or in the treatment pipeline either downstream or upstream.

In the future, in order to deliver the best possible treatments and outcomes for patients, we will need to move towards a more integrated view of the processes involved. Attention must be paid to the interplay between treatment optimization and treatment operations. We need to ensure that our carefully crafted solutions to micro-level problems such as individual treatments do not adversely affect the macro-level operations of the facility that is carrying out the treatments. By doing so, we will move towards optimality for both the individual patient and the larger health care system.

Bibliography

[1] Ron Alterovitz, Etienne Lessard, Jean Pouliot, I-Chow Joe Hsu, James F. O'Brien, and Ken Goldberg. Optimization of hdr brachytherapy dose distributions using linear programming with penalty costs. *Medical Physics*, 33(11):4012–4019, 2006.

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

[3] C. Baum, M. Alber, M. Birkner, and F. Nusslin. Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities. *Radiother. Oncol.*, 78:27–35, 2006.

[4] Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Prog. A*, 88(3):411–424, 2000.

[5] Aharon Ben-Tal and Arkadi Nemirovski. Robust optimization—methodology and applications. *Math. Prog. B*, 92(3):453–480, 2002. ISMP 2000, Part 2 (Atlanta, GA).

[6] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Oper. Res.*, 52(1):35–53, 2004.

[7] P. J. Biggs and W. U. Shipley. A beam width improving device for a 25 MV x ray beam. *Int. J. Radiat. Oncol. Biol. Phys.*, 12:131–135, 1986.

[8] T. Bortfeld, S. B. Jiang, and E. Rietzel. Effects of motion on the total dose distribution. *Semin. Radiat. Oncol.*, 14(1):41–51, 2004.

[9] T. Bortfeld and W. Schlegel. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Phys. Med. Biol.*, 38:291–304, 1993.

[10] Thomas Bortfeld. IMRT: A review and preview. *Phys. Med. Biol.*, 51:R363–R379, 2006.

[11] Thomas Bortfeld, Kimmo Jokivarsi, Michael Goitein, Jong Kung, and Steve B Jiang. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation. *Phys. Med. Biol.*, 47(13):2203–2220, 2002.

[12] K. M. Brock, J. M. Balter, L. A. Dawson, M. L. Kessler, and C. R. Meyer. Automated generation of a four-dimensional model of the liver using warping and mutual information. *Medical Physics*, 30(6):1128–1133, 2003.

[13] Timothy C Y Chan, Thomas Bortfeld, and John N Tsitsiklis. A robust approach to IMRT optimization. *Phys. Med. Biol.*, 51(10):2567–2583, 2006.

[14] Millie Chu, Yuriy Zinchenko, Shane G Henderson, and Michael B Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Phys. Med. Biol.*, 50(23):5463–5477, 2005.

[15] David Craft. Local beam angle optimization with linear programming and gradient search. *Physics in Medicine and Biology*, 52(7):N127–N135, 2007.

[16] M. Engelsman, G. C. Sharp, T. Bortfeld, R. Onimaru, and H. Shirato. How much margin reduction is possible through gating or breath hold? *Phys. Med. Biol.*, 50:477–490, 2005.

[17] Michael C. Ferris, Jinho Lim, and David M. Shepard. An optimization approach for radiosurgery treatment planning. *SIAM J. Optim.*, 13(3):921–937, 2003.

[18] R. George, P. J. Keall, V. R. Kini, S. S. Vedam, V. Ramakrishnan, and R. Mohan. Is the diaphragm motion probability density function normally distributed? *Med. Phys.*, 32(2):396–404, 2005.

[19] D. P. Gierga, J. Brewer, G. C. Sharp, M. Betke, C. G. Willett, and G. T. Y. Chen. The correlation between internal and external markers for abdominal tumors: implications for respiratory gating. *Int. J. Radiat. Oncol. Biol. Phys.*, 61(5):1551–1558, 2005.

[20] A. Holder. Designing radiotherapy plans with elastic constraints and interior point methods. *Health Care Manag Sci*, 6(1):5–16, 2003.

[21] I. A. Ibragimov. On the composition of unimodal distributions. *Theory Prob. Appl.*, 1(3), 1956.

[22] IMRT Collaborative Working Group. Intensity-modulated radiotherapy: current status and issues of interest. *Int. J. Radiat. Oncol. Biol. Phys.*, 51(4):880–914, 2001.

[23] P J Keall, G Starkschall, H Shukla, K M Forster, V Ortiz, C W Stevens, S S Vedam, R George, T Guerrero, and R Mohan. Acquiring 4D thoracic CT scans using a multislice helical method. *Physics in Medicine and Biology*, 49(10):2053–2067, 2004.

[24] Paul J. Keall, Gig S. Mageras, James M. Balter, Richard S. Emery, Kenneth M. Forster, Steve B. Jiang, Jeffrey M. Kapatoes, Daniel A. Low, Martin J. Murphy, Brad R. Murray, Chester R. Ramsey, Marcel B. Van Herk, S. Sastry Vedam, John W. Wong, and Ellen Yorke. The management of respiratory motion in radiation oncology report of AAPM task group 76. *Medical Physics*, 33(10):3874–3900, 2006.

[25] L. Y. Klepper. Electronic computers and methods of linear programming in the choice of optimial conditions for radiation teletherapy. *Medicinskay Radiol.*, 11:8–15, 1966.

[26] K. M. Langen and D. T. Jones. Organ motion and its management. *Int. J. Radiat. Oncol. Biol. Phys.*, 50(1):265–278, 2001.

[27] E. K. Lee, T. Fox, and I. Crocker. Simultaneous beam geometry and intensity map optimization in intensity-modulated radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 64(1):301–320, 2006.

[28] Eva K. Lee, Tim Fox, and Ian Crocker. Integer programming applied to intensity-modulated radiation therapy treatment planning. *Annals of Operations Research*, 119:165–181, 2003.

[29] J Lof, B K Lind, and A Brahme. Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy. *Inverse Problems*, 11(6):1189–1209, 1995.

[30] A. Lomax. Intensity modulation methods for proton radiotherapy. *Physics in Medicine and Biology*, 44(1):185–205, 1999.

[31] A. J. Lomax, T. Boehringer, A. Coray, E. Egger, G. Goitein, M. Grossmann, P. Juelke, S. Lin, E. Pedroni, B. Rohrer, W. Roser, B. Rossi, B. Siegenthaler, O. Stadelmann, H. Stauble, C. Vetter, and L. Wisser. Intensity modulated proton therapy: A clinical example. *Medical Physics*, 28(3):317–324, 2001.

[32] Daniel A. Low, Michelle Nystrom, Eugene Kalinin, Parag Parikh, James F. Dempsey, Jeffrey D. Bradley, Sasa Mutic, Sasha H. Wahab, Tareque Islam, Gary Christensen, David G. Politte, and Bruce R. Whiting. A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing. *Medical Physics*, 30(6):1254–1263, 2003.

[33] Weiguo Lu, Ming-Li Chen, Gustavo H Olivera, Kenneth J Ruchala, and Thomas R Mackie. Fast free-form deformable registration via calculus of variations. *Physics in Medicine and Biology*, 49(14):3067–3087, 2004.

[34] A. E. Lujan, E. W. Larsen, J. M. Balter, and R. K. Ten Haken. A method for incorporating organ motion due to breathing into 3D dose calculations. *Med. Phys.*, 26(5):715–720, 1999.

[35] Tim McInerney and Demetri Terzopoulos. Deformable models in medical image analysis: a survey. *Med. Image Anal.*, 1(2):91–108, 1996.

[36] Loren K. Mell, Amik K. Mehrotra, and Arno J. Mundt. Intensity-modulated radiation therapy use in the U.S., 2004. *Cancer*, 104(6):1296–1303, 2005.

[37] Radhe Mohan, Qiuwen Wu, Xiaohong Wang, and Jorg Stein. Intensity modulation optimization, lateral transport of radiation, and margins. *Medical Physics*, 23(12):2011–2021, 1996.

[38] A. Niemierko. Biological optimization. In Thomas Bortfeld, Rupert Schmidt-Ullrich, Wilfried De Neve, and David E. Wazer, editors, *Image-Guided IMRT*. Springer, 2006.

[39] Arinbjörn Ólafsson and Stephen J Wright. Efficient schemes for robust IMRT treatment planning. *Physics in Medicine and Biology*, 51(21):5621–5642, 2006.

[40] Tinsu Pan, Ting-Yim Lee, Eike Rietzel, and George T. Y. Chen. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Medical Physics*, 31(2):333–340, 2004.

[41] K Parodi, W Enghardt, and T Haberer. In-beam PET measurements of β^+ radioactivity induced by proton beams. *Physics in Medicine and Biology*, 47(1):21–36, 2002.

[42] Katia Parodi and Thomas Bortfeld. A filtering approach based on gaussian-powerlaw convolutions for local PET verification of proton radiotherapy. *Physics in Medicine and Biology*, 51(8):1991–2009, 2006.

[43] Felisa Preciado-Walters, Ronald Rardin, Mark Langer, and Van Thai. A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer. *Math. Prog. B*, 101:319–338, 2004.

[44] R. Reemtsen and M. Alber. Continuous optimization of beamlet intensities for photon and proton radiotherapy. Technical report, Brandenburgisch Technische Universität Cottbus, 2004.

[45] H Edwin Romeijn, Ravindra K Ahuja, James F Dempsey, Arvind Kumar, and Jonathan G Li. A new linear programming approach to radiation therapy treatment planning problems. *Oper. Res.*, 54(2):201–216, 2006.

[46] M. B. Sharpe, B. M. Miller, and J. W. Wong. Compensation of x-ray beam penumbra in conformal radiotherapy. *Medical Physics*, 27(8):1739–1745, 2000.

[47] Ke Sheng, Jing Cai, James Brookeman, Janelle Molloy, John Christopher, and Paul Read. A computer simulated phantom study of tomotherapy dose optimization based on probability density functions (pdf) and potential errors caused by low reproducibility of pdf. *Med. Phys.*, 33(9):3321–3326, 2006.

[48] David M. Shepard, Michael C. Ferris, Gustavo H. Olivera, and T. Rockwell Mackie. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Rev.*, 41(4):721–744, 1999.

[49] Ron S. Sloboda. Optimization of brachytherapy dose distributions by simulated annealing. *Medical Physics*, 19(4):955–964, 1992.

[50] S. Söderström and A. Brahme. Optimization of the dose delivery in a few field techniques using radiobiological objective functions. *Med. Phys.*, 20(4):1201–1210, 1993.

[51] J. C. M. Theuws, S. L. S. Kwa, A. C. Wagenaar, Y. Seppenwoolde, L. J. Boersma, E. M. F. Damen, S. H. Muller, P. Baas, and J. V. Lebesque. Prediction of overall pulmonary function loss in relation to the 3-D dose distribution for patients with breast cancer and malignant lymphoma. *Radiother. Oncol.*, 49(3):233–243, 1998.

[52] A. Trofimov and T. Bortfeld. Optimization of beam parameters and treatment planning for intensity modulated proton therapy. *Technol Cancer Res Treat*, 2(5):437–44, 2003.

[53] A. Trofimov, A. Knopf, H. Jiang, T. Bortfeld, and H. Paganetti. Proton dose calculation using monte-carlo-validated pencil beam database for KonRad treatment planning system. *Medical Physics*, 32:2030, 2005.

[54] Alexei Trofimov, Eike Rietzel, Hsiao-Ming Lu, Benjamin Martin, Steve Jiang, George T Y Chen, and Thomas Bortfeld. Temporo-spatial IMRT optimization: concepts, implementation and initial results. *Phys. Med. Biol.*, 50(12):2779–2798, 2005.

[55] Y. Tsunashima, T. Sakae, Y. Shioyama, K. Kagei, T. Terunuma, A. Nohtomi, and Y. Akine. Correlation between the respiratory waveform measured using a respiratory sensor and 3D tumor motion in gated radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 60(3):951–958, 2004.

[56] J. Unkelbach and U. Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Phys. Med. Biol.*, 49(17):4005–4029, 2004.

[57] M. van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque. The probability of correctr target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 47(4):1121–1135, 2000.

[58] S. S. Vedam, V. R. Kini, P. J. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan. Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker. *Med. Phys.*, 30(4):505–513, 2003.

[59] S. Webb. The physical basis of IMRT and inverse planning. *Br. J. Radiol.*, 76:678–689, 2003.

[60] H. R. Withers, J. M. G. Taylor, and B. Maciejewski. Treatment volume and tissue tolerance. *Int. J. Radiat. Oncol. Biol. Phys.*, 14(4):751–759, 1988.

[61] A. B. Wolbarst, L. M. Chin, and G. K. Svensson. Optimization of radiation therapy: Integral-response of a model biological system. *Int. J. Radiat. Oncol. Biol. Phys.*, 8(10):1761–1769, 1982.

[62] T. Zhang, R. Jeraj, H. Keller, W. Lu, G. H. Olivera, T. McNutt, T. Mackie, and B. Paliwal. Treatment plan optimization incorporating respiratory motion. *Med. Phys.*, 31(6):1576–1586, 2004.