Three Essays on Radiotherapy Treatment Planning Optimization

by

Victor W. Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2017

Doctoral Committee:

Associate Professor Marina A. Epelman, Co-Chair
Professor H. Edwin Romeijn, Co-Chair
Professor Brian T. Denton
Assistant Professor Martha M. Matuszak

Victor W. Wu

vwwu@umich.edu

ORCID iD: 0000-0001-5668-581X

For Mom, Dad, and Alan

ACKNOWLEDGEMENTS

First, I'd like to acknowledge my dissertation committee. Thanks to my PhD advisors, Professors Marina Epelman and Edwin Romeijn, I've not only grown as an operations researcher and engineer, but also as an intellect. It may sound silly, but they really taught me how to *think*: how to question, investigate, and communicate. I am indebted to Professor Martha Matuszak for her guidance in all things related to medical physics and clinical oncology. I also appreciate the valuable feedback and additional healthcare OR perspective of Professor Brian Denton. They were truly models of excellence who challenged me to pursue my own.

I would also like to thank Professors Randy Ten Haken, Mary Feng, Kristy Brock, Chris Deufel, Mustafa Sir, and Kalyan Pasupathy for their collaboration in my thesis research.

I discovered operations research first at Cornell and would like to thank Professors Robert Bland, Paat Rusmevichientong, Gennady Samorodnitsky, and Mark Lewis for the foundation in OR I came to Michigan with. I would like to especially thank Professor Shane Henderson for getting my feet wet in OR research and introducing me to OR in healthcare.

Thank you to Chris, Tina, Candy, Kristi, Rod, and Mint for all of their help throughout the IOE department.

I'm fortunate to have made many friends during my time in Ann Arbor, including Troy, Jeremy, Steph, Lauren, Colleen, Uriah, Joe, Min, Ryan, Wilmer, Karmel, Ilbin, Gregg, Elliot, Zohar, Drew, and Kyle. I especially owe Troy many thanks for taking me under his wing the first few years and for being such a big support throughout my studies (and beyond!).

Finally, I'd like to thank my mom, dad, and brother, for always inspiring me, believing

in me, and keeping me grounded.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES viii

LIST OF TABLES xii

LIST OF ABBREVIATIONS xiv

ABSTRACT xvi

CHAPTER

I. Introduction 1

1.1 Radiation Therapy 1
1.1.1 Intensity Modulated Radiation Therapy (IMRT) 2
1.1.2 High-Dose Rate Brachytherapy (HDR-BT) 3
1.1.3 Treatment Plan Optimization 4
1.1.4 Chapter Summaries 7

II. Optimizing Global Liver Function 10

2.1 Introduction 10
2.2 Mathematical Optimization Models for Liver SBRT Treatment Planning 12
2.2.1 Notation and Treatment Criteria 12
2.2.2 Optimization Models 14
2.2.3 Perfusion Avoidance Model 15
2.2.4 Global Liver Function Preservation Model 16
2.3 Experiments 20
2.3.1 Model Parameters and Implementation Details 21
2.3.2 Liver Perfusion Patterns 22
2.4 Results 25
2.4.1 2D Liver Cases 26

2.4.2 Clinical (3D) Liver Cases . . . 27
2.5 Discussion . . . 29
2.6 Conclusion . . . 36

III. Adaptive Liver Stereotactic Body Radiotherapy . . . 38

3.1 Introduction . . . 38
3.2 Related Literature . . . 40
3.3 Proposed Framework . . . 43
3.3.1 Patient Geometry Uncertainty . . . 44
3.3.2 Chapter Outline . . . 44
3.4 Treatment Plan Evaluation Criteria . . . 46
3.4.1 Dose Effect Over Time: Biologically Effective Dose . . . 46
3.4.2 Structure Criteria . . . 47
3.5 Deterministic Model . . . 48
3.6 Stochastic Models . . . 49
3.6.1 A More Concise Formulation . . . 51
3.7 Diminishing Horizon Heuristic . . . 53
3.7.1 History-Independent Stochastic Model . . . 53
3.7.2 Diminishing Horizon Heuristic Based on History-Independent Model . . . 57
3.8 Computational Experiment . . . 59
3.8.1 Representing Interfraction Motion . . . 59
3.8.2 Experimental setup . . . 59
3.8.3 Results . . . 62
3.9 Discussion . . . 64
3.9.1 Unique Number of Scenarios in DHH . . . 66
3.10 Conclusion . . . 67

IV. Multicriteria Optimization for Brachytherapy Treatment Planning 68

4.1 Introduction . . . 68
4.1.1 Related Literature . . . 69
4.2 Methods and Materials . . . 73
4.2.1 Patient Cases . . . 73
4.2.2 Trade-off Surface . . . 73
4.2.3 Notation . . . 76
4.3 Representing and Navigating the Trade-off Surface . . . 76
4.3.1 "Filling out" the Trade-off Surface via Interpolation . . . 77
4.3.2 Navigating the Trade-off Surface . . . 80
4.4 Generating Initial Library of Plans via Convex Optimization . . . 81
4.4.1 Metrics: Value at Risk and a Convex Approximation . . . 81
4.4.2 Multicriteria Optimization Treatment Planning Model . . . 83
4.4.3 Voxel Truncation . . . 84
4.4.4 Library Generation: ϵ-Constraint Method . . . 87

4.5 Experiments . . . 88
4.5.1 Choice of $1 - \Delta^{\mathrm{TCVaR}}$ and τ . . . 89
4.5.2 Choice of ϵ . . . 89
4.6 Results . . . 90
4.6.1 Plan Interpolation By Dwell Time . . . 90
4.6.2 Quality of a Trade-off Surface Generated with a $\gamma_\Delta, \gamma_\tau$ — One Truncation . . . 91
4.6.3 Quality of a Trade-off Surface Generated with a $\gamma_\Delta, \gamma_\tau$ — Iterative Truncation . . . 92
4.6.4 Truncation Accuracy . . . 95
4.6.5 Testing Parameter Choice . . . 96
4.6.6 Comparison of Plan Quality with Commercial System . . . 96
4.7 Discussion . . . 97
4.7.1 Interpolation . . . 101
4.7.2 Computational Efforts . . . 104
4.8 Conclusions . . . 104

BIBLIOGRAPHY . . . 105

LIST OF FIGURES

Figure

1.1 Therapy Flow 3
1.2 Multileaf collimator composed of tungsten leaves that mold the beam shape. 4
1.3 BT Delivery 4
1.4 BT Afterloader 4
1.5 Treatment plan for IMRT (left): blue and other colors inside represent patient geometry (tumor in pink); green blocks represent beam angles; gray corresponds to beam intensity profile. Treatment plan for BT (right): The prostate and tumor are depicted in red; green and blue lines represent catheters and alternating shades represent seed dwell locations. 5
2.1 Comparison between examples of true local response (blue) and simple logistic approximation (red) varied by pre-treatment perfusion levels. 19
2.2 Geometry of 2D liver case (axial view). The small contour represents the boundary of the PTV voxels and is surrounded by normal liver voxels. The large contour represents the boundary of the patient and contains normal tissue voxels that are not shown. 23
2.3 Commonly encountered perfusion patterns on DCE-MRI (left); Synthesized 2D liver cases with PTV in gray and surrounding liver voxels color-coded by relative importance (i.e., $g_j(f_j)$) values (right). 24
2.4 Pre- and post-treatment GLF achieved by ℓEUD, fEUD and GLF plans on 2D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function. (Note: all cases have some deficit in function at the beginning of treatment.) 28
2.5 Visualization of pre- and post-treatment liver function maps and dose wash differences for 2D liver perfusion pattern P1: There is noticeable reduction in dose to high-functioning regions using either fEUD or GLF objective. (2.5a), (2.5d)-(2.5f) are function maps; (2.5b)-(2.5c) are dose wash differences. 29
2.6 Visualization of pre- and post-treatment liver function maps and dose wash differences for 2D liver perfusion pattern P3: There is noticeable reduction in dose to high-functioning regions using GLF objective but limited difference using fEUD objective. (2.6a), (2.6d)-(2.6f) are function maps; (2.6b)-(2.6c) are dose wash differences. 30

2.7 Pre- and post-treatment GLF achieved by ℓEUD, fEUD, and GLF plans on 3D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function. (Note: all cases have some deficit in function at the beginning of treatment.) 31
2.8 Dose distributions for ℓEUD, fEUD, and GLF plans illustrated for Patient Case 1. 32
2.9 Dose distributions, and their differences, for ℓEUD, fEUD, and GLF plans illustrated for Patient Case 1. Liver is contoured in light brown and filled with its black-and-white perfusion map. 33
2.10 Distribution of voxels with high pre-treatment function by dose-bins of 10 Gy-EQD2 (one bin for dose-saturated voxels) along post-treatment function line-plot; 2D case P1. 35
2.11 Distribution of voxels with high pre-treatment function by dose-bins of 10 Gy-EQD2 (one bin for dose-saturated voxels) along post-treatment function line-plot; Patient Case 1. 36
3.1 Left: with the gall bladder filled, the colon (yellow arrows) is a safe distance from the high dose region around the PTV; Right: the gall bladder loses volume and the colon moves closer to the high dose region, potentially resulting in overdosing the colon. Red and green contours correspond to the 100% and 50% of prescription dose iso-dose lines. 39
3.2 Example of IGRT. Left: duodenum shifts toward PTV. Middle: shifted duodenum would be overdosed with more than 6 Gy in the original plan. Right: IGRT shifts the dose distribution to avoid overdosing the shifted duodenum. 39
3.3 Example scenario tree with 3 possible geometries at each fraction. One path from the first fraction to the last fraction is considered a scenario. 45
3.4 2D example of assignment of dose deposition coefficients for deformed geometries. The 3-by-3 grid represents static voxels corresponding to dose deposition coefficients (D_{ij}). In this example, the blue voxel moves from the center to the right middle location and is assigned the coefficients corresponding to the latter location. The red voxels are compressed and both will have the dose coefficients of the bottom center static voxel. 60
3.5 Distribution of PTV Coverage achieved by plans generated via DHH. In 90% of scenarios were able to meet the typical prescription dose of 132 Gy (black dashed line). Distribution is also categorized by number times the most limiting geometry, ω_3, occurred (bar colors). 63
3.6 PTV Coverage comparison: applying DHH versus using solution of (SO-HI), a.k.a. $\text{DHH}_{\text{1st Stage}}$. 65
4.1 Screen capture of example GUI. Left: trade-off(s) can be made by selecting which criterion to "heat" (increase dose) vs "cool" (decrease dose) at the expense of another criterion (other criteria) by the green bar selection; Right: Visualization of DVH, dose distribution, and dwell times corresponding to the current plan on the trade-off surface. 75

4.2 Populating a trade-off surface for the example breast case. Red asterisks are plans generated from optimization (points in $\mathcal{L}$ projected to form L; in this example, PTV D95 and PTV D115cc are the metrics used in defining L); a triangulation is generated based on these plans (red lines). Black crosses represent the set of additional plans, Q, used to increase the density of our representation of the trade-off surface; using the triangulation, the dwell times of the library plans are interpolated to create plans on the blue trade-off surface, which is plotted as a mesh. 78
4.3 Populating a trade-off surface for the example breast case. Red asterisks are plans generated from optimization (points in $\mathcal{L}$ projected to form L; in this example, PTV D95 and Urethra D10 dose are the metrics used in defining L); a triangulation is generated based on these plans (red lines). Black crosses represent the set of additional plans, Q, used to increase the density of our representation of the trade-off surface; using the triangulation, the dwell times of the library plans are interpolated to create plans on the blue trade-off surface, which is plotted as a mesh. 79
4.4 VaR and CVaR example on a differential DVH (left). $D_{5\%}$ is VaR at 5%; UCVaR at 5%; and TCVaR at 5% with 2% of the hottest voxels (purple volume) excluded. CVaR is the mean dose of the volume highlighted by the blue and purple. 82
4.5 Breast 1. Without truncation, the minimum difference from the best is 1% while others only go up to 0.5%. 93
4.6 Breast 2. Without truncation, the minimum difference from the best is 0.5%, but at worst 3%; although many are within 0.5% of the best, several go up to beyond 1% depending on the level of truncation. 93
4.7 Breast 3. Without truncation, the minimum difference from the best is 0.5%, but at worst 3%; although many are within 0.5% of the best, several go up to beyond 1% depending on the level of truncation. 94
4.8 Breast 1. Effects of iterative truncation. For a given truncation iteration (e.g., all plans obtained by truncating only twice), dose differences are computed using (4.19) and the distribution of percent differences are plotted as a box plot for each $\gamma_\Delta, \gamma_\tau$ combination. 95
4.9 Breast 1. Effects of iterative truncation. For a set of partial candidate plans, the $K+1^{\text{st}}$ metric is computed using the library by metric interpolation. For a given truncation iteration (e.g., all plans obtained by truncating only twice), dose differences are computed using (4.19) and the distribution of percent differences are plotted as a box plot for each $\gamma_\Delta, \gamma_\tau$ combination (truncation iteration is clarified with the choice of γ_Δ). 96
4.10 Breast 2. Truncation Accuracy by Iteration of Truncation (iteration 1 is truncating according to clinical plan); up to 3 iterations of truncation considered. $\gamma_\tau = 0.0625, 0.33, 0.75$. 98
4.11 Breast 2. Effects of iterative truncation. For a given truncation iteration (e.g., all plans obtained by truncating only twice), dose differences are computed using (4.19) and the distribution of percent differences are plotted as a box plot for each $\gamma_\Delta, \gamma_\tau$ combination. 98

4.12 Breast 3. Choosing $(\gamma_\Delta, \gamma_\tau) = (2, 0.0.33)$ to generate a library based on Breast 1 and Breast 2. 99
4.13 Breast 1 PTV Coverage Comparison with a Commercial System. A set of plans generated by an experienced medical physicist using the Eclipse Treatment Planning System. This set of plans was compared to the different libraries by fixing all other metrics and comparing PTV D95 achieved. Note these are sorted by TCVaR tail size and grouped by truncation size and truncation iteration. 100
4.14 Breast 2 PTV Coverage Comparison with a Commercial System. A set of plans generated by an experienced medical physicist using the Eclipse Treatment Planning System. This set of plans was compared to the different libraries by fixing all other metrics and comparing PTV D95 achieved. Note these are sorted by TCVaR tail size and grouped by truncation size and truncation iteration. 101
4.15 Example Prostate Case. Due to the wide range of values relative to the a particular library, either more truncations are needed or a better quality plan indicating hottest voxels is needed. 102
4.16 Prostate 1. Truncation Accuracy by Iteration of Truncation (iteration 1 is truncating according to clinical plan); up to 3 iterations of truncation considered. $\gamma_\tau = 0.0625, 0.33, 0.75$. 102
4.17 Example Cervix Case. Due to the wide range of values relative to the a particular library, either more truncations are needed or a better quality plan indicating hottest voxels is needed. 103
4.18 Cervix 1. Truncation Accuracy by Iteration of Truncation (iteration 1 is truncating according to clinical plan); up to 3 iterations of truncation considered. $\gamma_\tau = 0.0625, 0.33, 0.75$. 103

LIST OF TABLES

Table

2.1 Parameter values used in models for the 2D cases. 25
2.2 Parameter values used in models for the 3D cases. [a]Approximated lower and upper bounds based on clinical practice at University of Michigan Hospital. 25
2.3 $\ell\text{EUD}^{+}_{\text{fLiver}}$ (Gy) achieved by fEUD and GLF plans on 2D cases. 27
2.4 Pre- and post-treatment GLF achieved by ℓEUD, fEUD and GLF plans on 2D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function. 27
2.5 $\ell\text{EUD}^{+}_{\text{fLiver}}$ (Gy) achieved by ℓEUD, fEUD, and GLF plans on 3D cases. . . 28
2.6 Pre- and post-treatment GLF achieved by ℓEUD, fEUD, and GLF plans for 3D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function. 34
3.1 Summary of simplifications made in some of our models, their effects, and potential for future work. 61
3.2 Structures and their parameters for patient liver case. For measuring PTV coverage, $\lambda^{-}_{\text{PTV}} = 0.84$. 62
3.3 Per-fraction structure criteria resulting from the application of the nominal plan to each potential geometry (ω_1 is nominal). Intolerable levels of dose are in red. 62
3.4 DHH's average instance runtime by length of history τ. $K = 4; T = 5$. . . . 64
4.1 DVH Metric Clinical Protcol (% Tx Dose) for Breast Cases. Tx Dose = Prescribed dose to 95% of the tumor volume. There are three Tx doses listed, one for each breast case tested. 73
4.2 DVH Metric Clinical Protcol (% Tx Dose) for Cervix Cases. Tx Dose = Prescribed dose to 95% of the tumor volume. 73
4.3 DVH Metrics Clinical Protocol (% Tx Dose) for Prostate Cases. Tx Dose = Prescribed dose to 95% of the tumor volume. 74
4.4 Breast 1. Replicating Library Plans via Interpolation of Dwell Times: Max Absolute Error (%) of PTV D15cc. Note that most errors are below 1%. . 90
4.5 Breast 2. Replicating Library Plans via Interpolation of Dwell Times: Mean Error (%) of PTV D15cc. Note that most errors are below 1%. 91

4.6 Breast 2. Replicating Library Plans via Interpolation of Dwell Times: Max Error (%) of PTV D15cc (No iterative truncation). 91
4.7 Breast 1. Percent of plans that led to improved truncation accuracy over 3 truncation iterations (100% means for all instances of feasible ϵ, truncation accuracy improved from one iteration to the next iteration.) 97
4.8 Breast 2. Percent of plans that led to improved truncation accuracy over 3 truncation iterations (100% means for all instances of feasible ϵ, truncation accuracy improved from one iteration to the next iteration. See Figure 4.10 for trend in instances that improve in true positives. 97

LIST OF ABBREVIATIONS

RT Radiation therapy/radiotherapy

TP treatment planning

IMRT intensity modulated radiation therapy

SBRT stereotactic body radiation therapy

CT computed tomography

MR magnetic resonance

TPS treatment planning software

PTV planning target volume

ART adaptive radiation therapy

OAR organ at risk

BED biologically effective dose

IGRT image-guided radiotherapy

PMF probability mass function

CBCT cone beam CT

pCT planning CT

BT brachytherapy

GTV gross tumor volume

CTV clinical target volume

HDR high dose-rate

CS critical structure

GUI guided user interface

QA quality assurance

CVaR conditional value-at-risk

TCVaR truncated conditional value-at-risk

VaR value-at-risk

pCT planning CT

DOG day-of geometry

ABSTRACT

Radiation therapy is one of the most common and effective methods of treating cancer. There are two main types: external and internal. External to the patient, a linear accelerator aims beams of radiation toward the patient; internal to the patient, radioactive sources are placed temporarily or permanently at the treatment site to deposit dose locally. Both methods of treatment can deliver complex dose distributions to a patient. The radiation damages both tumorous tissue and nearby healthy organs; treatment planning optimization determines how to deliver a dose distribution that maximizes tumor kill while sparing nearby healthy organs as much as possible. This thesis studies three treatment planning problems: the first two are in the context of external radiation therapy and the third is in the context of internal radiation therapy.

Conventional planning is based on only the physical geometry of the patient anatomy. In chapter II, we propose two models that incorporate (additional) liver function information for planning liver cancer treatment to preserve as much post-treatment liver function as possible and compare this to a conventional approach that ignores liver function information.

Conventional plans assume the patient geometry does not change between the time of patient imaging and later treatments. Although the patient geometry can be updated at treatment for plan adaptation, current practice may lead to plans that result in significantly worse quality than originally intended due to its myopic nature. In chapter III, we propose a model that produces a plan that caters to each potential patient geometry while considering both day-of and cumulative impact.

In high-dose rate brachytherapy, the patient undergoes anesthesia due to the need to implant catheters for radiation source placement before planning. Consideration of multi-

ple conflicting criteria in treatment planning results in challenging optimization problems. Current commercial systems require iterative guess-and-checking of optimization input parameters to make trade-offs among criteria, but a plan must be finalized quickly to minimize anesthesia administration. In chapter IV, we develop a practical optimization engine that generates a trade-off surface and feeds into a graphical user interface that provides the clinician more control to make trade-offs without trial-and-error optimizations.

CHAPTER I

Introduction

1.1 Radiation Therapy

Radiation therapy/radiotherapy (RT) is one of the main methods of treating cancer. Radiation can be delivered to kill cancer cells from sources external or internal to the patient. External beam radiation therapy is a non-invasive procedure that typically involves a linear accelerator directing beams of radiation from a gantry aimed at the tumor site (step 4, Figure 1.1). In our discussion of external beam radiation therapy we will focus on intensity modulated radiation therapy (IMRT), which allows modulation of beam intensities to deliver a complex so-called *dose distribution*. Internal radiation therapy, also called *brachytherapy (BT)* (*brachys* means "short-distance" in Greek), is an invasive procedure that involves a radiation oncologist implanting a small radioactive source at the treatment site. We will focus on a specific version called high-dose rate high dose-rate (HDR) BT, with the oncologist first implanting a grid of catheters in the patient, and then using an afterloader that sends radioactive seeds through the catheters throughout the tumor cite so that a radiation dose is deposited in nearby tissue. Each modality has its own advantages: external beam RT is non-invasive, while brachytherapy delivers a more localized dose, harming fewer regions of healthy organs.* Radiation dose is often delivered in multiple treatments called *fractions*

*It is also possible that radiotherapy is combined with other therapies (e.g., surgery, chemotherapy, other radiotherapy).

to maximize biological effect: with time between fractions, healthy tissue typically recovers more quickly than cancerous tissue. Inevitably, in each fraction, dose is delivered to both healthy and cancerous tissue. In the next sections, we describe RT process flow and the role of treatment planning in each form of radiotherapy.

1.1.1 Intensity Modulated Radiation Therapy (IMRT)

Once the oncologist decides that IMRT is an appropriate course of treatment, the patient first comes in for imaging (e.g., computed tomography (CT) or magnetic resonance (MR) imaging) to enable construction of a 3D model of the patient geometry in the treatment planning software (TPS) before the first fraction. In the 3D model of the patient, a physician then identifies *structures* by manually contouring them. Structures can be categorized into two groups: each healthy non-tumorous volume is called a *critical structure (CS)* or an *organ at risk (OAR)*, and tumorous volumes. The *gross tumor volume (GTV)* is the visible tumor in the image; the *clinical target volume (CTV)* is the GTV with an added margin to account for disease spread that cannot be captured in the image; the *planning target volume (PTV)* is the CTV with another margin to account for geometric uncertainties in planning and/or treatment delivery. We will elaborate on these uncertainties later.

Using this patient model, treatment is planned to determine a fluence map (non-uniform intensity profiles for each beam, gray blocks in green grid in Figure 1.5 (left)) that delivers a dose distribution in the patient. These intensity profiles then go through a process called leaf sequencing (e.g., see *Langer et al.*, 2001; *Siebers et al.*, 2002; *Xia and Verhey*, 1998) in order to be translated into a set of deliverable apertures (beam shapes molded by the multileaf collimator, Figure 1.2); the intensity of an aperture is proportional to beam-on time of the linear accelerator while the aperture is exposed. Figure 1.1 illustrates each step of this process. A conventional treatment schedule consists of 30 daily fractions (skipping weekends); stereotactic body radiation therapy (SBRT) delivers dose in up to 5 larger daily fractions. Chapters II and III will assume leaf sequencing is part of a post-processing step

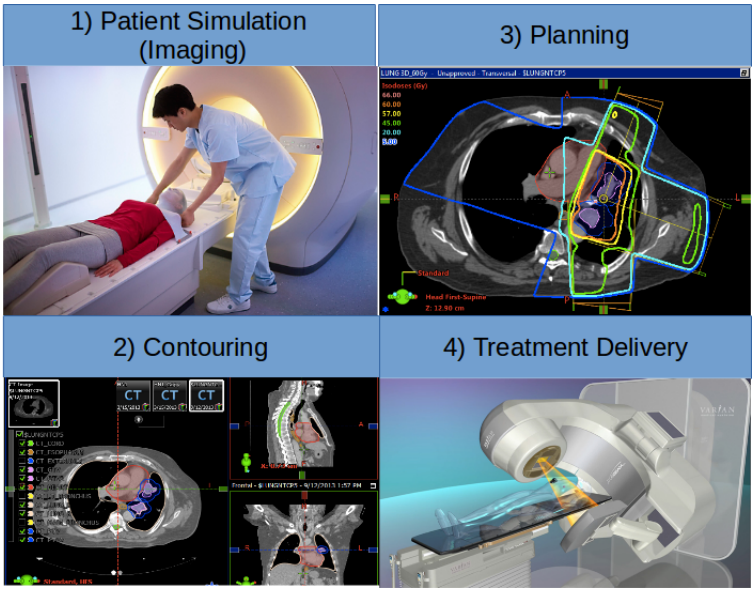


Figure 1.1: External beam RT treatment flow.[a] 1) Patient is imaged via CT scan; 2) Physician manually contours important structures on 2D slices of the 3D image; 3) Treatment plan optimization is performed (shown in the picture are 2D slices depicting patient's anatomy and dose distribution); 4) Treatment is delivered using a linear accelerator that emits beams of radiation through a gantry.

[a]Images from Varian Medical Systems, Palo Alto, CA

and focus only on fluence map optimization.

1.1.2 High-Dose Rate Brachytherapy (HDR-BT)

Once HDR-BT is chosen as the treatment modality, the patient comes in for a fraction and goes under anesthesia for a radiation oncologist to implant catheters, using ultrasound guidance, to the treatment site. Once catheters are implanted, imaging is performed to determine patient geometry and construct a patient model as well as to identify potential dwell locations for the seeds along the catheters and determine how long a seed dwells at various locations. Treatment is planned, and an afterloader (Figure 1.4) enters the room to deliver the radioactive seeds (to avoid extraneous human interaction) into the catheters (e.g., see Figure 1.3 for prostate treatment setup). Chapter IV will focus on this treatment modality.

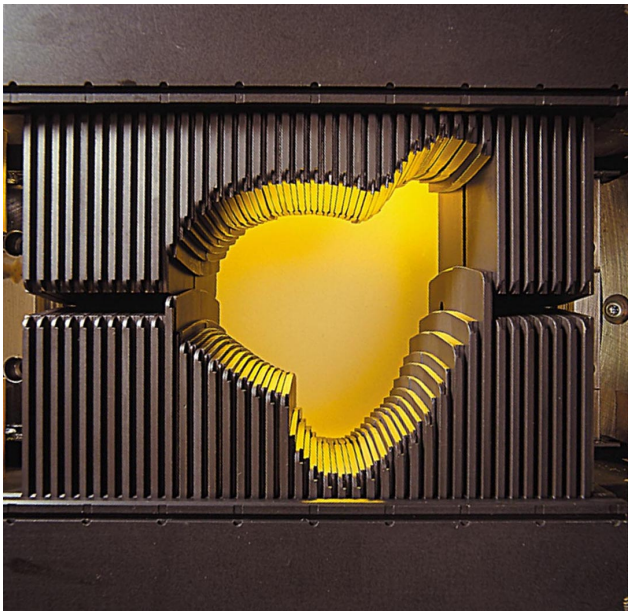


Figure 1.2: Multileaf collimator composed of tungsten leaves that mold the beam shape.[a]

[a]Varian Medical Systems, Palo Alto, CA

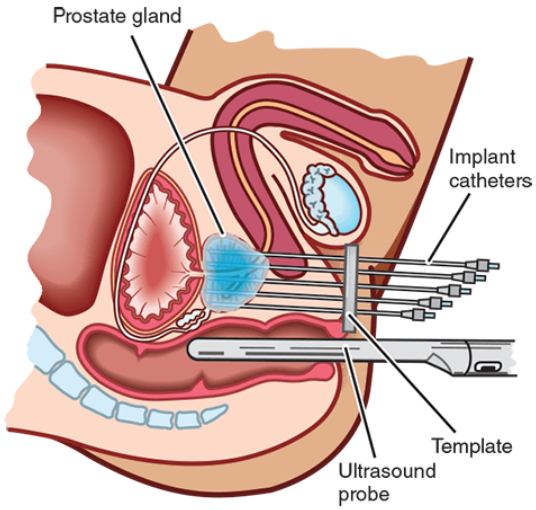


Figure 1.3: BT treatment delivery for prostate. Imaging, contouring, and planning are similar to IMRT[a]

[a]www.prostate.org.au

Figure 1.4: An afterloader delivers radioactive seeds into the catheters.[a]

[a]elekta.com

1.1.3 Treatment Plan Optimization

Treatment plan optimization uses mathematically-expressed metrics to balance the goals of eradicating the tumor and sparing the healthy organs. The output (i.e., optimized de-

cisions) of these models are machine instructions used to deliver treatment. For a more detailed overview of treatment planning in IMRT, see *Aleman et al.* (2013); *Bortfeld* (2006); *Romeijn and Dempsey* (2008). In the models used for the different treatment modalities, the reader will observe parallels in the relationships between treatment decisions and the dose delivered. We will introduce notation and models individually in each chapter.

Next, we provide an overview of current challenges in treatment planning.

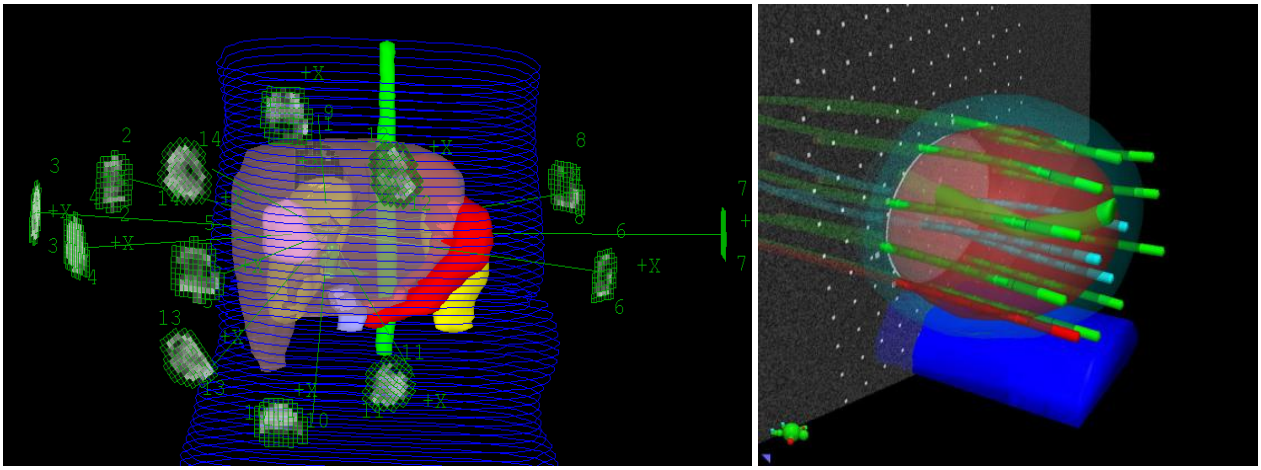


Figure 1.5: Treatment plan for IMRT (left): blue and other colors inside represent patient geometry (tumor in pink); green blocks represent beam angles; gray corresponds to beam intensity profile. Treatment plan for BT (right): The prostate and tumor are depicted in red; green and blue lines represent catheters and alternating shades represent seed dwell locations.

1.1.3.1 Challenges

There are many clinical challenges throughout the process of radiotherapy; often, when trying to account for these challenges in treatment planning, mathematical challenges arise in the treatment planning model.

Treatment is conventionally planned around the geometry of the patient anatomy, limiting planning to only physical features of the patient. With advances in biomarkers and imaging, organ function information, in addition to the physical geometry of organs, can be captured (*Ling et al.*, 2000) and used in treatment planning. Ultimately, the goal in reducing dose to OARs is to preserve organ function. A biological model that captures changes in

organ function with respect to dose is called a "dose-response model." The model is usually a (mathematical) function that predicts how (organ) function will change if a certain amount of dose is delivered, which can be used in treatment planning to maximize post-treatment organ function. Radiosensitivity (sensitivity to radiation) (*Fertil and Malaise*, 1981; *Long*, 2015; *Wang et al.*, 2016) can also impact the efficacy of treatment. Current practice does not explicitly account for such biological information in treatment planning; incorporating such information can lead to improved outcomes (measured by the biological model) but result in challenges for optimization, e.g., using non-convex metrics based on biological models to measure treatment quality.

Although there are many sources of uncertainty in treatment planning, we list a few examples: 1) interfraction motion, i.e., changes in patient geometry between fractions; 2) intrafraction motion, i.e., motion during a treatment, e.g., from breathing; and 3) set-up errors at treatment, i.e., the patient's position relative to the linear accelerator. These uncertainties cause errors (expected versus delivered) in the dose distribution (e.g., see *Stroom and Heijmen*, 2002; *Van Herk*, 2004 for an overview on how tumor margins are drawn in the structure delineation process and *Langen and Jones*, 2001 for general organ motion management to avoid some of these uncertainties). Two major challenges arise for these uncertainties: 1) changes in geometry mean the original plan will not treat the patient as expected, potentially affecting treatment outcome negatively; 2) re-imaging and re-planning on the day of the treatment, as well as performing plan quality assurance (QA), is operationally very challenging due to increased patient wait time, so alternative quicker strategies for adapting treatment are needed (keeping in mind both immediate and cumulative consequences).

During treatment planning, there are several competing dosimetric goals set by clinical protocol (e.g., tumor coverage versus chest sparing for a breast site) where tradeoffs must be made, especially if not all goals can be met simultaneously. Because many clinically relevant metrics are difficult to optimize directly, approximations are made in the model, and treatment planning becomes an iterative process, alternating between choosing optimization

model parameters and optimizing the resulting instance to obtain a modified plan. However, changing these parameters does not provide direct control of change in the resulting plans' metrics. This unpredictability results in a time-consuming trial-and-error process and can lead to loss in treatment plan quality. Several frameworks based on multicriteria optimization have been proposed to generate trade-off surfaces consisting of many plans to choose from, allowing more control over candidate plan generation. For an overview, we refer the reader to *Küfer et al.* (2009).

1.1.4 Chapter Summaries

1.1.4.1 Optimizing Global Liver Function in Radiation Therapy Treatment Planning, Chapter II

Typically, planning is done only based on the dose delivered to various structures and the geometry of patient anatomy (obtained from the planning CT). In each structure, the planner treats each voxel with the same level of importance and there is no distinguishing of voxels. However, with recent advances in imaging and use of biomarkers, additional information about a critical structure can be obtained. For example, functional imaging (e.g., via contrast injection and dynamic contract-enhanced MRI) can reflect local organ function and help prioritize which portions of the structure ought to be spared to preserve overall (global) organ function. We propose two novel treatment planning models of varying complexity that incorporate liver function information and liver dose-responses, and compare the achieved post-treatment global liver function with what is achieved by a conventional treatment planning (TP) model that ignores liver function information.

1.1.4.2 Adaptive Liver Stereotactic Body Radiotherapy, Chapter III

Treatment planning is typically done once, before any treatment begins, based on pre-treatment imaging data, and the plan is divided into equal fractions, assuming the patient comes to the clinic with the same geometry as the one that was imaged. However, inter-

fraction motion is a source of uncertainty and can drastically affect treatment efficacy if the geometry on a particular treatment day differs from what was expected: critical structures may shift or deform and be in the high dose region, receiving much more dose than originally intended or even exceeding allowable dose limits. In this chapter, we propose a multistage stochastic program that anticipates potential geometries to adapt treatment plans to the "geometry of the day" to maximize overall PTV coverage while satisfying dose limits (cumulative and at each fraction) no matter which geometry arises. Geometries can be considered *favorable* (typical PTV coverage or better can be achieved) or *unfavorable* (at best, the PTV is under-dosed). This RTTP model allows for compensating for PTV under-dosing in unfavorable geometries when favorable geometries arise. Operationally, this model produces a plan for each potential geometry at each fraction to give the clinician flexibility in adapting treatment; this is in contrast to previous works that produce a plan with respect to geometry-averaged model parameters, i.e., one plan that will be delivered no matter which geometry is realized.

1.1.4.3 Multicriteria Optimization for Brachytherapy Treatment Planning, Chapter IV

Unlike external beam RTTP, HDR BTTP treatment planning is done when the patient is present for treatment — recall that the patient is under anesthesia. This is because imaging is done *after* the patient arrives and catheters are implanted. Only then can we identify dwell locations. The operational pressure is to plan treatment by exploring allowable criteria tradeoffs quickly. Delays can lead to increased anesthesia use, patient discomfort, and treatment inaccuracy, influencing treatment outcome and toxicity. Current commercial planning systems require iterative guessing-and-checking of optimization parameters for re-optimization to make adjustments to the dwell times. This sequential generation of treatment plans and lack of control in making dosimetric tradeoffs (e.g., fixing a particular metric to be constant while making a tradeoff among other metrics) is often inefficient and can

unintentionally exclude candidate plans that would have been more desirable. This chapter addresses two questions:

1. Given a library of high-quality treatment plans for the current patient, how can we use this information to estimate a trade-off surface? Moreover, how can a physician navigate this trade-off surface in an intuitive and efficient manner, preferably not needing to interact with any parts of the process that require optimization?

2. With the patient waiting under anesthesia, how can we efficiently generate such a library of high-quality plans when the metrics considered are non-convex dose-volume metrics?

Although these two questions can be considered independently, we present them together in one chapter, where the overarching goal is developing an "optimization engine" that generates a library of high quality plans that form a foundation for a guided user interface (GUI) that the physician can use in the clinic to aid them in selecting a treatment plan for each patient.

CHAPTER II

Optimizing Global Liver Function

2.1 Introduction

Stereotactic body radiation therapy (SBRT) has become a popular method of treatment for liver cancer for its efficacy in local tumor control (*Liu et al.*, 2013). SBRT is generally used to treat small tumors (approximately 5 cm or less in diameter) due to the risk of normal tissue toxicity when treating larger volumes to high doses. High dose from treatment, patients' sensitivity to radiation, and their pre-treatment liver function all affect the likelihood of radiation-induced liver disease (RILD). The latter two factors are patient attributes that are not reflected in computed tomography (CT) scans typically used to characterize patient geometry for treatment planning. In this chapter, we focus on incorporating pre-treatment liver function into treatment planning decisions, motivated by the following considerations. The liver is well-known to function in a parallel-like fashion (*Jackson et al.*, 1995), e.g., a certain portion of the liver may be removed or damaged without losing overall function. However, local liver function is not uniform, and initial function as well as *which* functional portions are damaged influences the post-treatment global liver function. This implies the need to take into consideration during treatment planning 1) spatial features of a dose distribution, which would prioritize sparing of high-functioning portions of the liver, and 2) liver tissue dose response.

There has been ongoing work exploring use of physiological imaging to consider organ

function and/or tumor aggression information in treatment planning, see, e.g., brain target radiosensitivity from dynamic susceptibility contrast-enhanced MRI (*Chen et al.*, 2007), lung target proliferation rate from fluorine-18-fluorodeoxyglucose PET (*Das et al.*, 2004; *Feng et al.*, 2009), liver function from ^{99m}Tc-sulphur colloid (SC) SPECT/CT (*Bowen et al.*, 2015), and lung function from various imaging modalities (*Ireland et al.*, 2007; *Seppenwoolde et al.*, 2002). *Bowen et al.* (2015) segmented high-functioning liver into regions binned by intervals of SC uptake and redistributed dose by scaling the mean dose of each of these regions according to SC uptake in the objective function. *Ireland et al.* (2007) segmented functional lung using ^{3}He MRI and compared treatment plans from minimizing the volume receiving at least 20 Gy in the total lung vs. total functional lung. *Seppenwoolde et al.* (2002) showed potential of using voxel-based cost functions and classifying patients by perfusion pattern. Two advantages of using voxel-based information are the ability to view the relevant organ approximately as a continuous body, as opposed to (manually) segmenting it into disjoint functional and nonfunctional structures (e.g., as in *Bowen et al.*, 2015; *Ireland et al.*, 2007), and to use functional imaging information directly, which is operationally better suited for adaptive planning. In these previous works, all approaches used surrogates for post-treatment organ function and did not consider organ dose-response behavior in the optimization. *Zhang et al.* (2010) examined lung dose response and observed that post-treatment reduction in perfusion plateaus beyond a particular dose threshold, i.e., damage becomes saturated; the liver dose-response model from *Wang et al.* (2016) reflects similar behavior.

In this chapter, we propose two new treatment planning approaches aimed at better preserving liver function without compromising tumor coverage or exceeding acceptable limits to other critical structures. In particular, we propose replacing the traditional objective of minimizing linearized equivalent uniform dose (ℓEUD) to the liver with ones that account for liver function: 1) perfusion-weighted ℓEUD (fEUD) that avoids delivering dose to highly-perfused liver by explicitly incorporating voxel-based pre-treatment liver perfusion into a treatment planning model, resulting in a problem that is convex and can be efficiently

solved to global optimality, and 2) post-treatment global liver function (GLF) that explicitly captures global liver function using a model of liver dose-response based on pre- and post-treatment perfusion, resulting in a problem that is nonconvex and can be solved to local optimality. To measure liver function, our study uses voxel-based pre-treatment liver perfusion, computed from Dynamic-Contrast Enhanced MRI (DCE-MRI). *Cao et al.* (2013) have shown portal venous perfusion to be a good indicator of both local and global liver function.

The rest of this chapter is organized as follows. In Section 2.2, we formalize the three optimization models used, and describe our experimental setting for analyzing these models, including both 2D (sythensized data) and 3D (real patient data) instances that we used to test these models, in Section 2.3. In Section 2.4, we present and compare resulting dose distributions obtained from each of the models. In Section 2.5, we show how the fEUD and GLF models produce alternative dose distributions to the ones obtained by the ℓEUD model, and in particular show the potential benefits of each alternative dose distribution. We conclude the chapter in Section 2.6.

2.2 Mathematical Optimization Models for Liver SBRT Treatment Planning

2.2.1 Notation and Treatment Criteria

We discretize the relevant areas of the patient anatomy into a finite set of voxels V, and discretize the beams, whose orientation is chosen a priori by a dosimetrist, into a finite set of beamlets N. Let S be the set of structures, V_s be the set of voxels in structure $s \in S$, with $V = \cup_{s \in S} V_s$, and $D \in \mathbf{R}^{|V| \times |N|}$ be the dose deposition matrix, where an entry D_{ji} is the dose deposited in voxel $j \in V$ from beamlet $i \in N$ at unit intensity. In our treatment planning model, the decisions include $x_i,\ i \in N$, which is the intensity of beamlet i, and $z_j,\ j \in V$,

which is the (physical) dose delivered to voxel j computed as

$$z_j = \sum_{i \in N} D_{ji} x_i, \ j \in V.$$

Given dose distribution $z \in \mathbf{R}^{|V|}$, the generalized equivalent uniform dose (gEUD) (*Niemierko*, 1999), a biological criterion, can be used to evaluate a dose distribution to a structure $s \in S$ with voxel set V_s. The gEUD to a structure $s \in S$ is defined as

$$\text{gEUD}_s(z) = \left(\frac{1}{|V_s|} \sum_{j \in V_s} z_j^{a_s} \right)^{\frac{1}{a_s}}, \tag{2.1}$$

where a_s is a structure-specific parameter (*Li et al.*, 2012). For computational efficiency, we use a piecewise-linear approximation of the gEUD, namely, the linearized equivalent uniform dose (ℓEUD) (*Thieke et al.*, 2002), which is a convex combination of the mean and maximum dose, or of the mean and minimum dose, for structures where we are concerned with overdosing or underdosing, respectively:

$$\ell\text{EUD}_s^+(z) = \alpha_s^+ \frac{1}{|V_s|} \sum_{j \in V_s} z_j + (1 - \alpha_s^+) \max_{j \in V_s} z_j \qquad (\ell\text{EUD-overdose})$$

$$\ell\text{EUD}_s^-(z) = \alpha_s^- \frac{1}{|V_s|} \sum_{j \in V_s} z_j + (1 - \alpha_s^-) \min_{j \in V_s} z_j, \qquad (\ell\text{EUD-underdose})$$

where $\alpha_s^\pm \in [0, 1]$ is also structure-specific. We note that, in optimization models, typically $\ell\text{EUD}_s^+(z)$ is bounded from above to avoid hot spots (e.g., if s is a critical structure) and $\ell\text{EUD}_s^-(z)$ is bounded from below to ensure sufficient coverage and avoid any cold spots (e.g., if s is a target). From an optimization perspective, both of these ℓEUD-bounding constraints are convex, and therefore tractable.

2.2.2 Optimization Models

Using the treatment criteria discussed above, our treatment planning optimization model minimizes an objective function based on dose, subject to dose constraints for various structures. The model is as follows:

$$\underset{x,z}{\text{minimize}} \quad <\text{Objective Function of } z> \tag{2.2}$$

$$\text{subject to} \quad \ell\text{EUD}^{-}_{\text{PTV}}(z) \geq \overline{\ell\text{EUD}}_{\text{PTV}} \tag{2.3}$$

$$\ell\text{EUD}^{+}_{s}(z) \leq \overline{\ell\text{EUD}}_{s} \qquad s \in S\backslash\{\text{Liver, PTV}\}, \tag{2.4}$$

$$z_j = \sum_{i \in N} D_{ji} x_i \qquad j \in V, \tag{2.5}$$

$$0 \leq x_i \leq U \qquad i \in N. \tag{2.6}$$

Here, in (2.3), ℓEUD to the planning target volume (PTV), denoted $\ell\text{EUD}^{-}_{\text{PTV}}(z)$, is bounded below by the parameter $\overline{\ell\text{EUD}}_{\text{PTV}}$, which is typically the prescription dose. In our experiments, we only consider cases with one target, but more constraints of this type can be added to represent multiple targets. In (2.4), the parameters $\overline{\ell\text{EUD}}_{s}$ denote upper bounds on $\ell\text{EUD}^{+}_{s}(z)$ to the remaining critical structures $s \in S\backslash\{\text{Liver, PTV}\}$. Constraints in (2.5) link dose and beamlet intensity variables. Inequalities in (2.6) provide lower and upper bounds on beamlet intensities. Solutions to these models are treatment plans consisting of beamlet intensities x that deliver resulting dose distributions z, which together satisfy constraints (2.3)–(2.6). The function being optimized in (2.2) reflects an objective related to the liver dose; the standard approach is to minimize $\ell\text{EUD}^{+}_{\text{Liver}}(z)$ — we will refer to the model with this objective as the *ℓEUD model*. Note that in the ℓEUD model, the objective function gives uniform weight to the dose to each voxel in the liver, i.e., it considers every voxel as equally important to spare. In the following sections, we propose two alternative liver objectives: the first of these objectives reflects relative importance of liver voxels by considering their relative perfusion level to discourage dose to highly-perfused liver; the other objective accounts for

voxel-based dose response that considers "damage-resistant" and "damage-saturated" dose thresholds (global liver response is then the sum of appropriate terms over voxels).

2.2.3 Perfusion Avoidance Model

As previously mentioned, we can use liver perfusion maps computed from DCE-MRI as an indicator of local and global liver function (*Cao et al.*, 2013). We denote the perfusion vector for the liver by $f \in \mathbf{R}_+^{|V_{\text{Liver}}|}$, where the j^{th} component of f is the perfusion level of voxel $j \in V_{\text{Liver}}$. Liver perfusion values range from $0 \frac{\text{milliliters}}{\text{100 gram minute}}$ $\left(\text{abbreviated as } \frac{\text{mL}}{\text{100 g min}}\right)$ to over $100 \frac{\text{mL}}{\text{100 g min}}$. In this chapter, we assume there is no uncertainty in these measurements, e.g., due to noise or image registration. Letting $\mathbf{g} : \mathbf{R}_+^{|V_{\text{Liver}}|} \to \mathbf{R}_+^{|V_{\text{Liver}}|}$ be a vector-valued function $\mathbf{g}(f) = (g_1(f), \ldots, g_{|V_{\text{Liver}}|}(f))$, we define each component $g_j(f),\ j \in V_{\text{Liver}}$, as the *relative* importance of voxel j among all liver voxels, given perfusion vector f. In this work, we consider $g_j(\cdot)$ that only depend on f_j, and we require each $g_j(\cdot)$ to be non-decreasing in f_j. This gives greater importance to highly-perfused voxels, which we most want to preserve. In our selection of $\mathbf{g}(\cdot)$, we reflect the findings of previous studies (*Cao et al.*, 2008; *Wang et al.*, 2016) that there exists a perfusion threshold below which voxels do not have any functional value. Although there is not a consensus on the value of this threshold, in the following we use $30 \frac{\text{mL}}{\text{100 g min}}$ as a compromise between suggested values. Similarly, we assume that voxels with values measured over $100 \frac{\text{mL}}{\text{100 g min}}$ do not have higher function than those at $100 \frac{\text{mL}}{\text{100 g min}}$, as in *Cao et al.* (2006) and *Pandharipande et al.* (2005). Thus, values are rounded accordingly and the effective range of perfusion considered is $[30, 100]$. These values are then shifted down to $[0, 70]$ to further distinguish voxels with higher function from ones with lower function. Based on these considerations, we use

$$g_j(f) = \min\{100, \max\{30, f_j\}\} - 30, \quad j \in V_{\text{Liver}} \tag{2.7}$$

in our experiments. To incorporate perfusion information into treatment planning, we use a so-called "perfusion-weighted ℓEUD" (denoted fEUD), which is similar to the functional EUD described, e.g., by *Miften et al.* (2004) and used by *Seppenwoolde et al.* (2002) to reduce dose to high-functioning lung. We define fEUD to the liver by weighting the dose to each liver voxel by the voxel's relative importance:

$$\text{fEUD}^{+}_{\text{Liver}}(z;\mathbf{g}(f)) = \alpha^{+}_{\text{Liver}} \frac{1}{|V_{\text{Liver}}|} \sum_{j \in V_{\text{Liver}}} g_j(f) z_j + (1-\alpha^{+}_{\text{Liver}}) \max_{j \in V_{\text{Liver}}} g_j(f) z_j \tag{2.8}$$

(note: $\text{fEUD}^{+}_{\text{Liver}}(z;\mathbf{1})$=$\ell\text{EUD}^{+}_{\text{Liver}}(z)$). We refer to the optimization model (2.2)–(2.6) with $\text{fEUD}^{+}_{\text{Liver}}(z;\mathbf{g}(f))$ used as the objective in (2.2) as the *perfusion avoidance model*, or simply the *fEUD model*. The fEUD model is similar to the approach of *Bowen et al.* (2015) of differentiating functional liver regions by intervals of $^{99\text{m}}$Tc-sulphur colloid SPECT uptake and then minimizing a weighted sum of mean doses to each of the differentiated regions. However, we do not segment the liver by perfusion ranges and instead differentiate function at the voxel level.

Although we also introduce a liver dose-response model to optimize post-treatment global liver function in the next section, we include $\text{fEUD}^{+}_{\text{Liver}}(z,\mathbf{g}(f))$ in our experiments since it essentially reflects preference for delivering less dose to higher-perfused voxels, in order to show the potential benefits of a perfusion-conscious model. Moreover, because of the convex piece-wise linear structure of this objective function, the resulting optimization model is linear and can be solved much more efficiently than the more complex global liver function model.

2.2.4 Global Liver Function Preservation Model

The perfusion avoidance model described in the previous section may not fully capture characteristics of normal liver dose response. We thus propose an alternative optimization model, with an objective function designed to reflect post-treatment global liver function.

The model was developed using liver perfusion dose-response data from an Institutional Review Board-approved study. The study consisted of 24 patients who were treated with variable fractionation schemes and imaged using DCE-MRI at the following time points: pre-treatment, after 60% of planned treatment, and 1 month post-treatment. In the analysis, physical dose was converted to EQD2 (equivalent dose in 2-Gy fractions) to account for the variable fractionation.

Our model of perfusion-based post-treatment global liver function has two components. The first is a perfusion-based model of global liver function. According to *Wang et al.* (2016), given perfusion value f_j at voxel j, the corresponding so-called *probability of function,* p, for this voxel is computed as

$$p(f_j; F_{0.5}, n) = \frac{1}{1 + (\frac{F_{0.5}}{f_j})^n}, \tag{2.9}$$

where $F_{0.5}$ is the perfusion value at which the voxel functions with a likelihood of 50%, and n is a steepness parameter. Due to the nature of the logistic function, voxels with poor/high perfusion have similarly poor/high function probability, implying two perfusion thresholds: one below which all voxels have no function, and another above which all voxels are simply considered fully functioning, a behavior consistent with the weights used in the fEUD objective. Given the probability of function for each liver voxel, global liver function (assuming a parallel structure) can be computed as follows (see *Wang et al.*, 2016):

$$\bar{P}(f; F_{0.5}, n, \gamma) = \frac{1}{|\tilde{V}_{\text{Liver}}|} \sum_{\substack{j \in V_{\text{Liver}}: \\ p(f_j) > \gamma}} p(f_j) = \frac{1}{|\tilde{V}_{\text{Liver}}|} \sum_{\substack{j \in V_{\text{Liver}}: \\ p(f_j) > \gamma}} \frac{1}{1 + (\frac{F_{0.5}}{f_j})^n}, \tag{2.10}$$

where γ is the probability threshold under which voxels do not contribute to global liver function and $\tilde{V}_{\text{Liver}} = \{j \in V_{\text{Liver}} : p(f_j) > \gamma\}$.

The second component of our model captures the impact of dose on liver voxel perfusion. In the following, we use superscripts "pre" and "post" to denote pre- and post-treatment

measurements. Previous lung studies indicate a potential dose threshold beyond which damage is saturated, i.e., no more function is lost (*Zhang et al.*, 2010). To consistently reflect these thresholds for our liver response model, we assume a logistic form. Let $f^{\text{post}} : \mathbf{R}_+^{|V_{\text{Liver}}|} \to \mathbf{R}_+^{|V_{\text{Liver}}|}$ be a vector-valued function, with $f_j^{\text{post}}(\cdot)$ given by:

$$f_j^{\text{post}}(z_j) = \left(\frac{D_{50}^k}{D_{50}^k + \text{EQD2}(z_j)^k} \right) f_j^{\text{pre}}, \quad j \in V_{\text{Liver}}, \tag{2.11}$$

where D_{50} is the EQD2 dose required to reduce initial perfusion by 50%, k determines the steepness of the curve, and

$$\text{EQD2}(z_j) = \frac{z_j \left(\left(\frac{\alpha}{\beta} \right)_s + \frac{z_j}{T} \right)}{\left(\frac{\alpha}{\beta} \right)_s + 2}, \quad j \in V_s,\ s \in S, \tag{2.12}$$

where $\left(\frac{\alpha}{\beta} \right)_s$ is the alpha-beta ratio of structure s, and T is the total number of fractions. The parameters of this model were found from logistic fitting from the study's patient population.

Combining (2.10) and (2.11), we can express post-treatment global liver function in terms of EQD2 and pre-treatment perfusion:

$$\frac{1}{|\tilde{V}_{\text{Liver}}^{\text{post}}|} \sum_{j \in \tilde{V}_{\text{Liver}}^{\text{post}}} \left(1 + \left(\frac{F_{0.5}}{f_j^{\text{pre}}} \left(\frac{D_{50}^k + (\text{EQD2}(z_j))^k}{D_{50}^k} \right) \right)^n \right)^{-1}, \tag{2.13}$$

and, substituting (2.11), we obtain a model for post-treatment global liver function $\text{GLF}^{\text{post}}(z; f^{\text{pre}})$ as a function of physical dose z and pre-treatment perfusion f^{pre}:

$$\begin{aligned} \text{GLF}^{\text{post}}(z; f^{\text{pre}}) &= \bar{P}(f^{\text{post}}(z); F_{0.5}, n, \gamma)) \\ &= \frac{1}{|\tilde{V}_{\text{Liver}}^{\text{post}}|} \sum_{j \in \tilde{V}_{\text{Liver}}^{\text{post}}} \left(1 + \left(\frac{F_{0.5}}{f_j^{\text{pre}}} \left(\frac{D_{50}^k + \left(\frac{z_j \left(\left(\frac{\alpha}{\beta} \right)_{\text{Liver}} + \frac{z_j}{T} \right)}{\left(\frac{\alpha}{\beta} \right)_{\text{Liver}} + 2} \right)^k}{D_{50}^k} \right) \right)^n \right)^{-1}. \end{aligned} \tag{2.14}$$

Each term in the above summation (corresponding to voxel j) represents the post-treatment

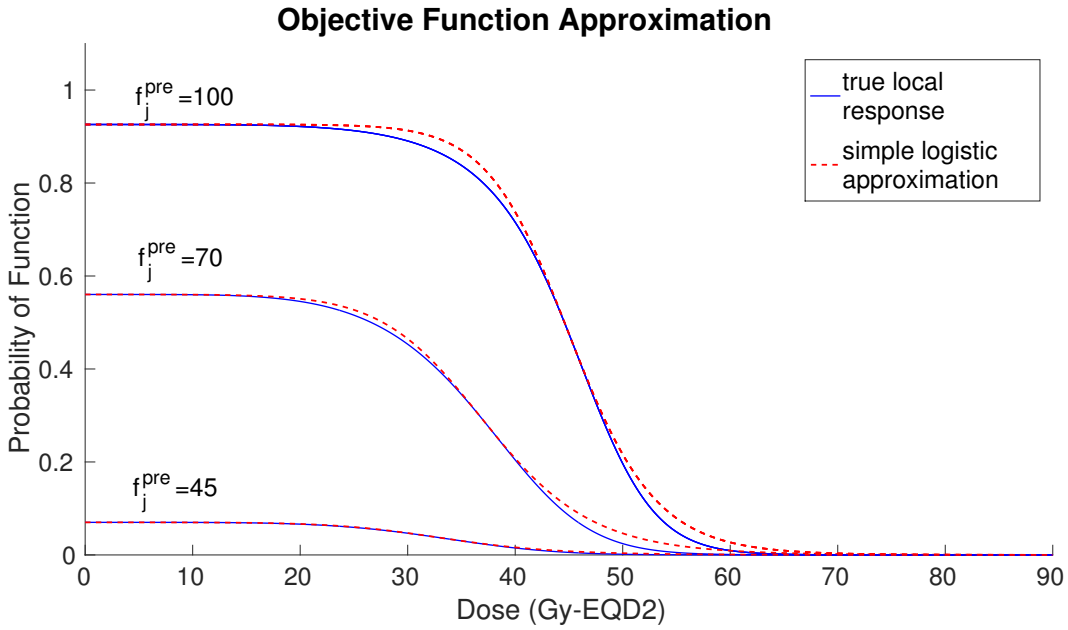


Figure 2.1: Comparison between examples of true local response (blue) and simple logistic approximation (red) varied by pre-treatment perfusion levels.

probability of function given dose z_j and pre-treatment perfusion f_j^{pre}.

We illustrate the shape of several terms in the sum (2.13), which reflect probability of function, in terms of EQD2, for voxels with various pre-treatment perfusion levels (solid blue curves), in Figure 2.1. There are two important dose thresholds concerning post-treatment function: below the *dose-damage threshold* no significant function is lost, and above the *dose-saturation threshold* no additional function is lost. The terms in the sum (2.14), if plotted with physical dose as the independent variable instead of EQD2, have similar S-shapes but are steeper due to the different scaling of the independent axis. For implementation purposes, we approximate each term of (2.14) by a simpler logistic function; details of the approximation are discussed in the next section. Note that since GLF should be maximized, to fit into our optimization framework, we (equivalently) minimize its additive inverse.

2.2.4.1 A Simple Logistic Approximation of GLF

To simplify implementation, we approximate $\text{GLF}^{\text{post}}(z; f^{\text{pre}})$ with the following simpler function $\widehat{\text{GLF}^{\text{post}}}(z; f^{\text{pre}})$:

$$\widehat{\text{GLF}^{\text{post}}}(z; f^{\text{pre}}) = \frac{1}{|\tilde{V}^{\text{pre}}_{\text{Liver}}|} \sum_{j \in \tilde{V}^{\text{pre}}_{\text{Liver}}} \frac{\bar{p}^{\text{pre}}_j(f^{\text{pre}}_j)}{1 + \text{e}^{\sigma_j(f^{\text{pre}}_j)(z - \bar{z}_j(f^{\text{pre}}_j))}}, \tag{2.15}$$

where $\tilde{V}^{\text{pre}}_{\text{Liver}} = \{j \in V_{\text{Liver}} : p(f^{\text{pre}}_j) > \gamma\}$. To derive this simplification, each term in the summation (2.14) (corresponding to voxel j) is replaced by a simple logistic function

$$\frac{\bar{p}^{\text{pre}}_j(f^{\text{pre}}_j)}{1 + \text{e}^{\sigma_j(f^{\text{pre}}_j)(z - \bar{z}_j(f^{\text{pre}}_j))}},$$

where $\bar{p}^{\text{pre}}_j$ is the pre-treatment probability of function of voxel j (computed using (2.9)), $\bar{z}_j$ is the dose that results, for this voxel, in reduction of pre-treatment function by 50%, and σ_j is the slope of the corresponding term in (2.14) at $\bar{z}_j$ ($\bar{z}_j$ is also approximately the point of change in the sign of curvature in the corresponding term in (2.14)). Note that the values of these parameters are based on pre-treatment perfusion f^{pre}_j, and in particular, that the steepness parameter σ_j reflects the sensitivity of the voxel, i.e., the magnitude of its response to dose. Also, in (2.15) we include all voxels whose *pre-treatment* probability of function exceeds γ, with minimal impact on the resulting values. The dotted red curves in Figure 2.1 illustrate the approximation (with the formulae adjusted to the EQD2 dose scale).

We refer to the optimization model (2.2)–(2.6) with $-\widehat{\text{GLF}^{\text{post}}}(z; f^{\text{pre}})$ used as the objective in (2.2) as the *global liver function preservation model*, or simply the *GLF model.*

2.3 Experiments

In Sections 2.2.2, 2.2.3, and 2.2.4 we proposed an optimization model (2.2)–(2.6) with three options for objective function (2.2): $\ell\text{EUD}^{+}_{\text{Liver}}(z)$, $\text{fEUD}^{+}_{\text{Liver}}(z; \mathbf{g}(f))$ of (2.8), and $-\widehat{\text{GLF}^{\text{post}}}(z; f^{\text{pre}})$ of (2.15). To test the effectiveness of considering varying degrees of func-

tion information during optimization, we compared dose distributions obtained from these three models (we refer to these dose distributions as the ℓEUD, fEUD, and GLF plans, respectively) for the same set of treatment parameters (i.e., the bounds $\overline{\ell\text{EUD}}_s$ that appear on the right-hand sides of (2.3) and (2.4)). Thus, the set of constraints was identical in the three models and the only difference was the objective function used in (2.2). All treatment planning performed for the purpose of the current study was retrospective.

2.3.1 Model Parameters and Implementation Details

Both radiation therapy simulation imaging and perfusion data were obtained from patients enrolled in a prospective study approved by an institutional review board. Plan setup and dose coefficient (D_{ji}) calculations were done in the University of Michigan's in-house treatment planning software packages, UMPlan and UMOpt (*Kessler et al.*, 2005; *Kim et al.*, 1995).

To evaluate dose to the PTV with gEUD in (2.1), it is common to use $a_{\text{PTV}} \approx -10$ (*Li et al.*, 2012). Using the piecewise linear approximation (ℓEUD-underdose) to evaluate dose to the PTV, we set $\alpha^-_{\text{PTV}} = 0.84$ so that $\ell\text{EUD}^-_{\text{PTV}}(z^{\text{delivered}})|_{\alpha=0.84} = \text{gEUD}_{\text{PTV}}(z^{\text{delivered}})|_{a=-10}$, where $z^{\text{delivered}}$ is a treatment plan delivered to a typical liver patient in the Department of Radiation Oncology at the University of Michigan. We simply used the mean dose, i.e., $\alpha^+_{\text{Liver}} = 1$, for $\ell\text{EUD}^+_{\text{Liver}}(z)$, as liver is widely considered to be a highly parallel organ. For all other structures, the maximum dose in structure s was used for $\ell\text{EUD}^+_s(z)$, i.e., $\alpha^+_s = 0,\ s \in S\backslash\{\text{Liver, PTV}\}$. The maximum intensity of any beamlet in (2.6) was bounded by $U = 40$ (according to institutional practice) to avoid extremely high beamlet intensities that would result in high monitor units and inefficient treatment delivery. This constraint typically does not restrict solution quality as beamlet intensities rarely reach this maximum value. In (2.11), $D_{50} = 51.7$ Gy-EQD2 and $k = 4.9$, which were determined from model fitting. In (2.9), $F_{0.5} = 67.4\frac{\text{mL}}{100\text{ g min}}$ and $n = 6.4$, both of which were obtained from *Wang et al.* (2016). To compute EQD2 dose for liver, we used $\left(\frac{\alpha}{\beta}\right)_{\text{Liver}} = 2.5$ Gy and $T = 5$

fractions (according to the protocol at the University of Michigan Hospital).

We remark that, in formulating our models, we have assumed that the liver perfusion data is known and not subject to uncertainty. In practice, uncertainty is undoubtedly present, in part due to possible measurement and registration imprecision, and can have a clinical impact. While this chapter focuses on incorporating functional imaging and using true dose-response models for capturing change in organ function in treatment planning, in our future work we plan to test the impact of parameter uncertainty on treatment plans obtained with the models presented here, and extend the models to make them robust. This would involve developing appropriate analytical models of the uncertainty and incorporating them into the optimization. As our preliminary work suggests, robust versions of the models considered in this chapter will be significantly more mathematically complex, and will require development of corresponding solution methods. We plan to report on the methodology and results of robust models in future publications.

Our implementation of all three models was done in C++, with Gurobi's primal simplex method (Gurobi Optimization, Houston, TX) used to solve the linear optimization problems resulting from ℓEUD and fEUD models. The GLF model results in a nonconvex nonlinear optimization problem, which was solved using IpOpt's primal-dual interior point method (*Wächter and Biegler*, 2006) and linear solver subroutine by *HSL* (2013). All experiments were performed on a custom-built machine with 3.5 GHz 8-Core Intel i7-3770K processor and 31.4 GB memory at 3901 MHz. Although the interior point method is only guaranteed to find locally optimal solutions, it was warm-started with the fEUD plan to guarantee a treatment plan that was at least as good (in terms of GLF objective) as the fEUD plan.

2.3.2 Liver Perfusion Patterns

The patient population varies widely in terms of liver perfusion patterns. Since 3D dose distributions are difficult to visualize, analyzing synthesized 2D cases with specific patterns allows for a comprehensive illustration of where dose is effectively reduced and how

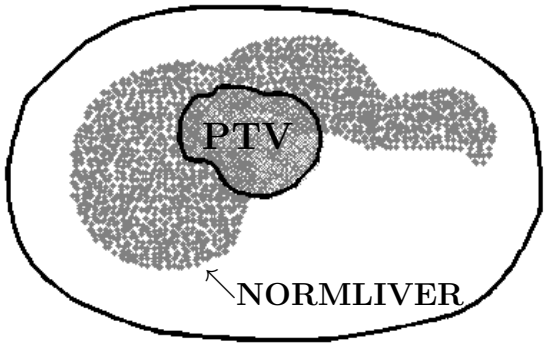
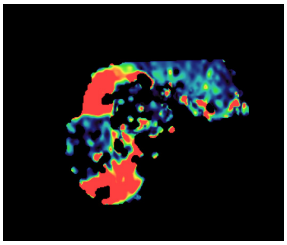
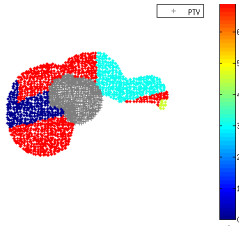


Figure 2.2: Geometry of 2D liver case (axial view). The small contour represents the boundary of the PTV voxels and is surrounded by normal liver voxels. The large contour represents the boundary of the patient and contains normal tissue voxels that are not shown.

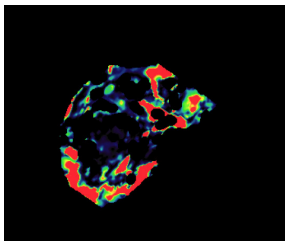
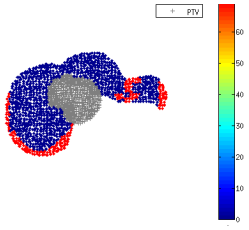
this reduction is compensated for to maintain target coverage. We identified oft-observed 3D perfusion patterns from DCE-MRI in real patient cases and, for one typical 2D liver geometry, replicated these perfusion patterns to generate diverse problem instances. We also applied our models to real 3D patient cases.

The 2D liver cases had 3 structures: PTV, normal liver, and unspecified normal tissue (Figure 2.2). Figure 2.3 shows the 5 perfusion patterns investigated within the same geometry. On the left are axial views of liver perfusion maps, color-coded by level of perfusion, that were observed in clinical cases; on the right are the synthesized cases where the PTV voxels (crosses) are contoured and the surrounding liver voxels (dots) are color-coded with respect to their perfusion-based relative importance (g) values, see (2.7). The unspecified normal tissue is not shown.

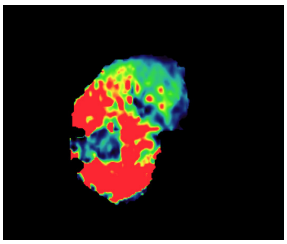
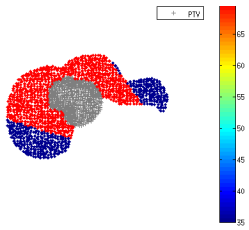
Table 2.1 summarizes parameter values used in the models for the 2D cases. We note that the upper bound for the normal tissue, $\overline{\ell\text{EUD}}_2$, is given (in physical dose) as 80 Gy in order to allow the fEUD and GLF models to focus on improving solely the dose distribution in the normal liver. However, when applying the models to real patient data, all bounds on $\ell\text{EUD}_s^+(z)$ to critical structures are based on clinical practice at the University of Michigan Hospital.

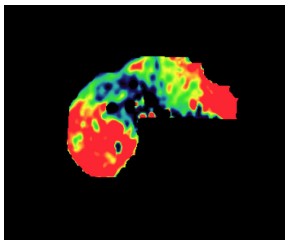
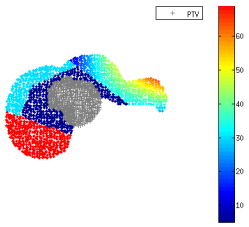
(a) Pattern P1: Large clusters of high and poor perfusion.

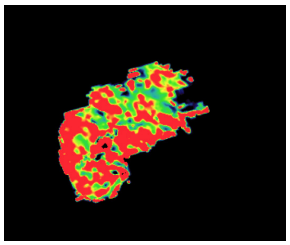
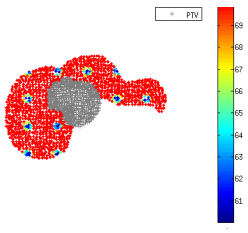
(b) Pattern P2: Small regions of high perfusion.

(c) Pattern P3: High perfusion around tumor.

(d) Pattern P4: Poor perfusion around tumor and high elsewhere.

(e) Pattern P5: High perfusion throughout.

Figure 2.3: Commonly encountered perfusion patterns on DCE-MRI (left); Synthesized 2D liver cases with PTV in gray and surrounding liver voxels color-coded by relative importance (i.e., $g_j(f_j)$) values (right).

We also considered five typical 3D patient cases. Table 2.2 summarizes parameter values used for all 3D cases. Lastly, in the 3D cases, 14 non-coplanar beams were used in order to allow flexibility to produce an appreciable alternative dose distribution (in the 2D cases, nine beams were used).

Structure	s	$\alpha_s^{\pm}$	$\overline{\ell\text{EUD}}_s$ (Gy)
PTV	0	$\alpha^- = 0.84$	60 (Rx dose)
PTV	0	$\alpha^+ = 0$	80
NORMLIVER	1	$\alpha^+ = 1.0$	Objective Function
NORMAL TISSUE	2	$\alpha^+ = 0$	80

Table 2.1: Parameter values used in models for the 2D cases.

Structure (s)	$\alpha_s^{\pm}$	$\overline{\ell\text{EUD}}_s$ (Gy)[a]
NORMLIVER	$\alpha^+ = 1.0$	Objective Function
PTV	$\alpha^- = 0.84$	60
PTV	$\alpha^+ = 0$	80
CORD	$\alpha^+ = 0$	25
LTKIDNEY	$\alpha^+ = 0$	27.3
RTKIDNEY	$\alpha^+ = 0$	27.3
STOMACH	$\alpha^+ = 0$	27.5
HEART	$\alpha^+ = 0$	32
DUODENUM	$\alpha^+ = 0$	30
BOWEL	$\alpha^+ = 0$	30
ESOPHAGUS	$\alpha^+ = 0$	27.5

Table 2.2: Parameter values used in models for the 3D cases. [a]Approximated lower and upper bounds based on clinical practice at University of Michigan Hospital.

2.4 Results

In this section we show the potential benefits and limitations of treatment plans obtained by the ℓEUD, fEUD, and GLF models. First, we discuss results for the sythensized 2D cases to provide intuition on differences in the resulting dose distributions. These examples also show which perfusion patterns have the most to gain from the models that use functional imaging information. Then, we discuss results for five representative real 3D patient cases.

In computing liver ℓEUD, we only concern ourselves with pre-treatment functional liver and denote it as $s = \text{fLiver}$:

$$\ell\text{EUD}^+_{\text{fLiver}}(z) = \frac{1}{|\tilde{V}^{\text{pre}}_{\text{Liver}}|} \sum_{j \in \tilde{V}^{\text{pre}}_{\text{Liver}}} z_j \tag{2.16}$$

2.4.1 2D Liver Cases

We compare the $\ell\text{EUD}^{+}_{\text{fLiver}}$ (as in (2.16)) achieved by the fEUD and GLF plans in Table 2.3. We present GLF values of the three plans in Table 2.4 and in Figure 2.4, along with pre-treatment GLF for each case. We use the "true" expression for GLF defined in (2.14) to calculate post-treatment global liver function values. Recall that, for each case, the solution process for the non-convex GLF model is warm-started with the solution of the fEUD model; after the numerical optimization reaches a pre-specified time limit (we use 30 hours), we report either the algorithm output or its warm-start input, whichever achieves higher post-treatment GLF, as the GLF plan.

The values reported in Table 2.4 include the pre-treatment (Pre-Tx) GLF, post-treatment (Post-Tx) GLF from the 3 plans (GLF, fEUD, ℓEUD) and an upper bound on optimal post-treatment GLF based on a piecewise linear concave upper bound on each term in (2.14) (GLF UB).* As expected, the GLF plan achieved the best GLF among the three plans. However, fEUD plans achieved varying improvement in GLF over ℓEUD plans. Although in our experiments the fEUD plan typically achieved a better GLF than the ℓEUD plan did, we warn the reader that this may not *always* be the case: the ℓEUD plan may achieve a better GLF than an fEUD plan in very select perfusion patterns (e.g., consider a liver perfusion pattern with alternating radial strips of high and poor perfusion, each rooted at the PTV).

We highlight two patterns, P1 (Figure 2.5a) and P3 (Figure 2.6a), for their noticeable and limited differences, respectively, in GLF achieved by the fEUD plan over the ℓEUD plan. Figures 2.5 and 2.6 provide a visual comparison of the ℓEUD, fEUD, and GLF plans for each of these two cases, and their impacts on liver function. The dose difference between the ℓEUD and GLF plans to each liver voxel is shown in Figures 2.5b and 2.6b, and the dose difference between the fEUD and GLF plans to each liver voxel is show in Figures 2.5c and

*For each term in (2.14), the upper bound is a function that coincides with the corresponding curve at 0, stays constant between 0 and the dose where function begins to fall off, and then decreases at a constant rate selected so that 0 is reached at the maximum allowable dose, where the voxel has 0 post-treatment function. To obtain an upper bound on optimal post-treatment GLF, we maximize the sum of these terms over all liver voxels, subject to constraints (2.3)–(2.6).

2.6c.

Perfusion Patterns	ℓEUD Plan	fEUD Plan	GLF Plan
P1: Large clusters of high and poor perfusion	18.9	19.0	21.6
P2: Small regions of high perfusion	8.6	6.9	10.7
P3: High perfusion around tumor	18.7	18.7	20.8
P4: Poor perfusion around tumor	14.5	14.5	17.0
P5: High perfusion throughout	6.5	5.8	10.2

Table 2.3: $\ell\text{EUD}^{+}_{\text{fLiver}}$ (Gy) achieved by fEUD and GLF plans on 2D cases.

Perfusion Patterns	Pre-Tx	CCV UB	GLF Plan	fEUD Plan	ℓEUD Plan
P1: Large clusters of high and poor perfusion	0.729	0.686	0.555	0.509	0.480
P2: Small regions of high perfusion	0.926	0.926	0.922	0.908	0.890
P3: High perfusion around tumor	0.769	0.715	0.558	0.524	0.521
P4: Poor perfusion around tumor	0.608	0.598	0.564	0.545	0.545
P5: High perfusion throughout	0.923	0.869	0.714	0.675	0.675

Table 2.4: Pre- and post-treatment GLF achieved by ℓEUD, fEUD and GLF plans on 2D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function.

2.4.2 Clinical (3D) Liver Cases

In this section we present results of applying the ℓEUD, fEUD, and GLF models to five real patient cases. We compare the $\ell\text{EUD}^{+}_{\text{fLiver}}$ (as in (2.16)) achieved by the ℓEUD, fEUD, and GLF plans in Table 2.5. GLF values are presented in Table 2.6 and Figure 2.7. Since these patient cases have perfusion patterns that are analogous to certain 2D patterns, we discuss in detail only patient case 1 (PC1), which is an example where the perfusion map led to an appreciable decrease in dose to high-functioning liver using the fEUD model, and an increase in global liver function using the GLF model. Using Eclipse (Varian Medical, Palo Alto, CA) to visualize dose distributions, we show in Figures 2.8a, 2.8b, and 2.8c the same axial view from the ℓEUD, fEUD, and GLF plans, respectively. In the background is an axial CT of the patient: the PTV is contoured in pink, and the liver is contoured in orange. The liver is represented separately in Figure 2.9a by its gray-scale DCE-MRI perfusion

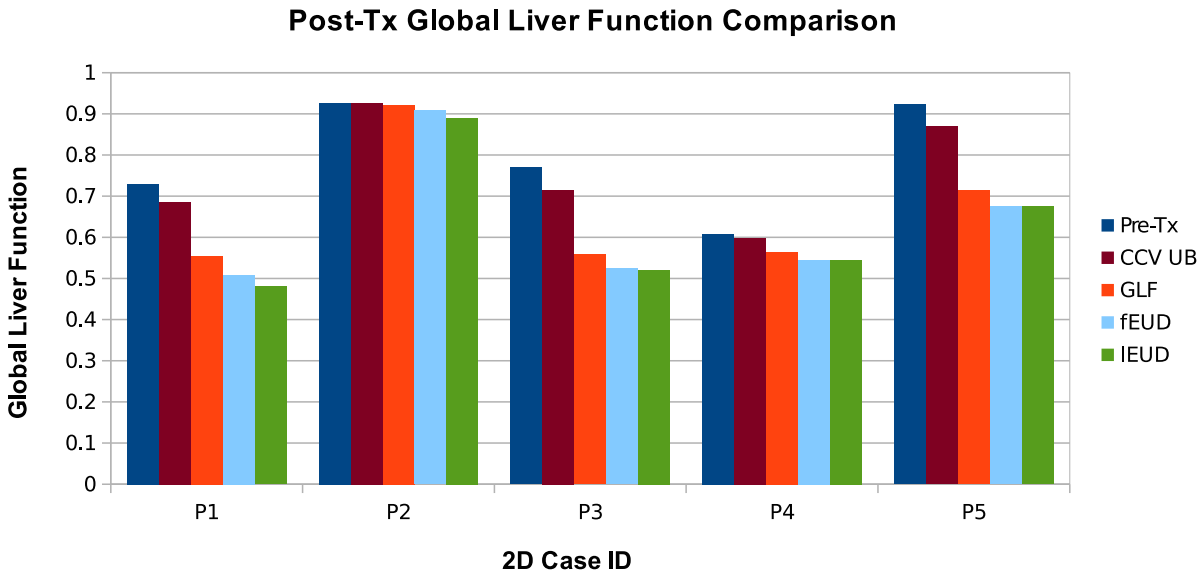
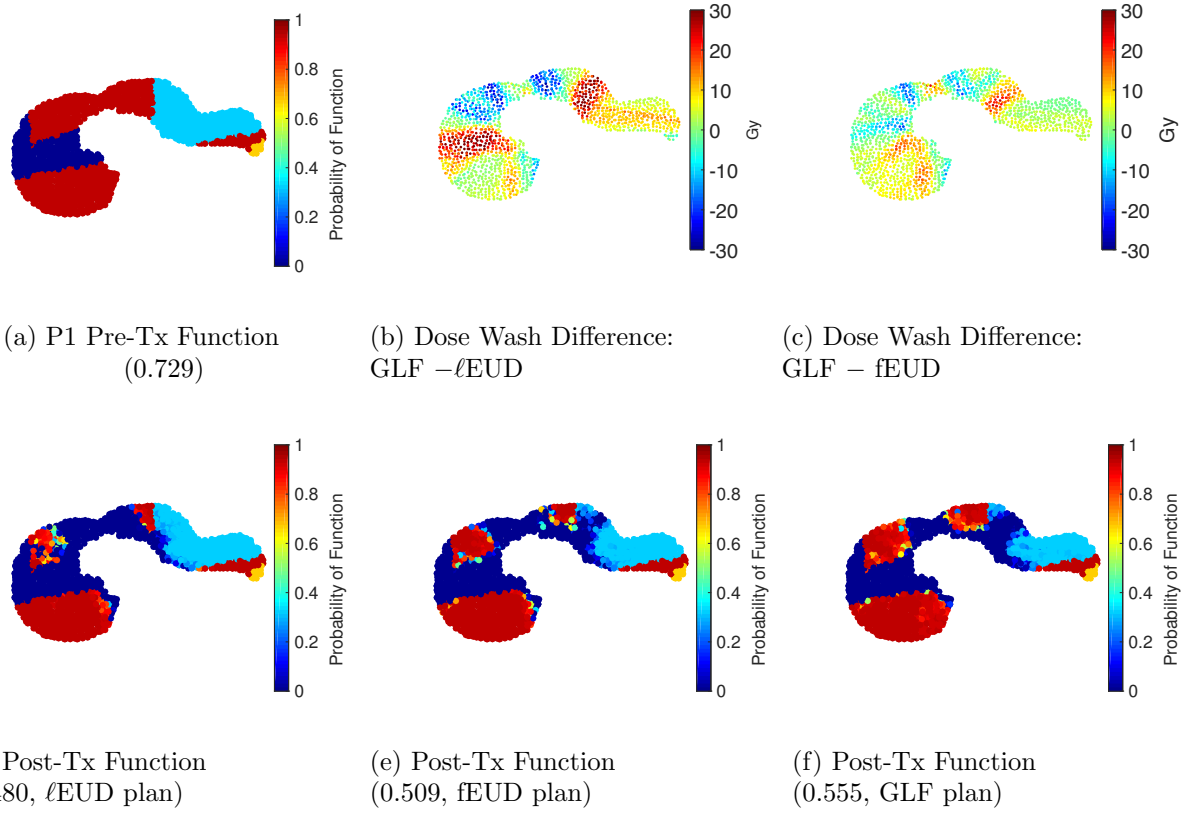


Figure 2.4: Pre- and post-treatment GLF achieved by ℓEUD, fEUD and GLF plans on 2D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function. (Note: all cases have some deficit in function at the beginning of treatment.)

map, which is also overlaid on the axial CTs (Figures 2.8a, 2.8b, and 2.8c). Bright/dark coloring represent high/poor perfusion. To see the change in dose at each voxel, we show two dose wash differences (fEUD plan minus ℓEUD plan, and GLF plan minus ℓEUD plan) in Figures 2.9b and 2.9c, where lighter color intensities correspond to larger dose differences, and hot/cold colors correspond to positive/negative dose differences.

Patient Case	ℓEUD Plan	fEUD Plan	GLF Plan
PC1	20.3	19.4	21.7
PC2	6.3	5.7	8.3
PC3	6.9	7.0	8.0
PC4	5.8	5.4	5.4
PC5	6.5	5.8	10.2

Table 2.5: $\ell\text{EUD}^{+}_{\text{fLiver}}$ (Gy) achieved by ℓEUD, fEUD, and GLF plans on 3D cases.



(a) P1 Pre-Tx Function (0.729)

(b) Dose Wash Difference: GLF $-\ell$EUD

(c) Dose Wash Difference: GLF $-$ fEUD

(d) Post-Tx Function (0.480, ℓEUD plan)

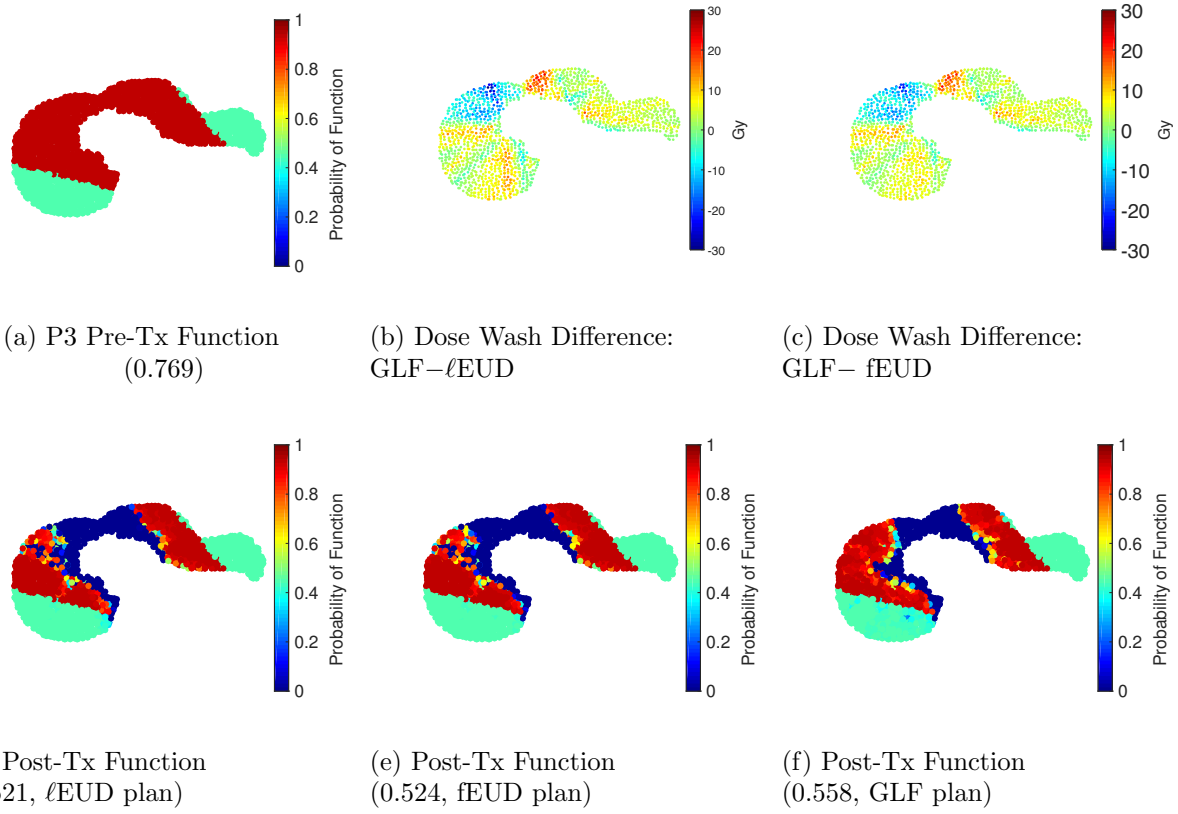
(e) Post-Tx Function (0.509, fEUD plan)

(f) Post-Tx Function (0.555, GLF plan)

Figure 2.5: Visualization of pre- and post-treatment liver function maps and dose wash differences for 2D liver perfusion pattern P1: There is noticeable reduction in dose to high-functioning regions using either fEUD or GLF objective. (2.5a), (2.5d)-(2.5f) are function maps; (2.5b)-(2.5c) are dose wash differences.

2.5 Discussion

Although it hasn't been studied which values of GLF are clinically acceptable, the GLF model does produce treatment plans that retain the most liver function compared to the fEUD and ℓEUD models. This is done by delivering dose exceeding the damage saturation threshold to fewer high-functioning voxels. We first discuss the results of 2D cases. Table 2.3 shows that, for most cases, the fEUD and GLF plans achieve $\ell\text{EUD}^{+}_{\text{fLiver}}$ comparable to the ℓEUD plan (albeit dependent on the specific perfusion pattern). Moreover, the plans are guaranteed to achieve adequate target coverage, which is enforced by constraint (2.3). In certain cases, although there is a notable increase in $\ell\text{EUD}^{+}_{\text{fLiver}}$ in the GLF plan from the ℓEUD plan, there is also notable gain in GLF (see Figure 2.4). From the dose wash



(a) P3 Pre-Tx Function (0.769)

(b) Dose Wash Difference: GLF−ℓEUD

(c) Dose Wash Difference: GLF− fEUD

(d) Post-Tx Function (0.521, ℓEUD plan)

(e) Post-Tx Function (0.524, fEUD plan)

(f) Post-Tx Function (0.558, GLF plan)

Figure 2.6: Visualization of pre- and post-treatment liver function maps and dose wash differences for 2D liver perfusion pattern P3: There is noticeable reduction in dose to high-functioning regions using GLF objective but limited difference using fEUD objective. (2.6a), (2.6d)-(2.6f) are function maps; (2.6b)-(2.6c) are dose wash differences.

differences, we observe that applying the fEUD and GLF models may be more effective in reducing dose to highly-perfused liver for particular perfusion patterns than for others. Specifically, the fEUD model follows the intuition that patterns P1, P2, and P4 have more beam paths to the target that are composed of low functioning liver voxels than patterns P3 and P5 do. In Figures 2.5b and 2.5c, we see that voxels with the higher perfusion mostly receive less dose, but in order to maintain target coverage, this dose is compensated for by boosting voxels with poor perfusion. However, for pattern P3, observe that although the fEUD plan achieves almost no improvement in post-treatment GLF over the ℓEUD plan (Figures 2.6b and 2.6c look similar due to subtracting similar plans, and consequently so do Figures 2.6d and 2.6e), the GLF plan is able to better preserve additional parts of high-

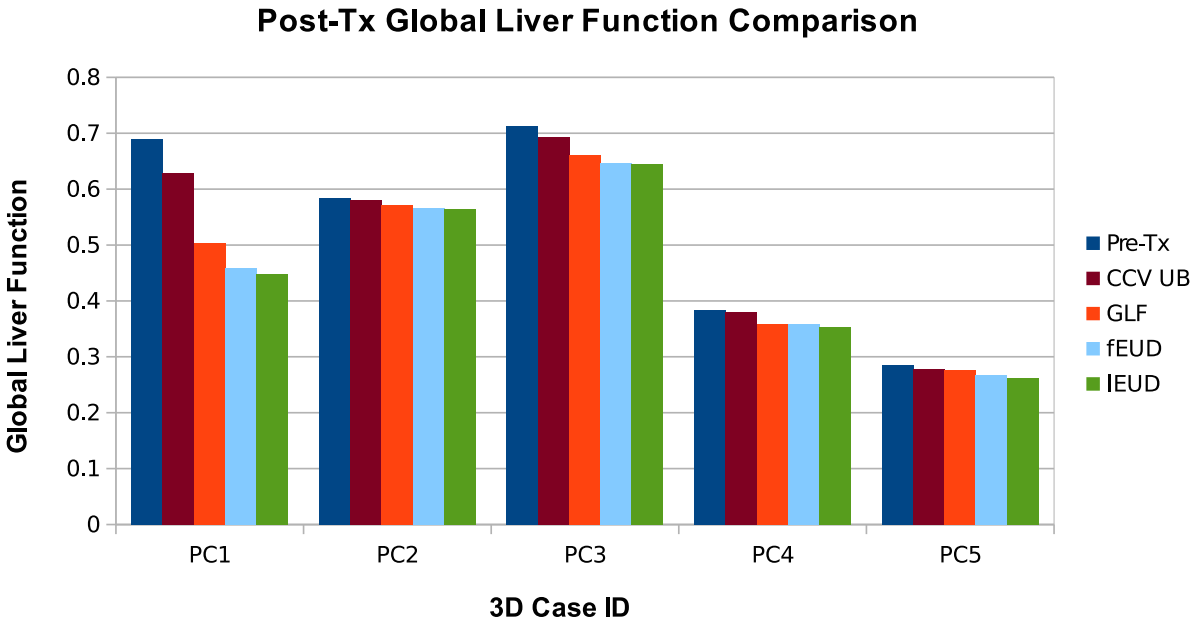
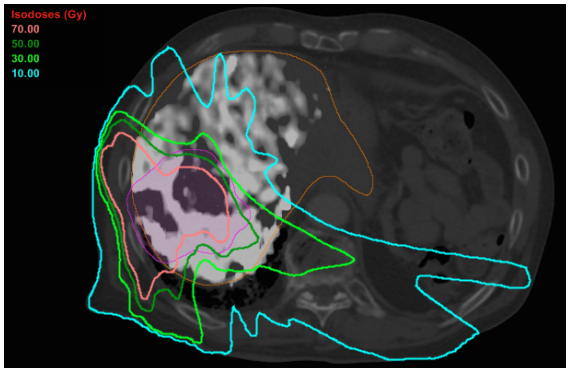


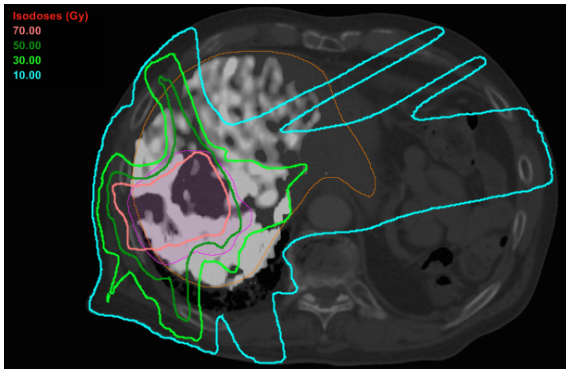
Figure 2.7: Pre- and post-treatment GLF achieved by ℓEUD, fEUD, and GLF plans on 3D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function. (Note: all cases have some deficit in function at the beginning of treatment.)

functioning liver and therefore achieves higher post-treatment GLF. Figures 2.5c and 2.6c contrast how the GLF and fEUD objectives prioritize where dose is deposited. The GLF plan achieves better GLF than the fEUD plan by delivering additional dose (red) to both a small region of high-functioning voxels, which are damage-saturated already in the fEUD plan, and large regions of the low-functioning voxels, which contribute little to global function. Thus, accounting for the two dose thresholds in the liver response allows a reduction in physical dose of up to 30 Gy (blue) in a large region of the high-functioning liver.

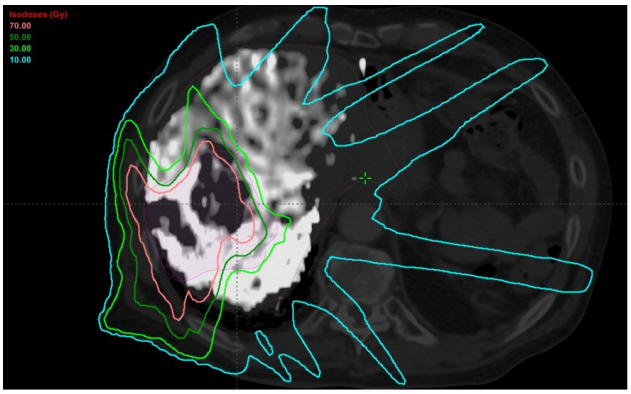
Figure 2.10 shows how the GLF plan is able to achieve better GLF than the fEUD plan does. The blue line represents a typical dose-response curve of a liver voxel with high pre-treatment function, i.e., a voxel that has pre-treatment probability of function of at least 0.5, which corresponds to perfusion of at least $F_{0.5}$ (the parameters for this particular dose-response curve are derived from using the average pre-treatment function of these voxels). Liver voxels are then grouped into bins by dose for each plan (bin size of 10 Gy-EQD2, with



(a) ℓEUD Plan, Post-Tx GLF 0.448

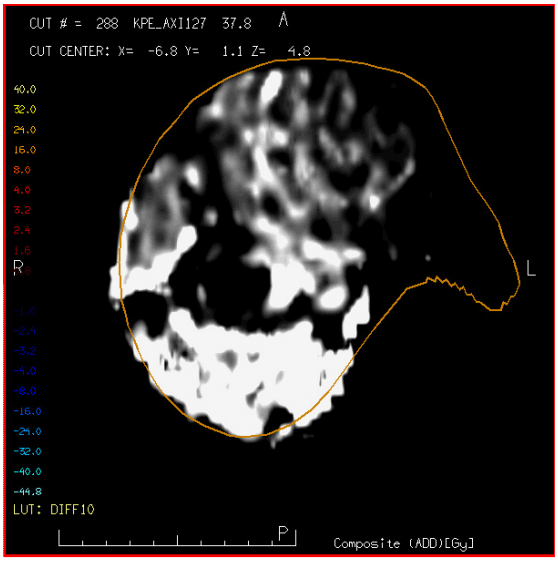


(b) fEUD Plan, Post-Tx GLF 0.459

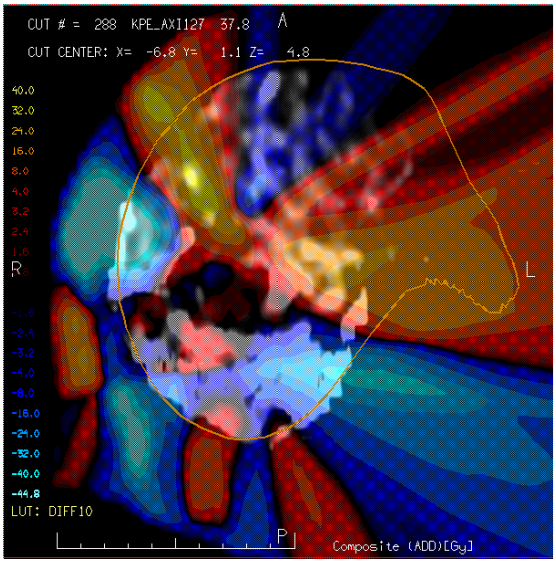


(c) GLF Plan, Post-Tx GLF 0.504

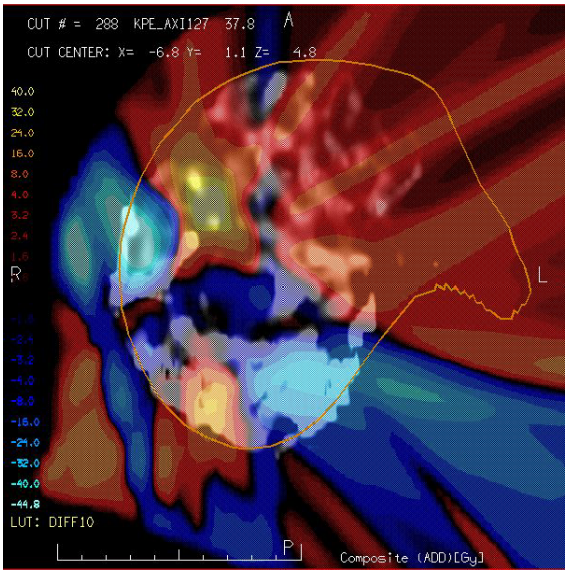
Figure 2.8: Dose distributions for ℓEUD, fEUD, and GLF plans illustrated for Patient Case 1.



(a) Perfusion map of liver (grayscale: light = high; dark = low), Pre-Tx GLF 0.689



(b) Dose wash difference (fEUD-ℓEUD) overlaid on perfusion map.



(c) Dose wash difference (GLF-ℓEUD) overlaid on perfusion map.

Figure 2.9: Dose distributions, and their differences, for ℓEUD, fEUD, and GLF plans illustrated for Patient Case 1. Liver is contoured in light brown and filled with its black-and-white perfusion map.

Patient Case	Pre-Tx	GLF UB	GLF Plan	fEUD Plan	ℓEUD Plan
PC1	0.689	0.628	0.504	0.459	0.448
PC2	0.585	0.581	0.572	0.567	0.565
PC3	0.712	0.693	0.661	0.647	0.645
PC4	0.383	0.380	0.359	0.359	0.353
PC5	0.285	0.278	0.276	0.268	0.262

Table 2.6: Pre- and post-treatment GLF achieved by ℓEUD, fEUD, and GLF plans for 3D cases. GLF UB is an upper bound on optimal post-treatment GLF obtained by maximizing a piecewise-linear concave relaxation of GLF dose response function.

all voxels beyond the damage-saturation threshold of 50 Gy-EQD2 grouped into a single bin) and each bin is represented in the figure by circles, where the diameter of each circle reflects the number of voxels in that bin (red circles correspond to the fEUD plan while green circles correspond to the GLF plan). The GLF plan delivers dose beyond the damage-saturation threshold to fewer voxels compared to the fEUD plan, and instead delivers *tolerable* amounts of dose (e.g., 15 to 20 Gy-EQD2) to *more* voxels to achieve comparable target coverage. Since tolerable amounts of dose insignificantly affect post-treatment function, the resulting post-treatment GLF is higher.

Although the differences in dose distributions from 3D patient cases are not as conspicuous as those from the 2D cases, we observe analogous behavior. Qualitatively, patient case 1 has a 3D perfusion map analogous to the 2D perfusion pattern P1 in Figure 2.3a: large clusters of various perfusion surrounding the tumor. Figures 2.8, 2.9, and 2.11 show this analogous benefit of using a GLF objective. In particular, comparing Figures 2.8a, 2.8b, and 2.8c, we observe that the bright-green 30 Gy iso-dose line (typically, where damage saturates) is pushed further in toward the PTV and covering less functional liver in the GLF plan, while more highly-perfused liver is covered by the 30 Gy line in the fEUD and ℓEUD plans. Moreover, using the ℓEUD plan as the baseline plan, the dose wash difference in Figure 2.9c (GLF$-\ell$EUD plan) shows a larger reduction in dose to the posterior region of the liver, which is the highest-functioning portion (as seen in Figure 2.9a), compared to the dose wash difference in Figure 2.9b (fEUD$-\ell$EUD plan).

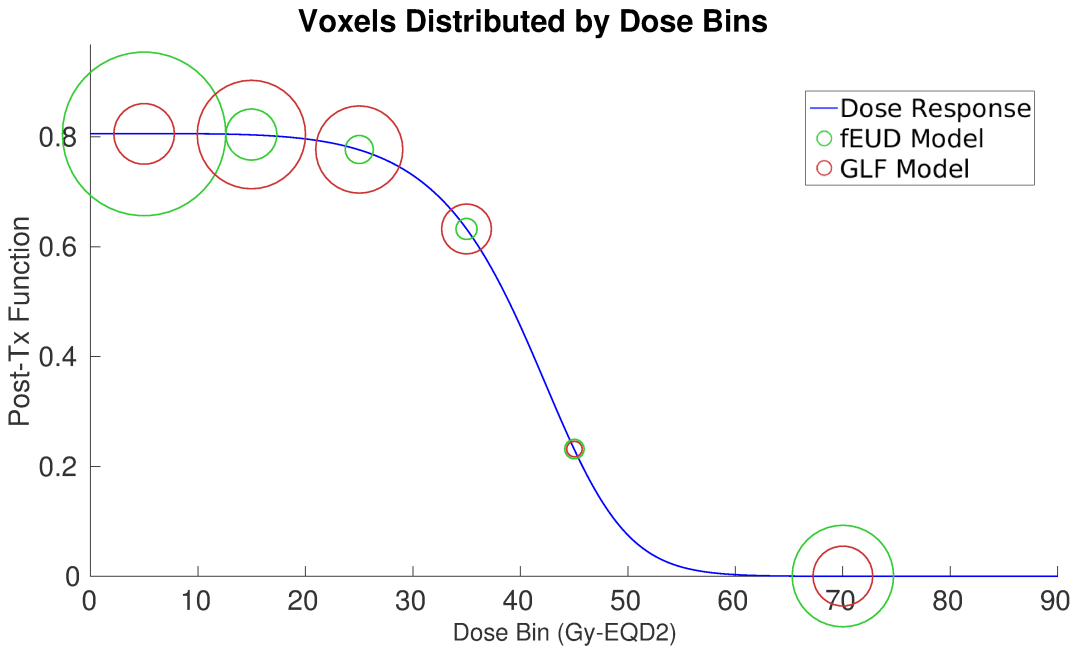


Figure 2.10: Distribution of voxels with high pre-treatment function by dose-bins of 10 Gy-EQD2 (one bin for dose-saturated voxels) along post-treatment function line-plot; 2D case P1.

The concave upper bounds (GLF UB) for certain perfusion patterns were not tight (e.g., P1, P3, P5, and PC1) due to overestimating the function on voxels receiving high dose, namely they should have no function at all from being saturated in dose. However, in other cases, where pre-treatment GLF was close to the achieved GLF values, we achieved post-treatment GLF that was also close to the concave upper bound, likely due to a sandwiching effect. Although this upper bound function is easy to optimize, in our future work we intend to investigate how to close the gap in the aforementioned cases.

Our current model considers voxel-based organ function information. A next step is to consider function at a higher level by preserving not only individual high-functioning liver voxels, but also so-called "highly functional subvolumes of liver," i.e., contiguous groups of voxels — identifying such subvolumes would supplement existing dose-based evaluation criteria. Another direction is to make our model robust by accounting for uncertainty in functional imaging data, e.g., from image registration errors and the perfusion measurements themselves. Finally, since response parameters are currently population-based, we intend

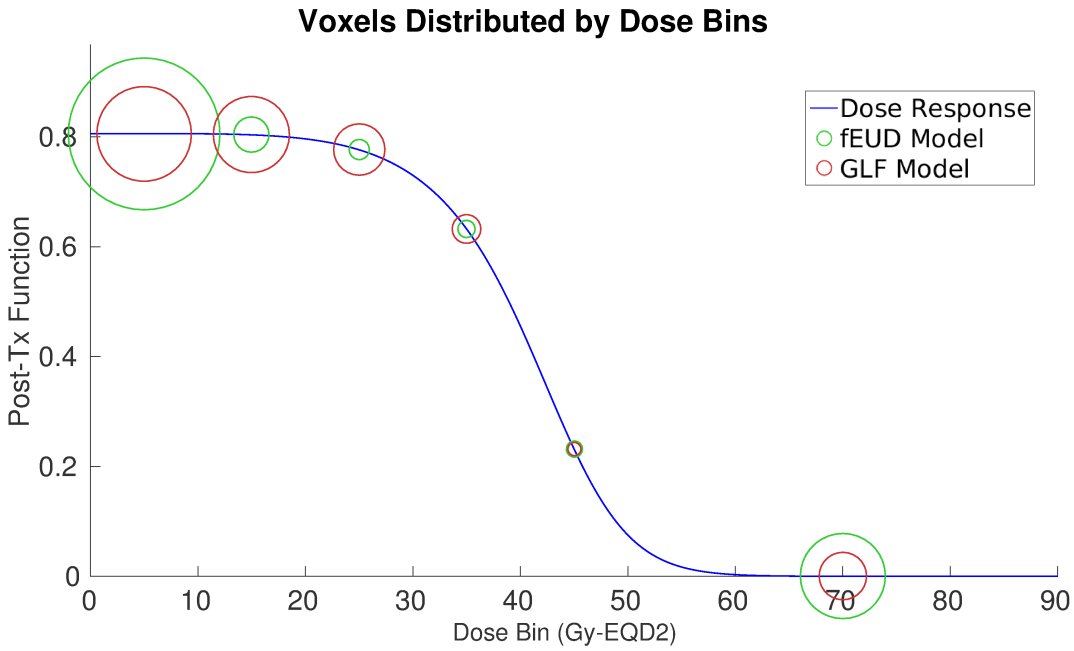


Figure 2.11: Distribution of voxels with high pre-treatment function by dose-bins of 10 Gy-EQD2 (one bin for dose-saturated voxels) along post-treatment function line-plot; Patient Case 1.

to further individualize planning through an adaptive framework by anticipating, part-way through treatment, patient-specific radiosensitivity with patient-specific response parameters (such response models have also been developed in *Wang et al.*, 2016).

2.6 Conclusion

We developed methods to explicitly incorporate voxel-level liver function information into SBRT treatment planning. The fEUD model can often effectively generate an alternative dose distribution that reduces dose to highly-perfused voxels and increases dose to poorly-perfused voxels (not necessarily by the same amount), while achieving the same target linearized equivalent uniform dose and satisfying dose limits to other critical structures. Although this model is computationally inexpensive in comparison to the GLF model, minimizing fEUD is only a surrogate for maximizing post-treatment global liver function and is not a uniformly good surrogate. Though computationally more expensive, the GLF model

directly optimizes expected post-treatment global liver function, a more clinically-relevant metric. The results of this work suggest a need to further investigate numerical methods that more efficiently optimize nonconvex objective functions such as GLF. Although these models are effective for certain types of perfusion patterns, future work consists of quantitatively identifying perfusion patterns that may indicate patient benefit from planning treatment with the fEUD or GLF models.

CHAPTER III

Adaptive Liver Stereotactic Body Radiotherapy

3.1 Introduction

In abdominal stereotactic body radiation therapy (SBRT), the patient may come in for treatment with a geometry that is different from the one in the planning CT. A change in patient geometry between fractions is called *interfraction motion* (not to be confused with *intrafraction motion*, which refers to changes in patient geometry during a fraction). On the day of a treatment, the patient may have a cone beam CT (CBCT) image to update the patient geometry. An example of interfraction motion can be seen in Figure 3.1. A treatment where treatment plans are adapted to such changes is called adaptive radiation therapy (ART). "Online" ART adapts plans during treatment delivery while "offline" ART adapts plans between treatments according to updated information from the previous fraction and adjusts plans for future fractions.

Typically, adapting by re-optimizing on the day of treatment is operationally taxing due to the amount of time required for both planning and a quality assurance (QA) check, extending the patient's visit. The current practice usually employs a simpler approach called image-guided radiotherapy (IGRT), which adjusts the plan if the image indicates that an OAR has moved into a region that would be overdosed by the original plan by shifting the original dose distribution (planned from the nominal geometry captured in the planning CT) in the direction of the OAR motion to avoid over-dosing it. A physician will guess-and-

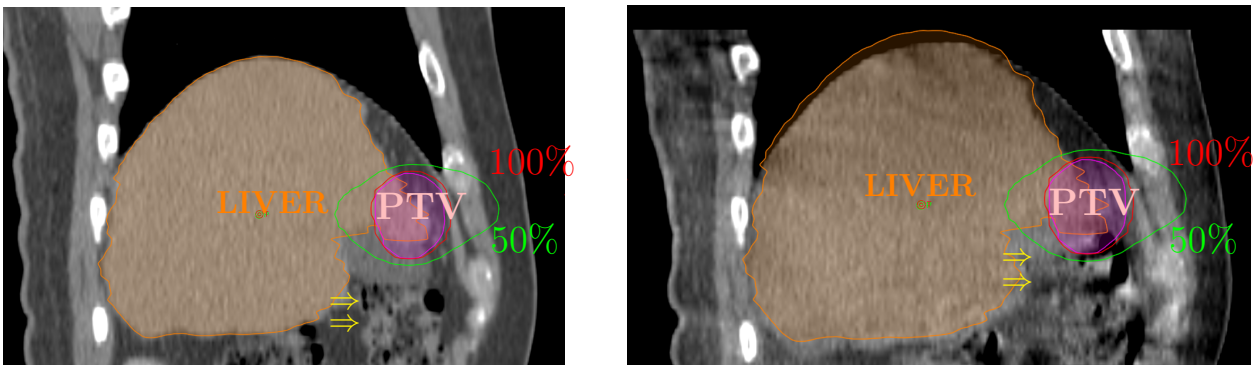


Figure 3.1: Left: with the gall bladder filled, the colon (yellow arrows) is a safe distance from the high dose region around the PTV; Right: the gall bladder loses volume and the colon moves closer to the high dose region, potentially resulting in overdosing the colon. Red and green contours correspond to the 100% and 50% of prescription dose iso-dose lines.

check shifts simply by looking at an updated patient geometry, which can be both inefficient and result in suboptimal adaptation. A shift in the dose distribution inevitably results in under-dosing the PTV in order to spare the OAR (see Figure 3.2). In addition to a lack of consideration of per-fraction dose limits in IGRT, there is no consideration of compensating for reduced PTV coverage in future fractions. Finally, if there are multiple OARs that surround the PTV closely, any shift away from one OAR would result in over-dosing another OAR, i.e., any shift of the dose distribution would result in a plan that is infeasible with respect to per-fraction dose limits.

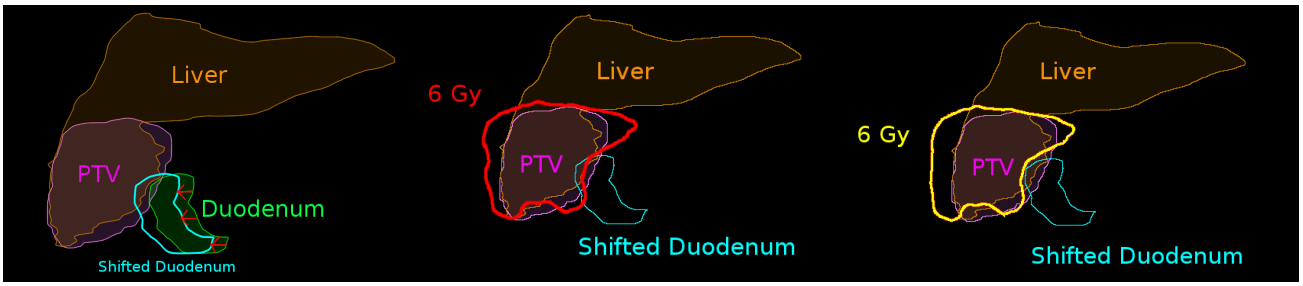


Figure 3.2: Example of IGRT. Left: duodenum shifts toward PTV. Middle: shifted duodenum would be overdosed with more than 6 Gy in the original plan. Right: IGRT shifts the dose distribution to avoid overdosing the shifted duodenum.

3.2 Related Literature

ART planning for both interfraction (between fractions, e.g., patient changes in geometry or setup error) and intrafraction (e.g., breathing) motion has been thoroughly studied (see *Chan and Mišić*, 2013; *McShan et al.*, 2006; *Men et al.*, 2012; *Sir et al.*, 2012; *Unkelbach and Oelfke*, 2005; *Wu et al.*, 2008). There are two main approaches in ART: offline and online. In online ART, the patient is imaged immediately before treatment and a plan is generated based on the day-of geometry (DOG), followed by plan delivery. In offline ART, planning is done before the patient comes in for treatment and is based on images of the patient taken in previous fractions and/or imaging sessions and the resulting plan is delivered in future fractions.

McShan et al. (2006) is one of the first works to consider multiple geometries but their models find one single plan that performs well on average; *Sir et al.* (2012) adds on an important feedback component which updates the treatment planning model with doses delivered as fractions progress, but also only generates one plan for future fractions that performs well averaged over several geometries. *Chan and Mišić* (2013) develop a framework for offline ART for breathing motion in lung cancer treatment: a collection of "breathing states" is defined based on images to reflect intrafraction motion, and for each state, the corresponding dose deposition matrix is computed. Their robust optimization model assumes the probability mass function (PMF) of breathing states of the lung belongs to an uncertainty set (where the probability of being in each state at any point in time is contained in some range). *Unkelbach and Oelfke* (2005) model an expected voxel dose value over a distribution that describes, for each voxel i, the probability that it moves to the position of a static voxel j, and tries to find a fluence map that minimizes a weighted sum of 3 terms: two-sided quadratic penalties on expected CTV voxel doses, the variance in dose to CTV voxels, and one-sided quadratic penalties on expected OAR voxel doses.* The resulting fluence map is delivered across a pre-specified number of equal fractions.

*This is not to be confused with expected quadratic penalties.

A more complex version of offline adaptive therapy consists of generating multiple plans, each of which accounts for a different realization of patient geometry. The "online" component is to use day-of imaging to decide which of the plans computed offline to use. *Burridge et al.* (2006); *Foroudi et al.* (2009); *Gill et al.* (2013); *Kuyumcian et al.* (2012); *Murthy et al.* (2011) take such an approach by considering multiple scenarios of one structure of interest that may arise the day of treatment (e.g., 3 plan for 3 bladder sizes, 6 plans for 6 PTV sizes, etc.), which are special cases of considering DOG. This approach of preparing multiple plans for multiple scenarios is fittingly called "plan of the day" ART. The benefit of this approach is that is saves online-adaptive computational efforts while still maintaining flexibility in adapting to a variety of geometries.

Wu et al. (2008) take an online adaptive approach by using CBCT images to collect updated geometry information just before the treatment and apply image deformation to the original dose distribution from the planning CT to derive what the deformed geometry *should receive*, i.e., how to deliver what was originally intended to each voxel. Linear voxel-based penalties have goal doses derived from the deformed-dose distribution, and their model finds a fluence map that delivers as close a dose distribution as possible to the deformed dose distribution. However, there is no guarantee that the deformed dose distribution is feasible, potentially resulting in unequal fractions. Note that their approach, similarly to IGRT, is dependent on the nominal (baseline) geometry seen on the planning CT (pCT).

Online ART considers the most up-to-date patient geometry, but the planning process leads to lengthened treatment time (i.e., patient waiting on the couch); offline ART approaches develop plans offline allowing for minimal additional waiting time for the patient, but do not consider a most up-to-date patient geometry. For prostate treatment planning, recent investigations take a combination of the two strategies by exploiting the benefits of each. *Lei and Wu* (2010) consider accounting for uncertainties in geometry due to deformation and rotation of structures in additional to the conventional translational structure motion. Offline, they attempt to account for systematic errors (e.g., prostate rotation) and

adjust margins to account for random errors. Doing so in addition to online adaptation allowed for moderate margin decrease compared to using online adaptation only. *Yang et al.* (2014) consider an approach that performs online ART only when necessary, i.e., when significant interfraction motion is detected, and in the meantime delivers the most updated plan otherwise, which reduces online computational efforts.

The works above all account for adapting treatment with respect to *only* physical dose and are likely to result in unequal fractions. Although they consider cumulative physical dose effects, there is no consideration of *biological* effects, measured by the so-called biologically effective dose (BED) (see *Barendsen*, 1982; *Fowler*, 1989; *Jones et al.*, 2001), which we define formally later. Unlike total physical dose to a voxel, i.e., the sum of doses received by the voxel over the course of a multi-fraction treatment, which is used in stationary geometry and equal-fraction setting, the BED is derived from the linear-quadratic cell survival model that captures radiation cell kill for a sequence of potentially unequal fractions (e.g., a non-stationary treatment schedule for a changing patient geometry).

BED-based planning has also been studied, but for the most part in a deterministic setting. *Kim et al.* (2009) take a Markov decision process approach to incorporate temporal effects via BED and then *Kim et al.* (2012) incorporate the possibility of adapting the dose distribution spatially, i.e., allowing fluence to change fraction to fraction. Neither of the above papers accounts for any sources of uncertainty. *Saberian et al.* (2015) propose a model that maximizes the BED of the average tumor dose subject to OAR BED limits in equal fractions. This objective function implicitly assumes that homogenous PTV coverage is available (i.e., BED of the average dose would be the same as the average BED). Although this leads to a formulation with convenient mathematical properties (i.e., they can maximize a concave objective), it is a simplified measure of PTV coverage. *Unkelbach et al.* (2012) consider one OAR vs one PTV tradeoff in BED and derive mathematical properties that help determine fractionation schedule (i.e., number of fractions). Both works assume no interfraction motion.

Chen et al. (2008) is the first work that proposed a treatment planning model that accounts for BED in an ART framework. They consider one tumor (whose dose distribution is assumed to be uniform) and one OAR, whose BED for each voxel is determined by the tumor dose and a ratio known as the "sparing factor" that is determined by the voxel's location relative to the tumor. They adapt to the sparing factor changing from fraction to fraction (to emulate voxel motion) by assuming that the sparing factor comes from a normal distribution in future fractions and do not consider adapting the fluence map.

At UCLA's department of radiation oncology, *Mutic and Dempsey* (2014) introduce ViewRay's (Los Angeles, CA) adaptive planning strategy. They adapt to the DOG by *re-planning* according to the current fraction and to the accumulated dose of previous fractions. This is done under the assumption that the current fraction's geometry is the geometry for remaining fractions, i.e., uncertainty in future fractions' geometries in unaccounted for.

3.3 Proposed Framework

Estimating possible geometries at each fraction (e.g., using biomechanical models as a predictive tool), we can anticipate sequences of geometries throughout treatment (we refer to each sequence as a *scenario*). Operationally, we optimize a set of plans — one for each potential geometry — before the patient arrives, and deploy the plan appropriate to the DOG. In this chapter, we propose a treatment planning model consistent with this approach. More specifically, the model anticipates possible deformations in patient geometry and tracks voxel motion in order to account for not only the dose to a voxel at a each fraction, but also the dose accumulation throughout all fractions. This so-called "daily-offline" ART generates plans for each possible geometry in future fractions with consideration of dose accumulative effects, i.e., (i) the treatment that the patient has received so far, and (ii) the potential treatments based on the potential geometries that may be realized in future fractions. In comparison to IGRT, our model enforces a robust constraint on dose limits: no matter the sequence of geometries realized, treatment dose limits (per fraction and cumulative) are

satisfied.

3.3.1 Patient Geometry Uncertainty

An important assumption in this framework is that the set of possible geometries used in the model realistically represents geometries that may be encountered when a patient comes in for treatment. Realistically, possible patient geometries belong to a continuum, but for problem tractability, we consider only a discrete set of geometries at each fraction. Furthermore, in this chapter we assume that the realized geometry in *each fraction* is independent and identically distributed over fractions,[†] which results in a symmetric scenario tree; we leave learning of geometries fraction to fraction and dynamic support (i.e., nonsymmetric scenario tree) for future work. We note that this framework is similar to *Sir et al.* (2012) with the difference being that we account for structures' BED and we generate *multiple* treatment plans, one for each geometry, and deliver the plan corresponding to the realized geometry (or the one that is most similar). This is in contrast to generating just one plan that is delivered no matter which geometry is realized and adapting a plan by myopically re-planning for the day's geometry without consideration of cumulative dose effects.

3.3.2 Chapter Outline

In the remainder of this chapter, we formalize our approach. We begin by introducing additional notation and relevant treatment criteria that will be used. Then, we construct the deterministic and stochastic planning models, and propose methods to solve them. We then present the patient data that was used in a case study and report on experimental results.

[†]However, treatments are *not* independent over fractions since they will depend on which geometries have been realized in previous and future fractions

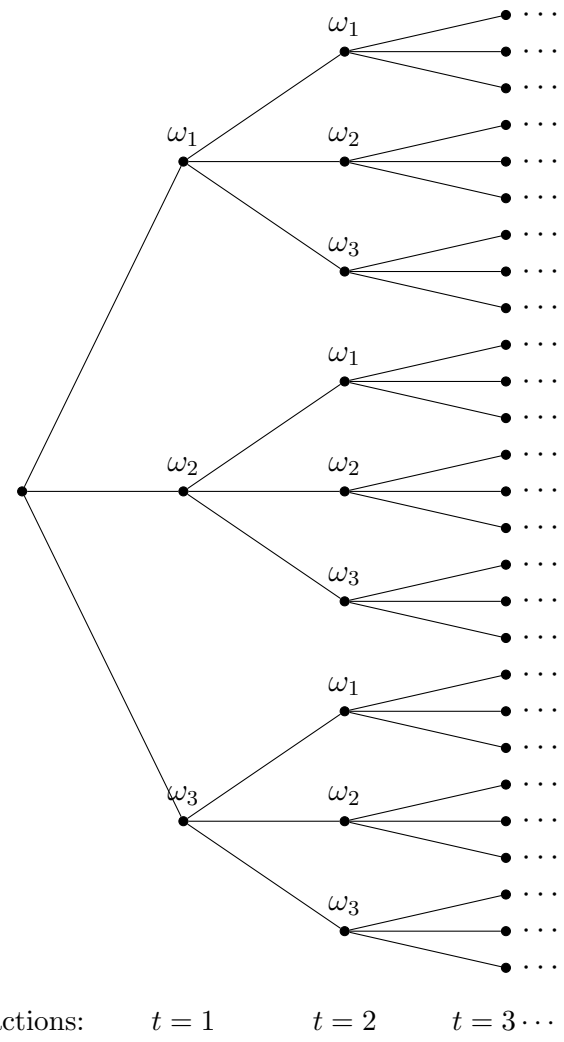


Figure 3.3: Example scenario tree with 3 possible geometries at each fraction. One path from the first fraction to the last fraction is considered a scenario.

3.4 Treatment Plan Evaluation Criteria

We largely continue to use notation of Chapter II, but introduce additional notation to capture the time component of adaptive planning: let $t = 1, \ldots, T$ index the fractions for some fixed time horizon T. Let $x^t \in \mathbb{R}_+^{|N|}$ and $z^t \in \mathbb{R}_+^{|V|}$ be the vectors of beamlet intensities and doses to voxels, respectively, in fraction t. The dose matrix D is also indexed by t to capture the fluence-dose relationship at fraction t:

$$z^t = D^t x^t, \qquad t = 1, \ldots, T. \tag{3.1}$$

Given a treatment consisting of T fractions, we denote by $z_j = (z_j^1, z_j^2, \ldots, z_j^T)$ the sequence of doses received by voxel j (note that this notation differs from notation in the previous chapter, where effects of fractionation were not being considered).

3.4.1 Dose Effect Over Time: Biologically Effective Dose

To evaluate the effect of dose delivered to a voxel over potentially unequal fractions, we use the BED (*Fowler*, 1989; *Jones et al.*, 2001), which is defined as:

$$\text{BED}_s(z_j) = \sum_{t=1}^{T} \left(z_j^t + \frac{(z_j^t)^2}{\left(\frac{\alpha}{\beta}\right)_s} \right), \quad s \in S, \tag{3.2}$$

where $\left(\frac{\alpha}{\beta}\right)_s \geq 0$ is the alpha-beta ratio specific to the structure. This value reflects the ratio of the linear to quadratic dose effects of a cell survival model. Equivalent dose in 2-Gy fractions (EQD2) to voxel j for the entire treatment is then computed as:

$$\text{EQD2}_s(z_j) = \frac{\text{BED}_s(z_j)}{1 + \frac{2}{\left(\frac{\alpha}{\beta}\right)_s}}. \tag{3.3}$$

Although SBRT treatments often deliver more than 2 Gy per fraction, this metric is still commonly considered by physicians evaluating them, because conventional RT treatments

were delivered in 2 Gy-fractions and are more easily interpretable. Note, however, that this metric is simply a scaled version of BED. Without impact on our optimization model, we will plan with respect to BED instead of EQD2.

3.4.2 Structure Criteria

To evaluate a treatment plan's effects on structures, we use the linearized EUD, whose argument is a measure with units in Gy, $d : \mathbb{R}_+^{|V_s|} \to \mathbb{R}_+$:

$$\ell\text{EUD}_s^+(d(z_j), j \in V_s) = \lambda_s \left(\frac{1}{|V_s|} \sum_{j \in V_s} d(z_j) \right) + (1 - \lambda_s) \max_{j \in V_s} \{d(z_j)\}, \quad s \in S, \tag{3.4}$$

$$\ell\text{EUD}_{\text{PTV}}^-(d_j(z_j), j \in V_s) = \lambda_{\text{PTV}} \left(\frac{1}{|V_{\text{PTV}}|} \sum_{j \in V_{\text{PTV}}} d(z_j) \right) + (1 - \lambda_{\text{PTV}}) \min_{j \in V_{\text{PTV}}} \{d(z_j)\}, \tag{3.5}$$

where we use $\lambda \in [0, 1]$ in place of α in equations (ℓEUD-underdose) and (ℓEUD-overdose) to avoid confusion with the alpha-beta ratio. Recall that if $d(z_j) = \sum_{t=1}^{T} z_j^t$, then ℓEUD_s^+ is convex in z for $s \in S$, and $\ell\text{EUD}_{\text{PTV}}^-$ is concave in z, and treatment optimization models formulated with these metrics typically result in convex problem, as was the case in Chapter II (where we made the additional assumption of equal fractions, i.e., $z_j^t = z_j^{t'} \ \forall t, t'$). If $d(z_j) = z_j^t$, these functions evaluate structure criteria for a single fraction t.

3.4.2.1 OAR Damage and PTV Hot Spots

To ensure critical structures receive tolerable amounts of dose throughout treatment, we enforce relevant BED upper bounds for all structures. For each structure s, we measure cumulative dose effect as:

$$\begin{aligned} &\ell\text{EUD}_s^+(\text{BED}_s(z_j), j \in V_s) = \\ &\lambda_s \left(\frac{1}{|V_s|} \sum_{j \in V_s} \text{BED}_s(z_j) \right) + (1 - \lambda_s) \max_{j \in V_s} \{\text{BED}_s(z_j)\}. \end{aligned} \tag{3.6}$$

This metric is typically bounded above, resulting in a convex constraint.

3.4.2.2 PTV Coverage

To measure PTV coverage across potentially unequal fractions, we consider $d(z_j) = \text{BED}_{\text{PTV}}(z_j)$ (with slight abuse of notation):

$$\ell\text{EUD}^{-}_{\text{PTV}}(\text{BED}_{\text{PTV}}(z_j), j \in V_{\text{PTV}}) = \tag{3.7}$$
$$\lambda_{\text{PTV}} \left(\frac{1}{|V_{\text{PTV}}|} \sum_{j \in V_{\text{PTV}}} \text{BED}_{\text{PTV}}(z_j) \right) + (1 - \lambda_{\text{PTV}}) \min_{j \in V_{\text{PTV}}} \{\text{BED}_{\text{PTV}}(z_j)\}.$$

Note, however, that maximizing (3.7) or bounding it below results in a non-convex problem, a challenge of BED-based planning (*Saberian et al.*, 2015; *Unkelbach and Papp*, 2015).

3.5 Deterministic Model

Our deterministic model, where we assume that geometry in all fractions is known a priori, can be formulated as:

$$\text{maximize } \ell\text{EUD}^{-}_{\text{PTV}}\left(\text{BED}_{\text{PTV}}(z_j), j \in V_{\text{PTV}}\right) \tag{DET-OPT-NCVX}$$

subject to

$$\ell\text{EUD}^{+}_{s}(\text{BED}_{s}(z_j), j \in V_s) \leq U^{\text{BED}}_{s} \qquad s \in S \tag{3.8}$$
$$\ell\text{EUD}^{+}_{s}(z^t) \leq u^{\ell\text{EUD}}_{s} \qquad s \in S, t = 1, \ldots, T \tag{3.9}$$
$$z^t = D^t x^t \qquad t = 1, \ldots, T \tag{3.10}$$
$$\sum_{t=1}^{T} x^t \leq b \qquad t = 1, \ldots, T \tag{3.11}$$
$$x^t \geq 0 \qquad t = 1, \ldots, T. \tag{3.12}$$

Here, constraints (3.8) enforce cumulative dose bounds for OARs, whereas (3.9) enforce per-fraction dose bounds; (3.10) model the fluence-dose relationship; and (3.11) prevent any one beamlet's intensity from being too high over the course of the treatment. Solving this problem with $D^t = D, t = 1, \ldots, T$, where D is the nominal geometry gives the *nominal treatment plan*. The above problem is non-convex because of its objective; we can approximate PTV coverage using physical dose instead of BED as the argument for ℓEUD:

$$\text{maximize } \ell\text{EUD}^-_{\text{PTV}}\left(\sum_{t=1}^{T} z_j^t, j \in V_s\right), \tag{DET-OPT-CVX}$$

resulting in a convex formulation.

3.6 Stochastic Models

In the stochastic model, we account for the uncertainty in patient geometry at each fraction. Let $\Omega := \{\omega_1, \ldots, \omega_K\}$ be the set of possible patient geometries at each fraction t. In a treatment with T fractions, a *scenario* is a sequence of T geometries, denoted by $\omega^{[T]} = (\omega^1, \ldots, \omega^T)$ (it also corresponds to a path in the scenario tree in Figure 3.3); the probability of $\omega^{[T]}$ occurring is denoted by $p_{\omega^{[T]}}$. Note that the total number of possible scenarios is $|\Omega|^T$. We also let, for $1 \leq t \leq T$, $\omega^{[t]} = (\omega^1, \ldots, \omega^t) \in \Omega^t$, and for $1 \leq t < t' \leq T$, we let $\omega^{[t:t']} = (\omega^t, \omega^{t+1}, \ldots, \omega^{t'}) \in \Omega^{t'-t+1}$ denote a sequence of geometries in fractions t through t'.

In order to track dose accumulation, we assume that structures retain the same number of voxels fraction to fraction. Then, to capture the relationship between dose in structures and various geometries, the dose deposition matrix will now explicitly depend on geometry: $D(\omega),\ \omega \in \Omega$. Then, the dose matrix at fraction t in scenario $\omega^{[T]}$ is the dose matrix corresponding to the geometry ω^t: $D^t(\omega^{[T]}) := D(\omega^t)$. Now that we have established notation for geometry realization, note that subscripts capture *which* geometry is realized and superscripts capture *when* a geometry is realized. Lastly, the dose distribution

and fluence at time t now also depend on $\omega^{[T]}$; let $x_i(\omega^{[T]}) = (x_i^1(\omega^{[T]}), \ldots, x_i^T(\omega^{[T]}))$ and $z_j(\omega^{[T]}) = (z_j^1(\omega^{[T]}), \ldots, z_j^T(\omega^{[T]}))$. Our stochastic model below maximizes expected tumor coverage subject to robust constraints on normal tissue sparing, ensuring the treatment is safe in any realized scenario:

$$\text{maximize} \sum_{\omega^{[T]} \in \Omega^T} \left(p_{\omega^{[T]}} \ell\text{EUD}^-_{\text{PTV}} \left(\text{BED}_{\text{PTV}}(z_j(\omega^{[T]})), j \in V_{\text{PTV}} \right) \right) \qquad \text{(STO-OPT-NCVX1)}$$

subject to

$$\begin{aligned}
\ell\text{EUD}_s^+(\text{BED}_s(z_j(\omega^{[T]})), j \in V_s) &\leq U_s^{\text{BED}} && \omega^{[T]} \in \Omega^T, s \in S \\
\ell\text{EUD}_s^+(z^t(\omega^{[T]})) &\leq u_s^{\ell\text{EUD}} && \omega^{[T]} \in \Omega^T, s \in S, \\
& && t = 1, \ldots, T \\
z^t(\omega^{[T]}) &= D^t(\omega^{[T]}) x^t(\omega^{[T]}) && \omega^{[T]} \in \Omega^T, t = 1, \ldots, T \\
\textstyle\sum_{t=1}^T x_i^t(\omega^{[T]}) &\leq b && i \in N, \omega^{[T]} \in \Omega^T \\
x^t(\omega^{[T]}) &\geq 0 && \omega^{[T]} \in \Omega^T, t = 1, \ldots, T \\
x^t(\omega^{[T]}) &= x^t(\omega'^{[T]}) && \omega^{[T]} \neq \omega'^{[T]} : \\
& && \omega^{[t]} = \omega'^{[t]}, \\
& && t = 1, \ldots, T-1.
\end{aligned}$$

The last family of constraints represents non-anticipativity restrictions on treatment decisions. This model consists of $|V|TK^T + |N|TK^T$ decision variables, $K^T|S| + TK^T|S| + TK^T|V| + |N|K^T + T|N|K^T + \sum_{t=1}^T (K^t - 1) = ((T+1)|S| + |N| + |V| + |N|)K^T + \sum_{t=1}^T (K^t - 1)$ constraints, and K^T terms in the objective. Note the significant increase in number of terms in the objective and how the consideration of scenarios multiplies the number of decision variables by K^T. Moreover, many of the sets of constraints also multiply by a factor of K^T in this formulation in comparison to the formulation in (DET-OPT-NCVX).

We attempted to solve instances of (STO-OPT-NCVX1) for a realistic-size patient case

with a variety of nonlinear optimization solvers, including the popular non-convex solver Knitro (Chicago, IL), using a Linux workstation with a 3.50GHz×8 processor and 32 GB of RAM. However, our attempts were unsuccessful: the machine ran out of memory. Next, we consider a more concise formulation that does not require non-anticipativity constraints.

3.6.1 A More Concise Formulation

Instead of keeping track of decision variables in each fraction for each scenario, which results in a large number of variables and non-anticipativity constraints, we use the following more concise formulation. This modification is based on the following observation: if we make x- and z-variables in every fraction explicitly dependent on geometry histories up to and including the current fraction, non-anticipativity constraints are no longer necessary. Let $x^{\omega^{[t]}} \in \mathbb{R}^{|N|}$ and $z^{\omega^{[t]}} \in \mathbb{R}^{|V|}$, for $\omega^{[t]} \in \Omega^t$, $t = 1, \ldots, T$, be the fluence map and dose distribution, respectively, at fraction t with a history of geometries $\omega^{[t]}$. (Here, $x_i(\omega^{[T]}) = (x_i^{\omega^{[t]}}, t = 1, \ldots, T)$ and $z_j(\omega^{[T]}) = (z_j^{\omega^{[t]}}, t = 1, \ldots, T)$.) Then problem (STO-OPT-NCVX1) can be equivalently reformulated as:

$$\text{maximize} \sum_{\omega^{[T]} \in \Omega^T} \left(p_{\omega^{[T]}} \ell\text{EUD}^-_{\text{PTV}} \left(\text{BED}_{\text{PTV}}(z_j(\omega^{[T]})), j \in V_{\text{PTV}} \right) \right) \quad \text{(STO-OPT-NCVX2)}$$

subject to

$$\begin{aligned}
\ell\text{EUD}^+_s(\text{BED}_s(z_j(\omega^{[T]})), j \in V_s) &\leq U^{\text{BED}}_s && \omega^{[T]} \in \Omega^T, s \in S \\
\ell\text{EUD}^+_s(z^{\omega^{[t]}}) &\leq u^{\text{lEUD}}_s && \omega^{[t]} \in \Omega^t, s \in S, \\
& && t = 1, \ldots, T \\
z^{\omega^{[t]}} &= D(\omega^t) x^{\omega^{[t]}} && \omega^{[t]} \in \Omega^t, \\
& && t = 1, \ldots, T \\
\textstyle\sum_{t=1}^T x_i^{\omega[t]} &\leq b && i \in N, \omega^{[T]} \in \Omega^T \\
x^{\omega^{[t]}} &\geq 0 && \omega^{[t]} \in \Omega^t, \\
& && t = 1, \ldots, T \\
x(\omega^{[T]}) &= (x^{\omega^{[t]}}, t = 1, \ldots, T) && \omega^{[T]} \in \Omega^T \\
z(\omega^{[T]}) &= (z^{\omega^{[t]}}, t = 1, \ldots, T) && \omega^{[T]} \in \Omega^T.
\end{aligned}$$

This alternate formulation has $|V| \sum_{t=1}^T K^t + |N| \sum_{t=1}^T K^t$ variables, $|S|K^T + |S| \sum_{t=1}^T K^t + |V| \sum_{t=1}^T K^t + |N|TK^t + |N| \sum_{t=1}^T K^t$ constraints,[‡] and K^T terms in the objective.

As we did with (DET-OPT-NCVX), we can approximate (STO-OPT-NCVX2) with a convex problem by considering only physical dose (instead of the BED) in the objective function measuring PTV coverage (and the same constraints as above):

$$\text{maximize} \sum_{\omega^{[T]} \in \Omega^T} p_{\omega^{[T]}} \ell\text{EUD}^-_{\text{PTV}} \left(\sum_{t=1}^T z_j^{\omega^{[t]}}, j \in V_{\text{PTV}} \right). \quad \text{(STO-OPT-CVX2)}$$

An implementation of even this more compact model with realistic-size instances on our prosumer machine with 64 GB RAM using commercial solvers led to running out of memory. Therefore, in the following section we propose a heuristic approach towards solving

[‡]Not counting the last two sets of constraints, which are included only to make explicit the notational convention.

(STO-OPT-CVX2).

3.7 Diminishing Horizon Heuristic

In this section, we propose a heuristic approach towards solving (STO-OPT-CVX2). We will limit our discussion to models that maximize objective functions based on physical dose to the PTV, and are thus convex. These were the models solved in our computational experiments, although the treatment plans they produced were evaluated using the more accurate BED-based coverage metric.

3.7.1 History-Independent Stochastic Model

The challenge presented by the stochastic models of Section 3.6 is the large number of variables and constraints resulting from considering the time and history in the decision variables. Existing methods in the literature for solving large multi-stage stochastic programs such as progressive hedging (*Rockafellar and Wets*, 1991) require solving instances for each scenario — for even a small realistic example patient, we have 4 possible geometries over 5 fractions, resulting in $4^5 = 1024$ scenarios, each of which takes approximately 10 minutes on a commercial solver (assuming we are using a convex approximation of the objective function), which would take approximately one week to solve if done sequentially.[§]

As a first step towards a heuristic approach for daily offline ART, we first formulate a simplified stochastic model. This model is *history-independent:* in it, each treatment plan $x(\omega) \in \mathbb{R}^{|N|}$ and corresponding dose $z(\omega) \in \mathbb{R}^{|V|}$ depend *only on the geometry* $\omega \in \Omega$ *realized* in a fraction, but *not* on the fraction number, or the history of geometries realized prior to the fraction. In other words, each geometry ω will have a corresponding treatment plan that will be deployed in every fraction this geometry is realized.

Although our decision space is reduced, we still want to be able to account for the

[§]We do not dismiss this approach entirely and leave parallelization of (approximately) solving subproblems for each scenario as a possible future investigation.

potential sequences of geometries in the counterparts to the first set of constraints and the objective of (STO-OPT-CVX2). The history-independent stochastic model that forms the basis of our heuristic is as follows:

$$\text{maximize} \sum_{\omega^{[T]} \in \Omega^T} \left(p_{\omega^{[T]}} \ell\text{EUD}^-_{\text{PTV}} \left(\sum_{t=1}^{T} z_j(\omega^t), j \in V_{\text{PTV}} \right) \right) \qquad \text{(SO-HI-Full)}$$

subject to

$$\ell\text{EUD}^+_s(\text{BED}_s(z_j(\omega^1), \ldots, z_j(\omega^T)), j \in V_s) \leq U_s^{\text{BED}} \qquad \omega^{[T]} \in \Omega^T, s \in S \qquad (3.13)$$

$$\ell\text{EUD}^+_s(z(\omega)) \leq u_s^{\text{lEUD}} \qquad \omega \in \Omega, s \in S \qquad (3.14)$$

$$z(\omega) = D(\omega)x(\omega) \qquad \omega \in \Omega \qquad (3.15)$$

$$\sum_{t=1}^{T} x(\omega^t) \leq b \qquad \omega^{[T]} \in \Omega^T \qquad (3.16)$$

$$x(\omega) \geq 0 \qquad \omega \in \Omega. \qquad (3.17)$$

In this formulation, the superscript t is only present to keep track of realizations of geometries in various scenarios $\omega^{[T]}$; recall that decisions depend on the geometry realized in the fraction, but no longer depend on time. Constraints (3.13) still enforce overall BED limits for any possible sequence of geometries. The number of constraints in (3.13) and (3.16) is quite large. We propose a reduced formulation below, with fewer constraints, based on the idea of "basic scenarios" introduced in *Sir et al.* (2012).

Given a scenario, i.e., a sequence of geometries $\omega^{[T]} = (\omega^1, \ldots, \omega^T)$, let us define a vector of K natural numbers $(N_k(\omega^{[T]}),\ k = 1, \ldots, K)$ where $N_k(\omega^{[T]})$ is the number of times geometry ω_k appears in $\omega^{[T]}$ (and thus $\sum_{k=1}^{K} N_k(\omega^{[T]}) = T$). Note that these vectors are the same for any scenarios that are permutations of one another. According to expression (3.3), BED of a voxel is affected by magnitudes of doses delivered in various fractions, but not the specific order in which they were delivered. Using this observation and the history-independent property of problem (SO-HI-Full) we can therefore specify cumulative

constraints in the model using distinct vectors $(N_k(\omega^{[T]}),\ k = 1, \ldots, K)$ instead of specific scenarios.

We can take this logic one step further and specify these constraints only for K basic scenarios, which *Sir et al.* (2012) define as scenarios where a single geometry is repeated T times (i.e, $N_k(\omega^{[T]}) = T$ for some k). The resulting formulation is

$$\text{maximize} \sum_{\omega^{[T]} \in \Omega^T} \left(p_{\omega^{[T]}} \ell\text{EUD}^-_{\text{PTV}} \left(\sum_{t=1}^{T} \left(z_j(\omega^t) \right), j \in V_{\text{PTV}} \right) \right) \tag{SO-HI}$$

subject to

$$\ell\text{EUD}^+_s(\text{BED}_s(\overbrace{z_j(\omega_k), \ldots, z_j(\omega_k)}^{T \text{ times}}), j \in V_s) \leq U^{\text{BED}}_s \qquad k = 1, \ldots, K,\ s \in S \tag{3.18}$$

$$\ell\text{EUD}^+_s(z(\omega)) \leq u^{\text{lEUD}}_s \qquad \omega \in \Omega,\ s \in S \tag{3.19}$$

$$z(\omega) = D(\omega)x(\omega) \qquad \omega \in \Omega \tag{3.20}$$

$$T \cdot x(\omega_k) \leq b \qquad k = 1, \ldots, K \tag{3.21}$$

$$x(\omega) \geq 0 \qquad \omega \in \Omega. \tag{3.22}$$

Proposition III.1. *Optimization problems* (SO-HI-Full) *and* (SO-HI) *are equivalent.*

Proof. The proof follows the same logic as in Proposition 1 of *Sir et al.* (2012). For simplicity, we focus on constraint (3.18) for one structure and the corresponding family of constraints (3.13), and drop the subscripts s; analysis of constraints (3.16) and (3.21) is similar and simpler.

These optimization problems have the same objective functions; the feasible region of (SO-HI-Full) is contained in (SO-HI) because the latter has fewer constraints (only ones for basic scenarios). It remains to establish the converse. Suppose $(x(\omega) \geq 0, z(\omega) \geq 0,\ \omega \in \Omega)$ is feasible for (SO-HI). In particular, $z(\omega_k)$ satisfies constraints (3.18) for $k = 1, \ldots, K$. Consider an arbitrary scenario $\omega^{[T]}$, which can be characterized by vector $(N_k(\omega^{[T]}),\ k = 1, \ldots, K)$ as discussed above. We will show that $(z(\omega^1), \ldots, z(\omega^T))$ satisfies global BED

constraint (3.13).

Based on (3.2), the BED of a voxel j in this scenario is:

$$\sum_{k=1}^{K} N_k(\omega^{[T]}) \left(z_j(\omega_k) + \frac{z_j(\omega_k)^2}{(\alpha/\beta)} \right),$$

and thus, for a structure's global BED constraint (3.13) for this scenario, we have

$$\begin{aligned}
&\ell\text{EUD}^+ \left(\sum_{k=1}^{K} N_k(\omega^{[T]}) \left(z_j(\omega_k) + \frac{z_j(\omega_k)^2}{(\alpha/\beta)} \right), j \in V \right) \\
=&\ell\text{EUD}^+ \left(\sum_{k=1}^{K} \frac{N_k(\omega^{[T]})}{T} \cdot T \left(z_j(\omega_k) + \frac{z_j(\omega_k)^2}{(\alpha/\beta)} \right) j \in V \right) \\
\leq& \sum_{k=1}^{K} \frac{N_k(\omega^{[T]})}{T} \ell\text{EUD}^+ \left(T \left(z_j(\omega_k) + \frac{z_j(\omega_k)^2}{(\alpha/\beta)} \right) j \in V \right) \\
\leq& \sum_{k=1}^{K} \frac{N_k(\omega^{[T]})}{T} U^{\text{BED}} = U^{\text{BED}},
\end{aligned}$$

where the first inequality is due to convexity of ℓEUD^+, and the second results from applying constraints (3.18) for each k. □

The resulting model in (SO-HI) contains $|V|K + |N|K$ decision variables, $|S|K + |S|K + |V|K + K + |N|K = (2|S| + |V| + |N| + 1)K$ constraints, and K^T terms in the objective. Note the significant reduction in numbers of decision variables and constraints from to those of (STO-OPT-CVX2). A summary of simplifications made to the original stochastic model (STO-OPT-NCVX2) to arrive at the history-independent model (SO-HI) can be found in Table 3.1. A brief discussion of how these simplifications can be addressed in future work is also included.

The following proposition considers the relationship between optimization problems (SO-HI) and (STO-OPT-CVX2).

Proposition III.2. *Optimization model* (SO-HI) *is a restriction of optimization model* (STO-OPT-CVX2).

Proof. Let $z(\omega), x(\omega), \omega \in \Omega$ be feasible to (SO-HI), and thus to (SO-HI-Full). For every sequence of geometries $\omega^{[T]} \in \Omega^T$, let $z^{\omega^{[t]}} = z(\omega^t)$ and $x^{\omega^{[t]}} = x(\omega^t)$ for $t = 1, \ldots, T$. The resulting solution is easily seen to be feasible to (STO-OPT-CVX2), and to have the same objective function value as $z(\omega), x(\omega), \omega \in \Omega$ has in problem (SO-HI). □

3.7.2 Diminishing Horizon Heuristic Based on History-Independent Model

We will use (SO-HI) as the foundation of a diminishing-horizon open-feedback loop control approach to daily offline ART. This approach is similar in spirit to the one in *Sir et al.* (2012). However, since the setting in the latter is one where imaging is done *after* a fraction is delivered, only a single plan was computed, to be used in the next fraction regardless of the realized geometry. Since in our setting we can take advantage of imaging performed just prior to each fraction, we compute, offline, a separate treatment plan for each possible geometry.

The diminishing horizon heuristic works as follows. Before the treatment starts, at "time 0," we solve the problem (SO-HI) to determine plans to be used for every possible geometry encountered in fraction 1. In fraction 1, the realized geometry ω^1 is observed via imaging, and the corresponding treatment will be delivered; let us denote by $(\bar{x}^1, \bar{z}^1)$ the beamlet intensities used and dose distribution delivered in this fraction.

To determine the geometry-dependent plans to be used in the second fraction, we modify the the model (SO-HI): we focus only on scenarios where the geometry realized in the first fraction is ω^1, and we fix values of x and z associated with the first fraction to $\bar{x}^1$ and $\bar{z}^1$, respectively, wherever appropriate. We solve the resulting (smaller) problem, and use its solution to deliver treatment appropriate to the geometry realized in fraction 2. We repeat this process for each remaining fraction by updating the above model with previously treated fractions; since in each step we consider scenarios with shorter sequences of future realizations, the problem instances we consider get smaller.

To formalize the above discussion, suppose in the first τ fractions we observed geometries

ω^t and used beamlet intensities $\bar{x}^t$ to deliver dose distributions $\bar{z}^t$, for $t = 1, \ldots, \tau$. The optimization problem solved prior to fraction $\tau + 1$ is:

$$\text{maximize} \sum_{\omega^{[\tau+1:T]} \in \Omega^{T-\tau}} \left(p_{\omega^{[\tau+1:T]}} \ell\text{EUD}^-_{\text{PTV}} \left(\sum_{t=1}^{\tau} \bar{z}_j(\omega^t) + \sum_{t=\tau+1}^{T} z_j(\omega^t), j \in V_{\text{PTV}} \right) \right) \tag{SO-HI(τ; $\omega^{[\tau]}$))}$$

subject to

$$\ell\text{EUD}^+_s \left(\text{BED}_s(\bar{z}^1_j, \ldots, \bar{z}^\tau_j, \overbrace{z_j(\omega_k), \ldots, z_j(\omega_k)}^{T-\tau \text{ times}}), j \in V_s \right) \leq U^{\text{BED}}_s \qquad k = 1, \ldots, K,\ s \in S \tag{3.23}$$

$$\ell\text{EUD}^+_s(z(\omega)) \leq u^{\text{lEUD}}_s \qquad \omega \in \Omega, s \in S \tag{3.24}$$

$$z(\omega) = D(\omega)x(\omega) \quad \omega \in \Omega \tag{3.25}$$

$$(T - \tau)x(\omega^t) \leq b - \sum_{t=1}^{\tau} \bar{x}^t \quad \omega^{[\tau+1:T]} \in \Omega^{T-\tau} \tag{3.26}$$

$$x(\omega) \geq 0 \qquad \omega \in \Omega \tag{3.27}$$

It is clear that the heuristic ensures that each plan for each geometry at each fraction satisfies per-fraction limits. The plans generated by this heuristic also satisfy overall BED limits because each plan is obtained through an instance of (SO-HI(τ; $\omega^{[\tau]}$)) that accounts for doses delivered so far in previous fractions, and for all possible geometries in future fractions, covering all potential remaining scenarios. As time progresses, the objective function of (SO-HI(τ; $\omega^{[\tau]}$)) requires fewer terms because of the diminishing horizon.

The comparison of the solutions obtained by the diminishing horizon heuristic to the stochastic models of Section 3.6 is not as clear-cut, since by its very nature, the diminishing horizon heuristic only generates plans appropriate to a particular realized scenario. However, one can envision a similar diminishing horizon implementation based on these stochastic models, in which case the recourse problem solved for each partial scenario $\omega^{[\tau]}$ would be a

relaxation of the corresponding instance of (SO-HI(τ; $\omega^{[\tau]}$)).

3.8 Computational Experiment

3.8.1 Representing Interfraction Motion

In order to apply adaptive planning methodology discussed in this chapter, we need to identify the dose deposition matrix $D(\omega)$ associated with every geometry realization incorporated into our model. *Janssens et al.* (2009) show the promise of computing dose accumulation via deformable image registration to track motion at the *voxel level*. To generate matrices $D(\omega)$, the pCT is registered to the image of a potential geometry (e.g., via CBCT or MRI) using MORFEUS (a deformable image registration algorithm based on biomechanical and finite-element modeling, see *Brock et al.*, 2005; *Samavati et al.*, 2015). The resulting deformation fields map the locations of each voxel (this is typically a point in the cube, such as the center or corner) on the pCT to a new location representing a deformed geometry. Recall that dose is energy divided by mass. To track dose to each voxel, we assume conservation of mass of voxels holds, which is consistent with conservation of energy for a fixed amount of dose. (*Zhong and Chetty* (2017) discuss how the law of conservation of energy is violated when there are changes in mass and changes in dose mapping is not properly applied.) A 2D example of voxel assignment is shown in Figure 3.4. This deformation model implies relationship $D(\omega) = B(\omega)D$, where D is the dose depositions matrix from pCT, and $B(\omega)$ is a binary matrix where $B_{ij} = 1$ if voxel i in the pCT moved to the location of spatial voxel j in geometry ω.[¶]

3.8.2 Experimental setup

In this subsection, we describe the characteristics of the patient case, along with some other model specification, in our computational experiment. Table 3.2 shows constraint

[¶] $B(\omega)$ is not necessarily a permutation matrix.

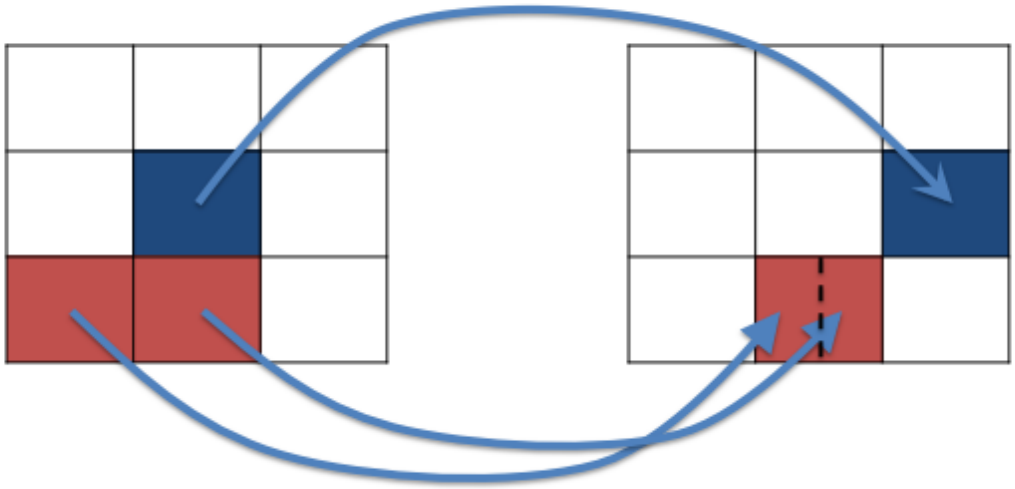


Figure 3.4: 2D example of assignment of dose deposition coefficients for deformed geometries. The 3-by-3 grid represents static voxels corresponding to dose deposition coefficients (D_{ij}). In this example, the blue voxel moves from the center to the right middle location and is assigned the coefficients corresponding to the latter location. The red voxels are compressed and both will have the dose coefficients of the bottom center static voxel.

parameters used for an example liver patient. We considered $T = 5$ (typical for liver SBRT treatment) and derived values of U_s^{BED} from physical dose limits from clinical protocol at the University of Michigan Department of Radiation Oncology.[‖] Values of $u_s^{\ell\text{EUD}}$ were chosen to be 10% higher than typical per-fraction physical dose limits to allow flexibility in boosting dose. The number of voxels in the patient was 15,615 and the number of beamlets was 989.

We considered $|\Omega| = 4$. For simplicity, we assumed each geometry to be equally likely, i.e., $p_{\omega^{[T]}} = \left(\frac{1}{|\Omega|}\right)^T = \frac{1}{1024}$.[**] Starting with the nominal geometry based on pre-treatment imaging, we applied 3 synthesized deformations, based on observations of a patient case, to generate the set Ω.[††] In ω_2, the liver deforms into what would be a higher dose region in

[‖]The total physical dose limits were divided into per fraction doses assuming equal fractions, and together used to obtain BED limits

[**]Recall that in Diminishing Horizon Heuristic, these probabilities are updated in the objective function after each fraction, e.g., after fraction 1, the number of remaining scenarios is 1024/4=256 scenarios, etc.

[††]One of these geometries, ω_2, was derived by applying a realistic deformation to this patient liver, but

Simplification	Impact	Future Work
Countably many geometries	Countably many plans to check; makes problem tractable	Patient geometry likely will not exactly match ones we considered, but we can also slightly modify one of the optimized plans; robust approach to account for variation around each extreme geometry
Equally likely geometries	Simplifies objective function	Base probability of geometries on historical data, biomechanical models for abdominal structures
Symmetric scenario tree	Set of patient geometries is static over fractions	Set of patient geometries and their probabilities may be updated fraction to fraction via a statistical learning model, implying the uncertainty is *fraction-dependent*
Convex objective	Reduces problem size (number of constraints, terms in objective, and variables); models efficiently solvable to global optimality	Develop a global optimization method to solve original non-convex problem

Table 3.1: Summary of simplifications made in some of our models, their effects, and potential for future work.

a treatment designed only for the nominal geometry, in which case it would exceed mean per-fraction dose bounds; in ω_3, we artificially moved the duodenum to compress against the PTV; in ω_4, the duodenum is farther away from the PTV and other OARs are within safe distances, allowing for potential boosting of dose to the PTV.

Structure s	λ_s^+ for ℓEUD_s^+	$(\alpha/\beta)_s$	$u_s^{\ell\text{EUD}}$	U_s^{BED}
Chest Wall	0	2.5	7.7	133
Duodenum	0	2.5	6.6	102
Heart	0	2.5	11.5	273
Kidney-Left	0	2.5	3.3	30
Liver	1	2.5	3.5	135
PTV	0	10	17.6	208
Stomach	0	2.5	6.6	102
Colon	0	2.5	7.0	114
Spinal Cord	0	2.5	5.5	75

Table 3.2: Structures and their parameters for patient liver case. For measuring PTV coverage, $\lambda_{\text{PTV}}^- = 0.84$.

Structure	Metric	Fx - Limit	ω_1-Nominal	ω_2	ω_3	ω_4
Chest Wall	Max	7.00	7.00	7.00	7.00	7.00
Duodenum	Max	6.00	3.73	3.24	16.00	3.44
Heart	Max	10.45	0.17	0.31	0.17	0.17
Kidney-Left	Max	3.00	0.88	0.65	0.88	0.88
Liver	Mean	3.50	3.50	3.88	3.50	3.50
Stomach	Max	6.00	3.84	3.84	3.84	3.84
Colon	Max	6.40	6.36	6.36	6.36	6.36
Spinal Cord	Max	5.00	1.76	1.76	1.76	1.76

Table 3.3: Per-fraction structure criteria resulting from the application of the nominal plan to each potential geometry (ω_1 is nominal). Intolerable levels of dose are in red.

We implemented all models using AMPL (Albuquerque, NM) and solved all instances with Mosek (*Mosek*, 2010), a general-purpose convex programming solver.

3.8.3 Results

To illustrate the potential importance of adaptive planning, we first obtained the nominal plan by solving an instance of (DET-OPT-CVX) using the nominal geometry in every

the deformation was observed on a different patient.

fraction. We then applied the nominal plan to the other three geometries, and calculated relevant metrics of the resulting dose distributions. Table 3.3 shows the per-fraction effects of applying the nominal treatment plan to a different patient geometry.

To analyze performance of the Diminishing Horizon Heuristic (DHH), we applied the heuristic to every possible scenario. Recall that, due to the robust nature of the constraints in our models, the treatment is guaranteed to be safe (i.e., meet dose constraints) for all OARs in all fractions and all scenarios. Therefore, we focus on studying PTV coverage (using the true measure of coverage from (3.7)) of plans obtained via the DHH in each scenario. As a benchmark, if the nominal geometry occurred at each fraction and we delivered the nominal plan, the coverage achieved would be 204.2 Gy. Our findings are summarized in Figure 3.5.

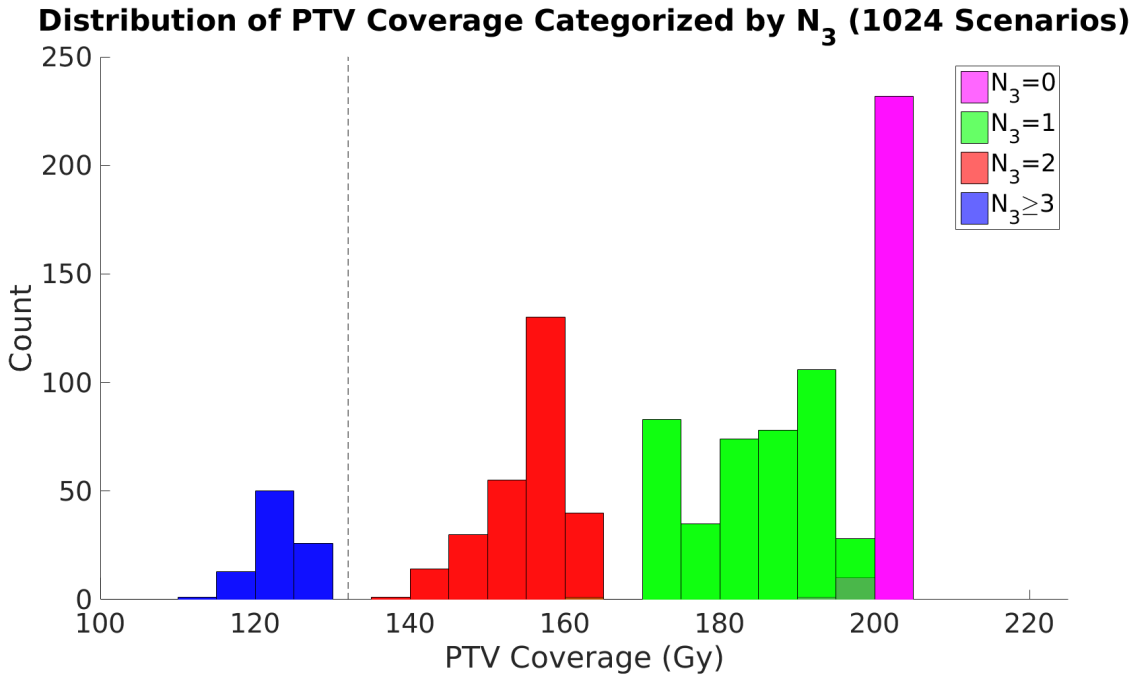


Figure 3.5: Distribution of PTV Coverage achieved by plans generated via DHH. In 90% of scenarios were able to meet the typical prescription dose of 132 Gy (black dashed line). Distribution is also categorized by number times the most limiting geometry, ω_3, occurred (bar colors).

Observe that since the maximum PTV dose at each fraction is bounded, we can obtain an upper bound on achievable PTV coverage. With a per-fraction bound of 17.6 Gy to each

τ	Number of Instances	Average Runtime (seconds)
0	1	17,300
1	4	4,900
2	16	3,400
3	64	2,600
4	256	500

Table 3.4: DHH's average instance runtime by length of history τ. $K = 4; T = 5$.

PTV voxel, the cumulative BED is:

$$5\left(17.6 + \frac{17.6^2}{10}\right) = 242.9 \text{ Gy}.$$

Moreover, our constraints include an upper bound on cumulative maximum PTV BED of 208 Gy (to constrol for hot spots), which is more restrictive than the above upper bound. Clearly, this bound is likely to be loose as rarely will all voxels accumulate this BED; most scenarios in Figure 3.5 achieve less coverage than this. On the other hand, for scenarios with $N_3 = 0$, our heuristic was able to achieve close to this level of coverage.

3.9 Discussion

Daily offline adaptive planning provides a library of plans for each potential geometry that arises. Our approach takes into consideration all combinations of potential geometries throughout treatment, resulting in an adaptive strategy that is *not* dependent on the nominal geometry and nominal treatment plan. This flexibility in adapting the dose distribution spatially (as opposed to shifting a pre-defined dose distribution) i) guarantees feasibility of both cumulative and per-fraction structure dose upper bounds, and ii) allows for boosting dose in fractions with favorable geometry, which can compensate for fractions with unfavorable geometry, should they occur. The reader may wonder if it is worth updating the model each fraction and applying updated plans over solving the problem (SO-HI) and simply re-using those plans. In Figure 3.6, we compare the quality of PTV coverage achieved by using plans

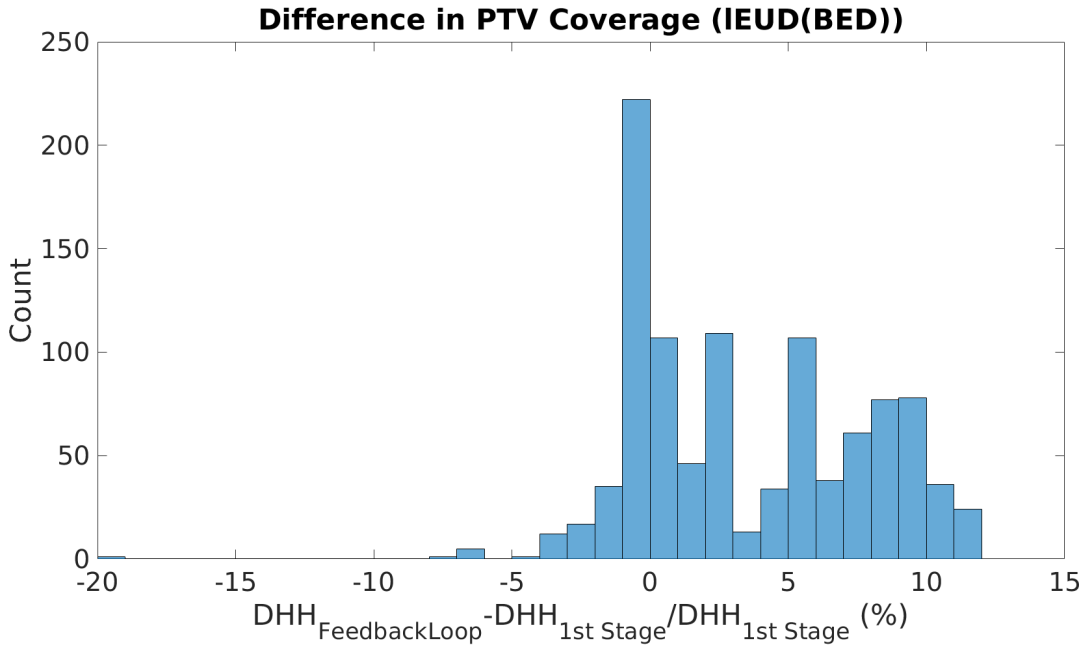


Figure 3.6: PTV Coverage comparison: applying DHH versus using solution of (SO-HI), a.k.a. $\text{DHH}_{\text{1st Stage}}$.

obtained by applying DHH to the coverage obtained using only the solution of (SO-HI). In over 90% of scenarios, the former achieves significantly better PTV Coverage.

Recall that IGRT relies solely on the nominal plan for the nominal geometry. Since there is no boosting of dose, if, based on the initial imaging, ω_3 were designated as the nominal geometry, then the coverage from the resulting plan would be low throughout the treatment horizon no matter which favorable geometry is realized after planning. Specifically, in this patient case, the nominal PTV coverage would be 63.3 Gy (in BED), which is an upper bound on coverage since in all other geometries, no structures are in danger and IGRT would not adapt. This would result in significant loss of valuable PTV coverage that could compensate fractions with poor coverage. However, in DHH, as seen in Figure 3.5, if ω_3 occurs at most 2 times throughout treatment, we are able to still achieve prescription dose thanks to boosting.

Although for evaluation purposes we considered all possible scenarios, a practical application of DHH would require significantly less computation. In particular, only *one* instance

of the model would need to be solved before each of T fractions, K plans per model would be obtained — one for each geometry. Whichever plan was delivered would then be used to update the instance used for the following fraction. Therefore, at each fraction, only K plans would need to go through quality assurance (QA) and leaf-sequencing (LS) before the patient arrives, resulting in a total of TK plans to be processed throughout the treatment horizon. This addresses the issue that daily replanning and QA lengthens the patient's visit.

Suppose daily replanning and QA were efficient procedures. The current replanning paradigm is myopic and adapts plans by assuming the observed geometry at the current fraction is the geometry for all remaining fractions. Although this updates information for the current fraction, our model is still more informative, realistic, and robust than replanning due to considering geometry uncertainty in future fractions.

Our model assumes treatment at each fraction, but there are also rare occasions when patients arrive with geometries that are not worth treating (e.g., ω_3 could be considered an example of such geometry). One future direction of this work is to consider the option of postponing treatment when such geometries arise. This also leads to questions such as, "what makes it 'worth' treating a patient on a particular day?" and "should skipped days be compensated later on, e.g., by allowing for more aggressive treatment (e.g., adapting per-fx bounds) in later fractions? if so, by how much?" Furthermore, in SBRT, or even conventional treatment schedules (30 fractions), the distribution of geometries can be updated and learned as treatment progresses: if favorable geometries are more likely, then skipping unfavorable geometries could be a risk worth taking, or if unfavorable geometries are more likely, then each fraction should be treated since only so much coverage can be achieved anyway.

3.9.1 Unique Number of Scenarios in DHH

Since the order in which geometries in future fractions occur does not affect the corresponding terms of the objective functions of (SO-HI) or (SO-HI(τ; $\omega^{[\tau]}$)), we can simplify these functions by considering the classic "Stars and Bars" problem (*Feller*, 1950). For ex-

ample, the objective function of (SO-HI) can be reduced from K^T to $\binom{T+K-1}{K-1}$ terms. When $T = 5$ and $K = 4$, this would mean a reduction from 1,024 to 56 terms. We found that this model reduction was not necessary in our experiments (recall the instance with no history took 17,300 seconds), but it may help to speed up computation when solving problems with a larger K or T (e.g., with conventional fractionation).

3.10 Conclusion

The proposed ART framework considers all potential geometries in all fractions. This look-ahead approach creates plans for those geometries that will always be feasible and maximize PTV coverage no matter what sequence of geometries is realized. Although the true stochastic optimization model presents computational challenges, a few simplifications leads to a tractable model (summarized in Table 3.1) while still achieving high treatment plan quality. One such simplification is considering an uncertainty set with a small number geometries; the true uncertainty set is realistically larger. In practice, the uncertainty set considered should cover a variety of extreme geometries so that a patient would arrive with a geometry close to one of them. Unlike IGRT, our framework does not rely on the geometry that the patient has when initially being imaged for treatment planning (i.e., the nominal geometry). This added flexibility allows for a set of treatment plans, each catered toward a potential geometry, to achieve the best PTV coverage possible. Finally, this approach does not require drastic change of treatment planning workflow (our heuristic can be done offline) and adds little additional effort in the clinic during the actual time of treatment.

CHAPTER IV

Multicriteria Optimization for Brachytherapy Treatment Planning

4.1 Introduction

In brachytherapy (BT) treatment planning, the result of a single application of commercial treatment planning software is often not the final plan that ultimately gets delivered. Commercial treatment planning systems usually produce plans based on the goal values of Dose Volume Histogram (DVH) metrics specified in the clinical protocol. If not all goals are simultaneously achievable (i.e., feasible), the physician must make trade-offs among feasible plans. Current commercial systems require the physician to guess-and-check input parameters (e.g., goal doses or objective weights) to explore trade-offs between candidate plans — a time-consuming process that is not intuitive. Because treatments need to be made essentially in real time (i.e., based on images taken after the patient has been put under anesthesia, planned and delivered while the patient is still under anesthesia), the physician does not have an opportunity to fully consider all trade-offs and may forgo promising plans.

The goal of this chapter is to provide an intuitive guided user interface for directly navigating trade-offs of DVH metrics among feasible plans for BT planning. Generating a dose-volume metrics trade-off surface requires solving several instances of a non-convex model, each of which is a mixed-integer program (*Lee et al.*, 2003; *Romeijn et al.*, 2003, 2006).

Due to the size of instances in radiotherapy (both external and interal source) treatment planning, they cannot be efficiently solved, especially given the limited time available for treatment planning in BT. Constraints on DVH metrics may also be familiar to the reader as *value-at-risk* constraints (*Romeijn et al.*, 2003).

4.1.1 Related Literature

Many aspects of BTTP optimization (deciding catheter locations and/or dwell times) have been studied. For work on source placement optimization and permanent implants, i.e., sources with short half-lives, see, e.g., *D'Souza et al.* (2001); *Lee and Zaider* (2003); *Lee et al.* (1999). We will focus on HDR-BT planning, which uses temporary implants. *Milickovic et al.* (2002) propose a multicriteria model that optimizes both dwell times and catheter selection for prostate cases by enforcing DVH criteria for problem with 3 structures and no more than 3000 voxels, and stop their MIP solver once they are satisfied with PTV coverage. Although solution times for these instances range from under one minute to 15 minutes, it is unclear how their method scales with more voxels and structures; although their model achieves desirable DVH criteria in their example patients, the variability in solution time also suggests it's unfit for trade-off surface generation. Hybrid Inverse Planning Optimization (HIPO) (*Karabis et al.*, 2005, 2009) uses voxel-based penalties, with a logistic function instead of the step function (1 if upper or lower bound doses are exceeded and 0 otherwise) is applied to smooth the objective. A simulated annealing-based heuristic is used to solve their model to a local optimum. *Giantsoudi et al.* (2013) propose a model that uses gEUDs to mimic DVH metrics, which was competitive or outperformed HIPO, and recommend gEUD parameters to use in BTTP to achieve higher dose conformity, decreased OAR dose, but also decreased PTV coverage. Although these approaches show promise in achieving high quality DVH metrics, there is a lack of discussion of how to control for trade-offs when DVH goals from clinical protocol are not simultaneously achievable. For other BT planning approaches, see *Dinkla et al.* (2015); *Lahanas et al.* (1999, 2003).

There has also been extensive research in multicriteria optimization (MCO) for external beam RT, especially in IMRT (e.g., see *Bokrantz and Forsgren*, 2013; *Breedveld et al.*, 2009; *Craft et al.*, 2005, 2012; *Long et al.*, 2012; *Romeijn et al.*, 2004). We highlight a few of the popular approaches because the multicriteria optimization concepts are analogous to multicriteria BTTP. *Craft et al.* (2005) discuss how much time can be saved if a planner uses an MCO-based planning system. *Romeijn et al.* (2004) propose a framework for multicriteria problems based on convex criteria and provide conditions under which non-convex treatment criteria can be transformed into equivalent convex ones (e.g., tumor control probability, normal tissue complication probability, gEUD, EUD), and state that criteria such as dose-volume metrics are "fundamentally nonconvex, i.e., transformations to convex criteria... do not exist." *Long et al.* (2012) rank the criteria in a specific order and consider tradeoffs between two consecutive criteria at a time (called a stage) based on this ranking. Once values are chosen for the metrics for the criteria considered in the current stage, those metrics are bounded by these values for subsequent stages, where other criteria that are lower in the ordering are considered. Providing a physician with two criteria to trade off at a time makes the task intuitive and simple. However, once a criterion's value is bounded, it is difficult in this process to return to that criterion to explore other potential trade-offs. *Breedveld et al.* (2009) show the equivalence between the weighted-sum method and ϵ-constraint method with respect to the Pareto-optimal solutions they generate for convex problems and explain how to find the relevant parameters in one formulation given a solution from the other. They build on a two-phase ϵ-constraint method (see *Breedveld et al.*, 2007) that is an auto-planning method, where prioritization of criteria and specification of goal doses for those criteria are required inputs. Once a candidate plan is found, it may be that the planner desires to re-adjust this plan (e.g., to adapt), which can be done by i) finding the equivalent weighted-sum problem that achieves the candidate plan and adding a constraint to reflect the minor adjustment to obtain a final plan.

There have also been a few interactive GUIs developed for (mostly) external beam treat-

ment planning (see, e.g., *Craft et al.*, 2007; *Monz et al.*, 2008; *Ripsman et al.*, 2015; *Thieke et al.*, 2007) *Craft et al.* (2007) introduce a GUI involving sliders for trading off multiple EUD criteria. *Ripsman et al.* (2015) develop an interface for Gamma Knife (Elekta, Stockholm) treatment planning. *Monz et al.* (2008); *Thieke et al.* (2007) propose an interface for IMRT planning, which consists of a "navigation star" that contains an axis for each structure criterion and an algorithmic foundation for generating the set of plans to be browsed by the planner on the star. The interface limits the user to certain portions of the star to limit the ranges of values of each axis (i.e., metric). However, no pairwise metric trade-off information is portrayed (or even obtainable). *Ruotsalainen et al.* (2010) consider MCO problems for HDR-BT, but do not explain their claim of obtaining Pareto-optimal solutions when non-convex objectives are included.

To approximately convexify DVH, or VaR, constraints in IMRT planning, *Romeijn et al.* (2003) is the first to propose using *conditional value-at-risk* (CVaR) (a distribution's tail average, instead of a quantile) constraints in an RTTP model. *Clark et al.* (2008) incorporate CVaR metrics in a prioritized optimization framework and propose finding a quantile that correlates well to rectum V40 (percent volume receiving at least 40% of the prescribed dose) according to their clinical archive of prostate treatment plans, suggesting that good choices of CVaR quantiles may *not* be the same as VaR quantiles, and existence of an "ideal" quantile parameter (or range of parameters) when using CVaR approximations. *Holm et al.* (2013) extended this approach to brachytherapy in conjunction with an objective function to improve homogeneity of dose in the PTV. Finally, *Engberg et al.* (2017) have proposed using CVaR objectives in a multi-objective framework for IMRT and recommend this approach over conventional voxel-based quadratic penalties. However, their choice of parameters for CVaR to approximate VaR only provides bounds on, rather than direct control of, VaR metrics, which may be too conservative and exclude plans that achieve better values, especially if the framework transitions to HDR-BT.

Our work makes the following additional contributions:

- Parameter choice. Quantile parameters for CVaR in previous works have been manually tuned, and there has been no general discussion of what are good choices; we conduct such an analysis and make recommendations on appropriate choices in the context of BT for the breast treatment site. We add to our test bed an example case for the prostate and for cervix treatment sites. All investigations are empirical.

- Improving CVaR approximation. In BT, certain structures (e.g., PTV) may have long upper tails in their DVHs, causing the mean-tail-dose to be skewed by extremely high doses. We propose an improvement in convex approximation to VaR by excluding voxels likely to receive high doses, resulting in a modified CVaR called *truncated CVaR* (TCVaR).

- Trade-off surface generation. We provide a process for "sketching" a trade-off surface of relevant DVH metrics (for the general case of 3 or more metrics), with controlled spacing between plans, and describe an interpolation of dwell-times for further populating the surface with more plans without re-optimization, while preserving plan quality.

- Displaying and navigating the trade-off surface. Due to the limited number of criteria considered in HDR BT planning (typically no more than 6), we are able to show more information for making trade-offs. We introduce a guided user interface that is simple and intuitive for the physician to easily make trade-offs with controlled step sizes during real-time planning by providing local information on pairwise metric trade-offs surrounding the current plan. The plan chosen is deliverable, i.e., it is composed of feasible dwell times.

4.2 Methods and Materials

4.2.1 Patient Cases

Our treatment planning approach is retrospectively tested on breast, cervix, and prostate treatment sites (3 of each, differing in geometry). Tables 4.1–4.3 show typical DVH metric goals for clinical protocols at the Mayo Clinic (Rochester, MN) for these sites. The expression "Dq," where $q \in (0, 100)$ is the dose to q% of the structure volume while "D$x$cc" is the dose to $x \geq 0$ cubic centimeters of the structure.

DVH Metric	Breast ($\times$ Tx Dose)
Tx Dose (cGy)	[700, 340, 400]
PTV D95	$\geq 100\%$
PTV D15cc	$\leq 200\%$
Pectoralis Max	$\leq 125\%$
Chest Max	$\leq 100\%$
Skin Max	$\leq 100\%$

Table 4.1: DVH Metric Clinical Protcol (% Tx Dose) for Breast Cases. Tx Dose = Prescribed dose to 95% of the tumor volume. There are three Tx doses listed, one for each breast case tested.

DVH Metric	Cervix
Tx Dose (cGy)	700
PTV D95	$\geq 100\%$
Bladder D2cc	$\leq 80\%$
Rectum D2cc	$\leq 70\%$
Small Bowel 2cc	$\leq 50\%$
Large Bowel 2cc	$\leq 70\%$
Ovoid-L and R Max	$\leq 200\%$

Table 4.2: DVH Metric Clinical Protcol (% Tx Dose) for Cervix Cases. Tx Dose = Prescribed dose to 95% of the tumor volume.

4.2.2 Trade-off Surface

Our ultimate goal is to design an intuitive GUI to help the physician in making trade-offs among several treatment planning criteria, or metrics. This would require construction of a

DVH Metric	Prostate
Tx Dose (cGy)	1500
PTV D1cc	$\leq 400\%$
PTV D95	$\geq 100\%$
Bladder D1cc	$\leq 75\%$
Rectum D1cc	$\leq 75\%$
Urethra D10	$\leq 115\%$
Urethra Max	$\leq 125\%$

Table 4.3: DVH Metrics Clinical Protocol (% Tx Dose) for Prostate Cases. Tx Dose = Prescribed dose to 95% of the tumor volume.

trade-off surface of treatment plans in the space of dimension equal to the number of metrics considered. The desired qualities of the trade-off surface and plans that populate it include:

1. The trade-off surface should be non-dominated, i.e., any plan on the surface cannot be improved in one metric without doing worse in some other metric.

2. To navigate the trade-off surface, the clinician should be able to choose one metric to improve/worsen as a trade-off with one or more other metrics (to worsen or improve, respectively).

3. When considering trade-offs resulting from changing a particular metrics, the clinician should be able to choose a "step size," i.e., the magnitude of desired change in this metric, while the consequent changes in the other metrics may vary (since trade-off rates are not known a priori).

4. Although the clinician should be able to consider trade-offs resulting from very small changes, the trade-off surface has to be generated within limited time allowed for computation (e.g., solving optimization problems), and thus has to be represented by only a finite (and fairly small) number of distinct plans. Other plans on the trade-off surface should be generated "on-the-fly," with minimal computational effort.

5. All treatment plans generated in the process of exploring the trade-off surface should be deliverable, i.e., there should exist a set of dwell times that achieve the metric values

selected on the trade-off surface. Moreover, and the dwell times for the chosen plan should be easy to compute.

Figure 4.1: Screen capture of example GUI. Left: trade-off(s) can be made by selecting which criterion to "heat" (increase dose) vs "cool" (decrease dose) at the expense of another criterion (other criteria) by the green bar selection; Right: Visualization of DVH, dose distribution, and dwell times corresponding to the current plan on the trade-off surface.[a]

[a]This GUI was conceptualized by a certified medical physicist.

An example of a possible implementation of such a GUI is depicted in Figure 4.1 for a case with three metrics of interest concerning PTV, urethra, and rectum. On the right-hand side are dose washes, the DVH, and dwell times corresponding to the current plan in the trade-off space. On the left-hand side are 3 rows, each corresponding to a structure-and-metric pair of interest (called "structure of interest" for brevity). In the leftmost column are values of the current metric and buttons "Heat" and "Cool," which are clinical jargon to increase and decrease the metric, respectively. The metric is changed by trading off with at least one

other metric, which can be chosen by selecting one or multple green bar(s) in the middle columns of the corresponding row. Also in the middle columns are graphs: each graph is a trade-off line plot between the row's structure of interest and the column's structure of interest. Each of these plots relates how two metrics vary on the trade-off surface while *all other metrics are held constant.*

4.2.3 Notation

We use notation similar to what is found in *Deufel and Furutani* (2014). Let S be the set of potential dwell locations and V — the set of all voxels. Let $d \in \mathbb{R}^{|V|}$ be the dose distribution, $s \in \mathbb{R}^{|S|}$ — the source strengths, and let $G \in \mathbb{R}^{|V| \times |S|}$ denote the dose deposition matrix, where an entry G_{ji} is the dose deposited in voxel j from source i at unit source strength. Similarly to IMRT, d and s are related by:

$$d = Gs. \tag{4.1}$$

Let $\vec{f} = (f_0, f_1, \ldots, f_K) : \mathbb{R}_+^V \to \mathbb{R}_+^{K+1}$ be a vector-valued function of d, with $K+1$ components ($K \geq 1$), where each component corresponds to a metric reflecting a criterion of interest. Let $V_k \subseteq V,\ k = 0, \ldots, K$ the set of voxels relevant to criterion k.

4.3 Representing and Navigating the Trade-off Surface

Because of the need for real-time BT planning, it is not possible to solve an optimization problem for every plan that will be considered on the trade-off surface. As a compromise, we propose obtaining, via optimization, a finite subset of the trade-off surface with sufficient granularity to "span" the surface; the choice of the level of granularity can be made by the user. This initial set of plans can then be used to navigate the continuous trade-off surface via interpolation.

Suppose that we have obtained the initial finite set, or *library* of plans $\mathcal{L}$. For each plan $l \in$

$\mathcal{L}$, s^l is the vector of dwell times of this plan, and $\vec{f^l} \equiv \vec{f}(Gs^l) = (f_0(Gs^l), f_1(Gs^l), \ldots, f_K(Gs^l))$ denotes the values of metrics of interest. We assume that all plans in the library are non-dominated, and thus corresponding metric vectors belong to the tradeoff surface. In this section, we will discuss an interpolation methodology that can be used to quickly "fill out" the surface based on this library of plans without solving additional optimization problems, and a process a clinician can use to navigate the surface while exploring trade-offs between the metrics; the process we used to generate the library is the topic of the following section.

4.3.1 "Filling out" the Trade-off Surface via Interpolation

As before, let $\mathcal{L}$ be the finite library of plans such that points $\vec{f^l},\ l \in \mathcal{L}$ belong to the trade-off surface. Since this library is likely to be rather small, these points provide a fairly sparse representation of the surface. To "fill in," or provide a denser representation of the surface, we propose the following technique.

Let $L = \{(f_1^l, \ldots, f_K^l),\ l \in \mathcal{L}\} \subset \mathbb{R}_+^K$ be the set of values of metrics 1 through K of the library plans, and compute a triangulation of $\mathbb{R}^K$ with vertices in L. Although a variety of triangulation methods exist, we chose the Delaunay triangulation method, which typically forms triangles that are as close to equilateral as possible, as opposed to "thin" scalene ones. Let $\psi^{\mathrm{TRI}(L)} : \mathrm{conv}(L) \to L^{K+1} \times [0,1]^{K+1}$ be an oracle that, given a point $q \in \mathrm{conv}(L)$, returns the $K+1$ vertices of tetrahedron $t(q) \in L^{K+1}$ that contain q, according to the triangulation, and $b_{t(q)} \in [0,1]^{K+1}$ that are the corresponding barycentric coordinates, so that

$$q = \sum_{i=1}^{K+1} b_{t_i(q)} t_i(q). \tag{4.2}$$

We implemented this oracle using MATLAB's `delaunayTriangulation` class which contains a method called `pointLocation` that returns the tetrahedron containing a user-specified point.

We can now easily easily construct a dense approximation of the trade-off surface (see Algorithm 1). Let $Q \subseteq \mathrm{conv}(L) \subset \mathbb{R}_+^K$ be a set of points in the space of metrics 1 though

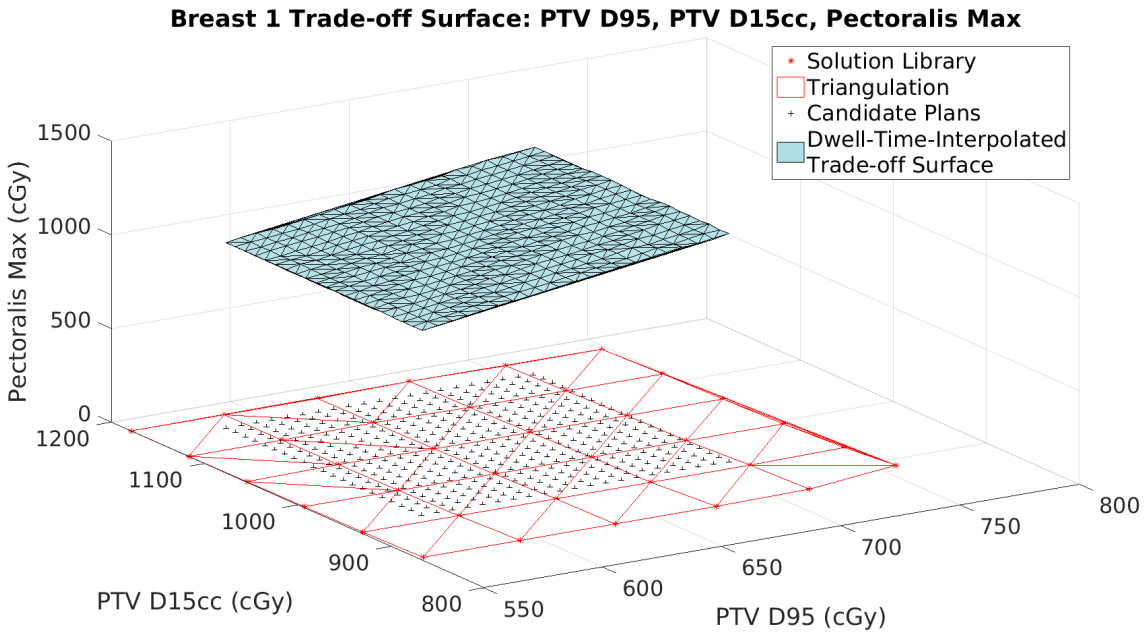


Figure 4.2: Populating a trade-off surface for the example breast case. Red asterisks are plans generated from optimization (points in $\mathcal{L}$ projected to form L; in this example, PTV D95 and PTV D115cc are the metrics used in defining L); a triangulation is generated based on these plans (red lines). Black crosses represent the set of additional plans, Q, used to increase the density of our representation of the trade-off surface; using the triangulation, the dwell times of the library plans are interpolated to create plans on the blue trade-off surface, which is plotted as a mesh.

K. For each $q \in Q$, we can obtain a deliverable treatment plan by interpolating dwell times based on the triangulation:

$$s^q = \sum_{i=1}^{K+1} b_{t_i(q)} s^{t_i(q)}. \tag{4.3}$$

Moreover, the point $(f_0(Gs^q), q) \in \mathbb{R}^{K+1}$ is an approximation of a new point on the trade-off surface based on this interpolation.

Generate a triangulation TRI(L).
for $q \in Q$ **do**
  Compute $b_{t(q)}, t(q)$ using $\psi^{\mathrm{TRI}(L)}$
  Interpolate dwell times of plans $t_i(q)$ according to (4.3)
  Add point $(f_0(Gs^q), q)$ to the approximate trade-off surface
end

Algorithm 1: Trade-off Surface Generation

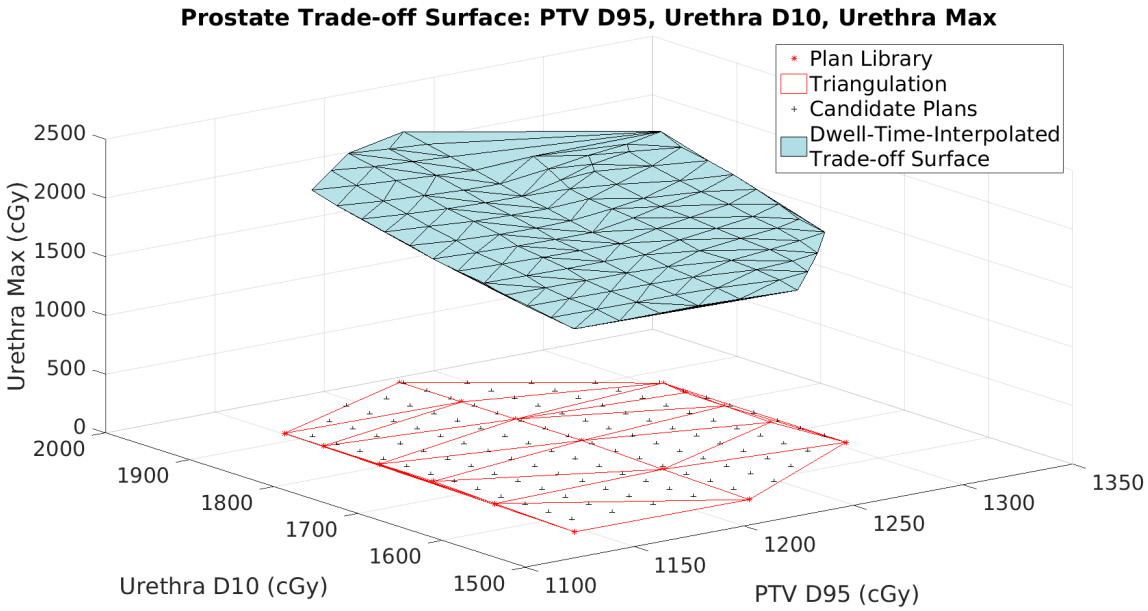


Figure 4.3: Populating a trade-off surface for the example breast case. Red asterisks are plans generated from optimization (points in $\mathcal{L}$ projected to form L; in this example, PTV D95 and Urethra D10 dose are the metrics used in defining L); a triangulation is generated based on these plans (red lines). Black crosses represent the set of additional plans, Q, used to increase the density of our representation of the trade-off surface; using the triangulation, the dwell times of the library plans are interpolated to create plans on the blue trade-off surface, which is plotted as a mesh.

4.3.2 Navigating the Trade-off Surface

Here, we describe a process that can be used to explore trade-offs between various metrics using the representation of the trade-off surface described above. The user begins at an initial plan l and corresponding point $\vec{f}^l$ on the trade-off surface. Suppose the user wants to explore the possibility of finding a plan with better values of metric f_{imp} (for ease of exposition, we assume that improvement corresponds to smaller values of the metric, although, depending on the nature of the criterion represented by this metric, this may correspond to, in the clinical jargon, "heating" or "cooling" of the underlying structure). For example, for the current plan in Figure 4.1, the user may want to improve the metric Target D95. Then, the user would select some of the remaining metric indices $W \not\ni \text{imp}$ to "worsen" (i.e., to trade off against f_{imp}); in the figure, the selected metric is Rectum D2cc, indicated by the highlighted green bar. Metrics that are not included in W are meant to remain constant during this trade-off exploration. It is possible that their values will deviate slightly in practice, but, as demonstrated in our empirical results, those deviations were within 1% of the specified values.

The "increment size," denoted $\eta * 100\%$ (0.1% in Figure 4.1), refers to the magnitude of increase (i.e., deterioration) in the selected metrics $f_i,\ i \in W$ that the user wants to allow in one step of trade-off exploration. After choosing which metrics to improve, worsen, or hold constant, and the magnitude of trade-off, we let $f_0 = f_{\text{imp}}$ and construct the triangulation and perform the interpolation according to Algorithm (1) to find the dwell times s^{new} and metric values l^{new} resulting from the trade-off. Specifically, let $p = (f_i(s^l), i \neq \text{imp}) \in \mathbb{R}^m$ and compute:

$$s^{\text{new}} = \sum_{i=1}^{m} b_{t_i((1+\eta e^{\text{wor}})p)} s^{t_i((1+\eta e^{\text{wor}})p)} \tag{4.4}$$

$$l^{\text{new}} = f(Gs^{\text{new}}) \tag{4.5}$$

where $e_i^{\text{wor}} = 1$ if $i \in W$ and 0 otherwise. The user can refer to the trade-off graph in Figure

4.1 to anticipate the trade-offs between the selected metrics, or use the "Heat" or "Cool" button corresponding to f_{imp} to move to the modified plan, and view the DVH plots and dose contour figures in the right-most pane of the GUI.

This process then can be repeated for multiple increments, as well as for other metrics to be improved. It is more important to be able to control how much metrics worsen than how much one metric improves. In this case the improvement is less predictable, but can be controlled by adjusting the step size appropriately.*

4.4 Generating Initial Library of Plans via Convex Optimization

In the previous section, we have discussed methods for generating and navigating a trade-off surface based on the initial finite library of plans. In this section, we discuss an approach for generating such a library.

4.4.1 Metrics: Value at Risk and a Convex Approximation

Many metrics used to evaluate treatment plans in BT are so-called *DVH metrics*, also known as *value-at-risk (VaR)*. Recall that these are quantiles of the dose vector to the structure in question, and thus are non-convex functions of d. We introduce a related, but convex, metric called *conditional value-at-risk (CVaR)*, which was first applied to IMRT treatment planning by *Romeijn et al.* (2003). Let index k refer to the criterion in question, and recall that V_k is the set of voxels relevant for this criterion. Given dose distribution d, we define for a structure k the Δ upper value-at-risk (VaR) as:

$$\mathbf{VaR}^+ : \ \delta_k^+(d; 1-\Delta) = \min_{\nu_k} \left\{ \nu_k \in \mathbb{R} : \frac{|j \in V_k : d_j \geq \nu_k|}{|V_k|} \leq 1-\Delta \right\} \tag{4.6}$$

*Though this is a trial-and-error procedure, as long as the step size is not large, the trade-offs can be controlled and a new plan is immediately obtained via interpolation, as opposed to waiting for a new optimization.

Sorting voxels in order of dose, this is the $100(1-\Delta)\%$ hottest dose. The upper CVaR at Δ is defined as the mean of all doses that exceed the upper Δ-VaR:

$$\textbf{CVaR}^{+}: \min_{\nu_k}\{\overline{\delta}_k^{+}(d,\nu_k;1-\Delta)\} = \min_{\nu_k}\left\{\nu_k + \frac{1}{(1-\Delta)|V_k|}\sum_{j\in V_k}\max\left(0,d_j-\nu_k\right)\right\}. \quad (4.7)$$

Note that CVaR is a piecewise-linear convex function in the arguments d, ν_k (*Pflug*, 2000).

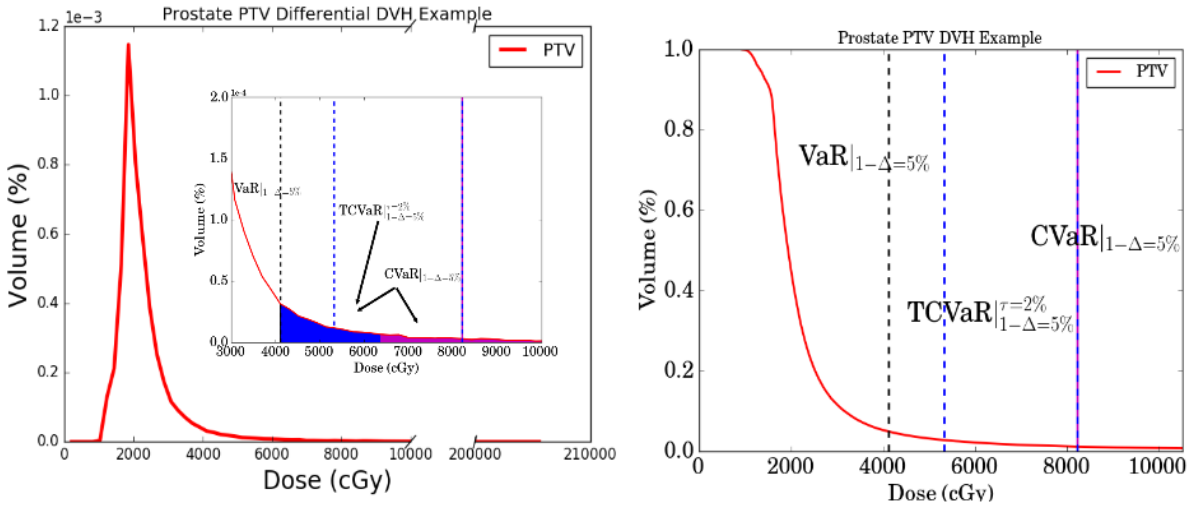


Figure 4.4: VaR and CVaR example on a differential DVH (left). $D_{5\%}$ is VaR at 5%; UCVaR at 5%; and TCVaR at 5% with 2% of the hottest voxels (purple volume) excluded. CVaR is the mean dose of the volume highlighted by the blue and purple.

For example, assuming 2 cubic centimeters (cc) is 5% of the total PTV volume, we can limit the dose to the 5^{th} percentile: constraint

$$\delta^{+}_{\text{PTV}}(d;0.05) \leq U$$

requires the *minimum* dose of the hottest 2cc to be no more than U centiGray (cGy). The CVaR approximation of this non-convex constraint will instead require the *average* dose to the hottest 2cc to be no more than U cGy and can be stated as:

$$\overline{\delta}^{+}_{\text{PTV}}(d,\nu_k;0.05) \leq U. \quad (4.8)$$

Observe that if there exists a feasible (d, ν_k) that satisfies (4.8), then it guarantees that the corresponding VaR also satisfies it. Similarly, for the lower CVaR at Δ, we have:

$$\mathbf{VaR}^- : \delta_k^-(d; 1-\Delta) = \max_{\nu_k}\{\nu_k \in \mathbb{R} : \frac{|j \in V_k : d_j \leq \nu_k|}{|V_k|} \leq 1-\Delta\} \tag{4.9}$$

and

$$\mathbf{CVaR}^- : \max_{\nu_k}\{\overline{\delta}_k^-(d, \nu_k; 1-\Delta)\} = \max_{\nu_k}\{\nu_k - \frac{1}{(1-\Delta)|V_k|}\sum_{j \in V_k}\max(0, \nu_k - d_j)\}. \tag{4.10}$$

4.4.2 Multicriteria Optimization Treatment Planning Model

Recall that we assumed that we are considering $K+1$ criteria for $K \geq 1$; suppose the CVaR (DVH) metrics representing these criteria are such that we would like metrics $0, 1, \ldots, K_1$ to have small values, and metrics $K_1+1, \ldots, K+1$ — to have large values. Let $\delta : \mathbb{R}_+^{|V|} \to \mathbb{R}_+^{K+1}$ be the vector-valued function of criteria. We can state our problem as a multicriteria optimization problems:

$$\begin{aligned} \text{mininimze}_{d,s} \quad & \delta(d) = \left\{\delta_0^+(d;\Delta), \ldots \delta_{K_1}^+(d;\Delta), -\delta_{K_1+1}^-(d;\Delta), \ldots, -\delta_K^-(d;\Delta)\right\} \\ \text{s.t.} \quad & d = Gs \\ & d \in \mathcal{D} \\ & s \geq 0, \end{aligned} \tag{NCVX-MCO-BT}$$

where $\mathcal{D}$ is a convex (typically, polyhedral) set of restrictions on dose that represent "hard constraints" of the treatment protocol, which dose distributions must *always* satisfy (e.g., bounds on minimum, mean, maximum, and linearized EUD doses). We would like to generate a set of plans that are efficient, i.e., belong to the Pareto surface of (NCVX-MCO-BT), making them suitable for the library $\mathcal{L}$ of Section 4.3.

Since VaR metrics are non-convex, using common multicriteria optimization methods (e.g., approaching (NCVX-MCO-BT) via a weighted sum objective or ϵ-constraint methods)

requires solution of multiplies instances of a large mixed-integer program that can not be done efficiently (*Deasy*, 1997; *Lee et al.*, 2003; *Romeijn et al.*, 2003). Instead, we approximate VaR-based metrics with CVaR-based ones, so that the resulting problems instances are convex, and thus efficiently solvable.[†] However, to achieve an accurate approximation, we next motivate a modification of CVaR in the context of **BTTP!** (**BTTP!**).

4.4.3 Voxel Truncation

It is clear that for any dose distribution and fixed quantile Δ, VaR and CVaR satisfy the following relationship:[‡]

$$\overline{\delta}^{+}(d, \Delta) \geq \delta^{+}(d, \Delta) \qquad . \tag{4.11}$$

The nature of dose delivery in BT results in long, and often heavy, upper tails in dose distributions to each structure due to extreme doses to (typically PTV) voxels which are near the sources. These values "skew" the upper CVaR, which makes (4.11) a poor upper bound, and makes selecting $\Delta^{\text{CVaR}} = \Delta^{\text{VaR}}$ a poor choice for approximation.[§] In this section we address this issue by discussing the concept of a "truncated CVaR" that tries to exclude (most) of the hottest dose contributions to the mean or entire voxels' doses. (Incidentally, if we knew exactly which voxels were the $100(1-\Delta)\%$ hottest (or coldest), then we could exclude all of those voxels from the structure, and use the (convex) maximum dose metric to capture VaR exactly!)

[†]Note that because the criteria are approximated as CVaR, these solutions are not necessarily Pareto optimal with respect to the true criteria of interest (VaR), which means some solutions may be dominated. In this work we assume all solutions included in the library are non-dominated (e.g., selected via non-dominated sorting). We then can use triangulation and interpolation of the trade-off surface with respect to the true metrics of the non-dominated solutions.

[‡]The same can be said for the lower tail relationship, but we will focus on the upper tail due to the nature of our problem.

[§]One solution to this is to exclude the mean dose of the hottest portion of the tail, resulting in a difference of convex functions. However, this re-introduces non-convexity into our problem.

4.4.3.1 Excluding Large (Small) Voxel-Source Coefficients in Computing Upper (Lower) CVaR

One possible approach to selecting which voxels to remove from the calculation of CVaR for each structure is based on the following consideration. Since most or all sources are implanted in or near the PTV, OAR voxels that will receive high dose are likely to be in close proximity of some sources in the PTV. We could "reshape" the tail of the dose distribution by excluding extreme contributions to the dose at each voxel. When calculating the dose at each voxel, we could exclude (for the purposes of CVaR evaluation) the contribution to the dose from sources that are located within a short distance of this voxel; since dose coefficients are proportional to the inverse squared distance (TG-43) between voxel and source, this would exclude excessive dose contributions from the calculation of the metric.

Our experiments showed that excluding such partial contributions to dose in CVaR calculations did not improve, but rather worsened the values of VaR metrics of plans obtained from CVaR-based optimizations. This is likely because we are still including all voxels in the tail calculation and only excluding high voxel-source pairs. Although dose is being accounted for, volume is not. Therefore, entire voxels must be considered and voxel-source proximity is not sufficient and are only secondarily related to dose.

4.4.3.2 Exclude Voxels by a Starting Solution

Another possible approach is to calculate CVaR for a truncated the dose vector, obtained by removing entire *voxels* from the structure in question, if they are believed to receive extremely high doses that would skew the tail of the distribution. Again, it is not known a priori which voxels in the structure are the hottest, but they can be selected, for example, as the hottest voxels in a high-quality baseline plan, obtained, e.g., by solving an unmodified CVaR-based optimization problem or by using the commercial planner.

Lack of precise knowledge of the right voxels to exclude can still be an issue. In the extreme case, we truncated all but one voxel incorrectly, resulting in one of the remaining

voxels receiving an extremely high dose (in fact, the maximum dose to the structure). Then CVaR of the truncated dose vector will still be a poor approximation of VaR due to the original issue of fat distribution tail. Instead, we consider an even *larger* tail size (i.e., quantile) and choose to truncate a *smaller* (but non-zero) portion of this tail, and consider the mean dose of the remaining voxels. In doing so, we may not truncate all of the correct voxels, but by truncating some of the hottest voxels we can significantly reduce the contribution of the remaining hot voxels on the average tail dose, bringing the approximation closer to the actual VaR. This achieves a balance between the two goals of i) closely approximating VaR with a convex metric, and ii) making the approximation robust when the high dose may shift to different voxels among different plans. Since most of the hottest voxels will be those closest to dwell locations, we make an assumption that the CVaR-optimal solution and VaR-optimal solution both share similar voxels receiving high doses.

We now make this idea more precise. Let τ denote the percent of structure volume truncated. We define the corresponding truncated upper CVaR (TCaR) metric to be:

$$\overline{\theta}_k(d, \nu_k; \bar{d}, \Delta, \tau) = \frac{(1-\Delta)}{1-\Delta'}\nu_k + \frac{1}{(1-\Delta')|V_k|} \sum_{j \in V_k : \bar{d}_j \leq \bar{D}^\tau} \left(\max\left(0, d_j - \nu_k\right)\right), \tag{4.12}$$

where truncated voxels are chosen based on the baseline plan $\bar{d}$, $\bar{D}^\tau = \delta^+(\bar{d}; 1-\tau)$, and $\Delta' = \Delta + \tau$. (The same idea can be applied for a lower tail with cold voxels.) Our computational experiments showed that truncation makes little improvement for PTV lower tail and OAR upper tail metrics (i.e., metrics of structures that are not close to dwell locations) since the voxels in question are unlikely to receive extreme doses.

Moreover, the initial reference plan $\bar{d}$ may not be a good indicator of the hottest voxels in plans across the trade-off surface, leading to poor quality plans since our metric is poorly approximated. The more voxels we incorrectly truncate, i.e., including them in the TCVaR approximation, the worse our approximation is.[¶] We discuss methods for improving the

[¶]The dose values of the truncated hot voxels also matter, e.g., if $1 - \Delta^{\text{VaR}} = 10\%$ and we wanted to truncated 5%, then we'd prefer the 5% hottest than, e.g., the next hottest 5%.

truncation, and hence, the approximation, later in this chapter.

4.4.4 Library Generation: ϵ-Constraint Method

We solve our multicriteria problem (NCVX-MCO-BT) with truncated conditional value-at-risk (TCVaR) metrics replacing VaR metrics in the objective by using the *ϵ-constraint* method (*Haimes et al.*, 1971). In this method, we consider instances of (single-objective) optimization problems, with one of the metrics serving as the objective, and the remaining metrics represented as constraints. As before, we subscript the metric in the objective with 0 and assume smaller values are preferred. For convenience, let the right-hand sides of TCVaR constraints be $\epsilon = (\underline{\epsilon}, \bar{\epsilon}) \in \mathbb{R}^K$ for the upper and lower tail constraints:

$$\text{minimize}_{d,s,\nu_k,k=0,\ldots,K} \quad \overline{\theta}_0^+(d, \nu_k; \bar{d}_k, \Delta, \tau) \tag{4.13}$$

$$\text{s.t.} \quad \overline{\theta}_k^+(d, \nu_k; \bar{d}_k, \Delta, \tau) \leq \bar{\epsilon}_k \qquad k = 1, \ldots, K_1 \tag{4.14}$$

$$\overline{\theta}_k^-(d, \nu_k; \bar{d}_k, \Delta, \tau) \geq \underline{\epsilon}_k \qquad k = K_1 + 1, \ldots, K \tag{4.15}$$

$$d = Gs \tag{4.16}$$

$$d \in \mathcal{D} \tag{4.17}$$

$$s \geq 0 \tag{4.18}$$

This problem is piecewise-linear convex and each instance can be efficiently solved with off-the-shelf linear programming solvers. If an instance is feasible, let a corresponding optimal plan be $(d^*(\epsilon), s^*(\epsilon))$.[‖] We consider a set $E \in \mathbb{R}_+^K$ of values of ϵ to create and solve multiple instances of (4.18). The solutions are then evaluated with respect to the true metrics of interest (VaR) and form the library discussed in Section 4.3, $\mathcal{L}(E)$. Finally, note that we can use different baseline plans to select voxels for truncation in each TCVaR metric.

[‖]For the library of solutions, we only consider feasible plans and infeasible instances will be ignored.

4.4.4.1 Iterative Truncation

As discussed before, the impact of truncation based on a baseline solution strongly depends on the quality of this solution.

One way to assess the quality of a particular choice of a subset of voxels to truncate is to use the resulting TCVaR to find a treatment plan by solving an instance of (4.18), and identify the hottest voxels in the resulting plan. We can compare this set of voxels to the ones initially truncated (i.e., those predicted to be the hottest based on a baseline plan), identifying the rate of "true positives," which measures the quality of approximation that was used — the closer it is to 100%, the closer we are to some value between the $100(1-\Delta^{\text{TCVaR}})\%$ and $100(1-\tau)\%$ quantiles. (Recall that, if we could guess the hottest voxels perfectly, we would use $\Delta^{\text{TCVaR}} = \Delta^{\text{VaR}}$ and τ such that $(1-\Delta^{\text{VaR}}-\tau)|V_k| = 1$, i.e., exactly the volume considered is one voxel.)

Therefore, we also investigate the following: if the initial solution is a poor indicator of the hottest voxels, can we improve our true positive rate by using the plan resulting from TCVaR optimization to update the baseline plan, i.e., update $\bar{d}$ to be the solution from the TCVaR optimization in (4.12) and repeat the optimization? Although there is no guarantee that the true positive rates improve (as we will see), having a solution that is a better baseline for guessing the hottest voxels can potentially improve the quality of our solutions.

4.5 Experiments

In the following section we will report on the number of empirical experiments we conducted to assess the feasibility of the overall approach we discussed above to perform analysis of trade-offs between various feasible treatment plans. To streamline that presentation, we dedicated this section to some preliminaries, describing the experiments we performed and the values of parameters chosen in those experiments.

Our experiments consisted of retrospective application of our method to several patient

cases. For the purposes of analysis, we will only focus on testing truncation of *one* upper tail metric for each treatment site, since skewed lower tails are rarely observed in BT. Furthermore, we assume truncation and tail size parameters across multiple structure-criteria are independent. We leave analysis of inter-structure-criteria effects for a future discussion.

We implemented the TCVaR optimization model (4.18) on a machine with 3.50GHz×8 processor and 32GB memory in Python using Gurobi's primal simplex method (Gurobi Optimization, Houston, TX). Trade-off surface population was implemented in Matlab (MathWorks, Natick, MA).

4.5.1 Choice of $1-\Delta^{\mathbf{TCVaR}}$ and τ

We approximate a DVH metric with quantile $1-\Delta^{\mathrm{VaR}}$ by TCVaR with quantiles $1-\Delta^{\mathrm{TCVaR}}$ and τ. We parametrize the latter relative to the former by using scalars $\gamma_\Delta, \gamma_\tau$: $(1-\Delta^{\mathrm{TCVaR}}) = \gamma_\Delta(1-\Delta^{\mathrm{VaR}}),\ \gamma_\Delta \geq 1$ and $\tau = \gamma_\tau(1-\Delta^{\mathrm{VaR}}),\ \gamma_\tau \leq 1$, respectively.** We consider multiple values $\gamma_\Delta \in [1.5,\ 2.5]$†† and$\gamma_\tau \in [0.0625,\ 0.75]$. We will discuss how we evaluate the performance of a choice of $\gamma_\Delta, \gamma_\tau$ after describing the rest of the process for plan generation.

4.5.2 Choice of ϵ

We choose ranges of values for ϵ according to ideal values of each metric as the "center," denoted as $\theta^0 \in \mathbb{R}_+^K$, of our trade-off surface (i.e., this is a treatment plan around which we would like to consider trade-offs). Let $\mu \in \mathbb{R}_+^K$ be the increments in of ϵ, and $\sigma \in \mathbb{R}_+^K$ be the total deviation from the central value for which we would like to explore potential trade-offs (typically, this will be $15-20\%$, which results in approximately a $30-40\%$ window for each metric). Then, our choices of ϵ belong to $E = \{\theta^0 \pm n\mathrm{diag}(\mu)\theta^0 \mid n \in Z, |n\mu_k| \leq \sigma_k, k =$

**$\gamma_\Delta < 1$ guarantees an overestimate and $\gamma_\tau > 1$ potentially truncates the entire quantile.

††For initial cases, we considered several values that were similar and for later cases consider fewer for illustration purposes

$1 \ldots, K\}$.[‡‡]

4.6 Results

4.6.1 Plan Interpolation By Dwell Time

We begin with justification of plan interpolation. After generating a library of plans, we validate the quality of linearly interpolated plans (by dwell time) to try and recoup each plan, other than vertices of the convex hull of L. Specifically, we remove one library plan and reconstruct the dwell times that achieve (almost) the same DVH values by applying Algorithm 1 on the remaining plans in the library. Table 4.4 show the maximum absolute errors between the achieved DVH values of an original plan in the library and the DVH values of the plan obtained by dwell time interpolation. Most errors are less than 1%, indicating that, given the density of library plans generated by our choice of set E, interpolation accurately predicts metric values for additional plans.

$\gamma_\Delta \backslash \gamma_\tau$	0	0.25	0.33	0.50	0.67	0.75
2.50	0.6893	0.4372	0.2884	0.1673	0.1894	0.1441
2.25	0.2771	0.3177	0.2547	0.1078	0.1937	0.1441
2.00	1.3191	0.3695	0.3428	0.1299	0.1579	0.1097
1.75	0.8626	0.3650	0.4273	0.2104	0.1656	0.1197
1.50	0.2047	0.9038	1.6033	0.9363	0.1364	0.1713

Table 4.4: Breast 1. Replicating Library Plans via Interpolation of Dwell Times: Max Absolute Error (%) of PTV D15cc. Note that most errors are below 1%.

In Tables 4.5 and 4.6, we present mean and maximum errors of interpolation for Breast 2. In contrast to Breast 1, the maximum errors are higher: interpolation of dwell times leads to up to $2-5\%$ worse PTV D15cc dose than the original optimized (i.e., library) plan. Though most errors are close to zero, there are extreme outliers in replicating library plans

[‡‡]For implementation, criteria chosen include PTV coverage (lower tail) and upper tail structure metrics, which all compete with PTV coverage. We start with the highest values of ϵ_k and iterate through in decreasing order. Therefore, for a fixed level of PTV coverage CVaR (corresponding metric constrained in 4.15) and values of upper tail bounds, if the problem is infeasible, then all remaining upper bound values of the upper tail criterion will be skipped since PTV coverage cannot be achieved for any smaller values of upper tail bounds.

$\gamma_\Delta \backslash \gamma_\tau$	0	0.0625	0.125	0.25	0.33	0.50	0.67	0.75
2.50	0.3277	0.3409	0.2600	0.2553	0.2528	0.3569	0.3585	0.3592
2.25	0.2855	0.3012	0.2561	0.2478	0.2436	0.3702	0.3479	0.3388
2.00	0.2859	0.3142	0.2103	0.2366	0.2678	0.3370	0.3781	0.3702
1.75	0.2920	0.3147	0.2172	0.2444	0.2643	0.3633	0.3481	0.3587
1.50	0.2947	0.3183	0.2852	0.2532	0.2462	0.3019	0.3496	0.3283

Table 4.5: Breast 2. Replicating Library Plans via Interpolation of Dwell Times: Mean Error (%) of PTV D15cc. Note that most errors are below 1%.

$\gamma_\Delta \backslash \gamma_\tau$	0	0.0625	0.125	0.25	0.33	0.50	0.67	0.75
2.50	2.4484	3.9284	3.0356	2.9997	3.1130	4.3064	4.5070	3.0097
2.25	2.1067	2.9605	3.0365	2.9500	2.8553	4.1388	4.4182	3.0590
2.00	2.0857	3.9905	1.9340	2.9439	2.9725	3.9399	4.3689	5.1139
1.75	2.2075	3.7677	1.8825	3.1200	3.1029	4.0345	4.2912	5.2296
1.50	2.1925	3.9839	4.2777	3.1827	3.1034	3.7926	4.3379	3.0928

Table 4.6: Breast 2. Replicating Library Plans via Interpolation of Dwell Times: Max Error (%) of PTV D15cc (No iterative truncation).

via dwell times; this may be due to the poor approximation of dose-volume metrics (i.e., the truncation is not effective either due to size and/or accuracy of truncation); we will return to formally define what is meant by "truncation accuracy." Fewer than 5% of replicated library plans had more than 1% worse values of this metric.

4.6.2 Quality of a Trade-off Surface Generated with a $\gamma_\Delta, \gamma_\tau$ — One Truncation

We begin with how to evaluate the quality of a library generated by a particular choice of $\gamma_\Delta, \gamma_\tau$ (recall that this determines the choice of TCVaR quantile and truncation). We begin by showing the benefits of using truncation versus not truncating. We will do this by considering a set of "partial" candidate plans: for each such plan, we have K metrics' values (recall the trade-off surface is in $\mathbb{R}^{K+1}$) and "look up" in the library (i.e., via plans' *metric* interpolation[§§]) the $K+1^{\text{st}}$ metric. This way, we are evaluating each library with the exact same partial plans, and compare them by a single metric, namely the $K+1^{\text{st}}$ one.

[§§]Although we have shown that interpolating by dwell time and by DVH metric achieve similar values of DVH metrics, we choose to interpolate by DVH metrics so that the 0^{th} metric can be compared among libraries since the other metrics will be exactly the same also *among* libraries.

Because we are interpolating, the partial candidate plans should be within the convex hull of the library plans projected onto $\mathbb{R}_+^K$ corresponding to the appropriate K metrics. In our experiments, we used 1% spacing for each metric up to $\pm 10\%$ in each direction of the central values (the libraries were generated with $\pm 15\%$ spacing).

In Figures 4.5–4.7, each box plot corresponds to a TCVaR tail size and the number of times truncation occurred, represented by γ_Δ-number of truncations. Then each large plot corresponds to a truncation size γ_τ. For each combination of $(\gamma_\Delta, \gamma_\tau)$, a plan library was generated with the same set of right-hand sides in constraints on TCVaR metrics. A set of candidate plans Q were constructed by metric interpolation for each library to determine the value of metric δ_0, via metric interpolation. For each candidate plan q, we define the best achieved metric to be:

$$\delta^{0*}(q) = \min_{(\gamma_\Delta, \gamma_\tau)} \{\delta^0(q; (\gamma_\Delta, \gamma_\tau)\}. \tag{4.19}$$

The value of $(\gamma_\Delta, \gamma_\tau)$ that is the minimizer is the best combination of parameters to use to construct the dwell times for a candidate plan. Therefore, we use values of $\delta^{0*}(q)$ to normalize what candidate plans in other libraries achieve, defining the dose difference (%) of a candidate plan as:

$$\frac{\delta^0(q, (\gamma_\Delta, \gamma_\tau)) - \delta^{0*}(q)}{\delta^{0*}(q)} \geq 0, \qquad \forall q \in Q, \forall (\gamma_\Delta, \gamma_\tau).$$

Each individual box plot represents the ranges of dose differences achieved by all candidate plans from a library specified by $(\gamma_\Delta, \gamma_\tau)$.

4.6.3 Quality of a Trade-off Surface Generated with a $\gamma_\Delta, \gamma_\tau$ — Iterative Truncation

In this section we analyze the effects of iterative truncation, i.e., we solve the instance (i.e., (4.13)-(4.18) with the same ϵ and TCVaR parameters) multiple times, but each time we

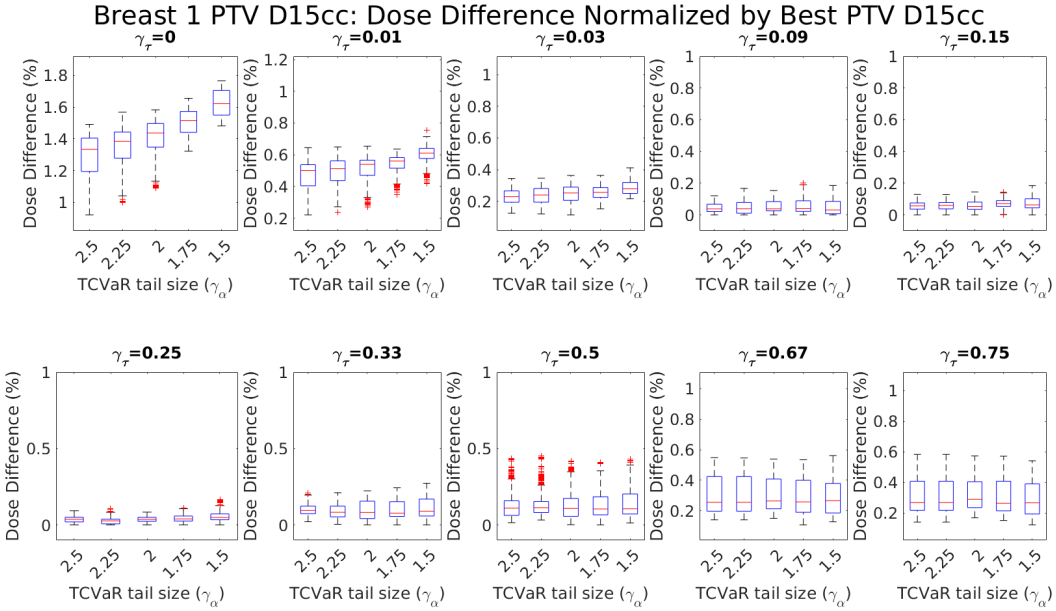


Figure 4.5: Breast 1. Without truncation, the minimum difference from the best is 1% while others only go up to 0.5%.

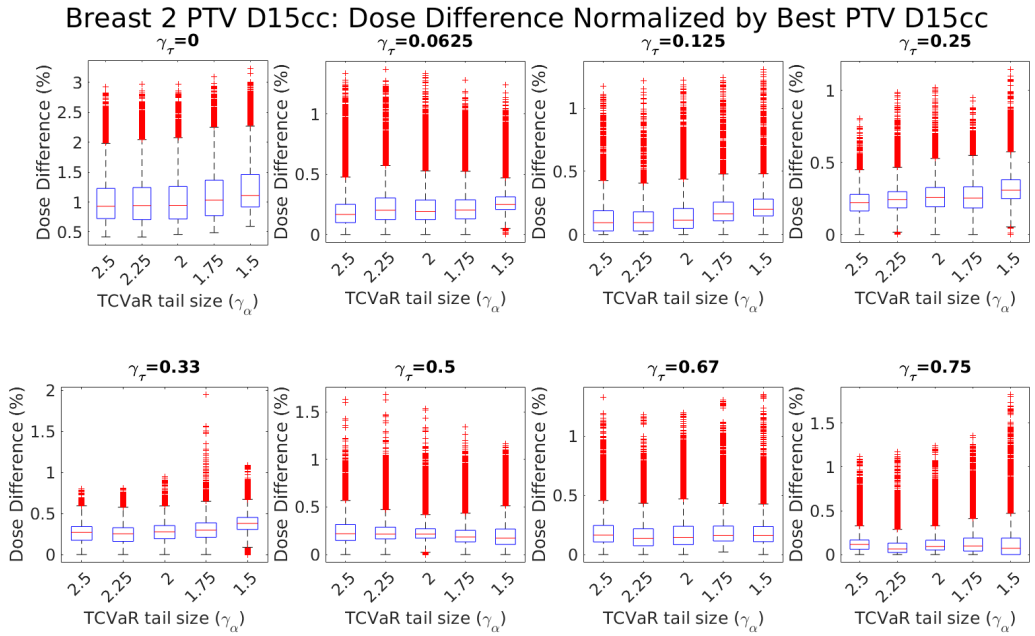


Figure 4.6: Breast 2. Without truncation, the minimum difference from the best is 0.5%, but at worst 3%; although many are within 0.5% of the best, several go up to beyond 1% depending on the level of truncation.

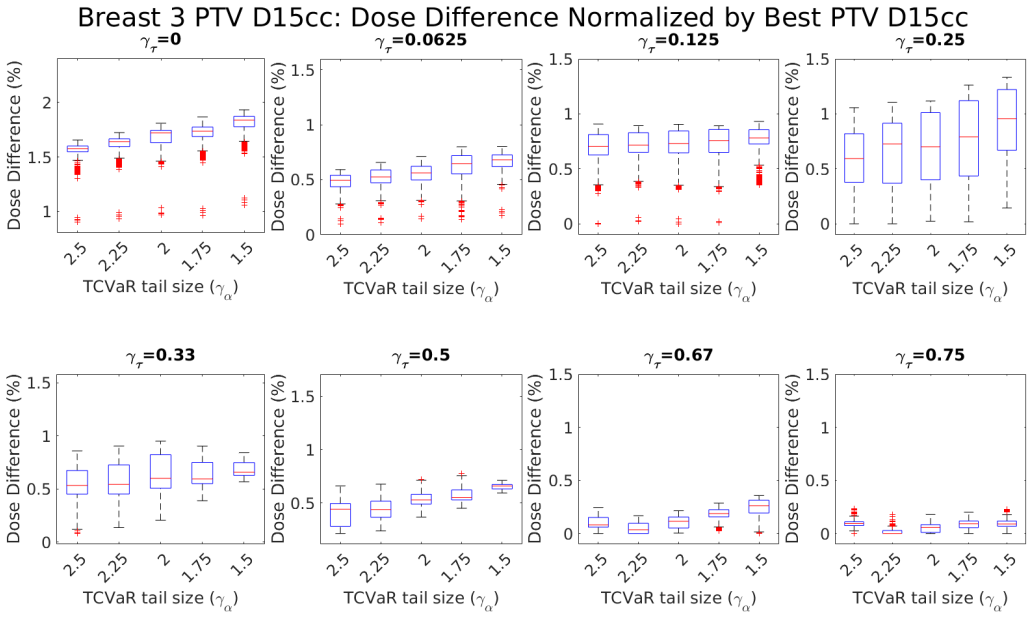


Figure 4.7: Breast 3. Without truncation, the minimum difference from the best is 0.5%, but at worst 3%; although many are within 0.5% of the best, several go up to beyond 1% depending on the level of truncation.

update what we consider to be the $100\tau\%$ hottest voxels based on the solution obtained by the previous optimization for the metric we're truncating. We investigated up to 3 iterative truncations (i.e., 3 solves for each instance).

For a set of partial candidate plans, the $K+1^{\text{st}}$ metric is computed using metric interpolation according to each library generated. For a given truncation iteration, Figure 4.11 shows the distribution of dose differences (%) computed by (4.19). A moderate truncation value ($\gamma_\tau = 0.33$) performs the best with many families of libraries being no more than 0.1% of the best values. For a given truncation level and iteration, the affect of γ_Δ does not differ in distribution of dose differences. In Figure 4.9 is a rearrangement of Figure 4.11 by showing how iterative truncation affects a pair of TCVaR parameters. If the truncation is small (e.g., $\gamma_\Delta = 0.0625$), iterative truncation overall does not achieve better values, while if the truncation is large, iterative truncation shrinks the range of dose differences and pushes the median dose difference toward 0.

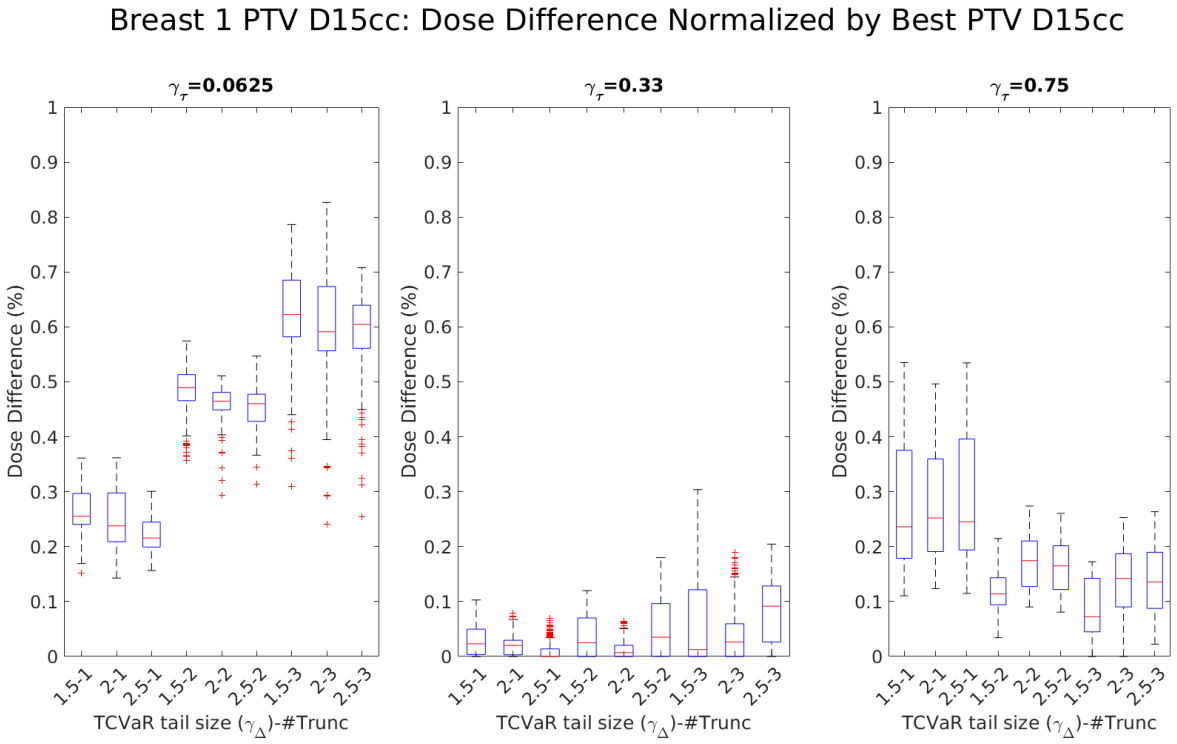


Figure 4.8: Breast 1. Effects of iterative truncation. For a given truncation iteration (e.g., all plans obtained by truncating only twice), dose differences are computed using (4.19) and the distribution of percent differences are plotted as a box plot for each $\gamma_\Delta, \gamma_\tau$ combination.

4.6.4 Truncation Accuracy

As a proxy for evaluating the quality of our approximation, we consider the $100\tau\%$ of voxels that were truncated according to a baseline plan and were actually $100\tau\%$ hottest in the optimized plan. We present some example progressions of accuracy by iteration of truncation in Tables 4.7 and 4.8 and Figure 4.10. Figure 4.10 shows the percent of hottest voxels correctly predicted based on the previous iterations plan (at 1, we use the clinical plan) and each set of lines corresponds to a combination of truncation size and instance of the problem solved (one plot per truncation size). We specifically present Breast 2 to show that iterative truncation improves the approximation (reflected in the increase in ability to predict the hottest voxels), but note that there is no guarantee of improvement (as demonstrated by some plans).

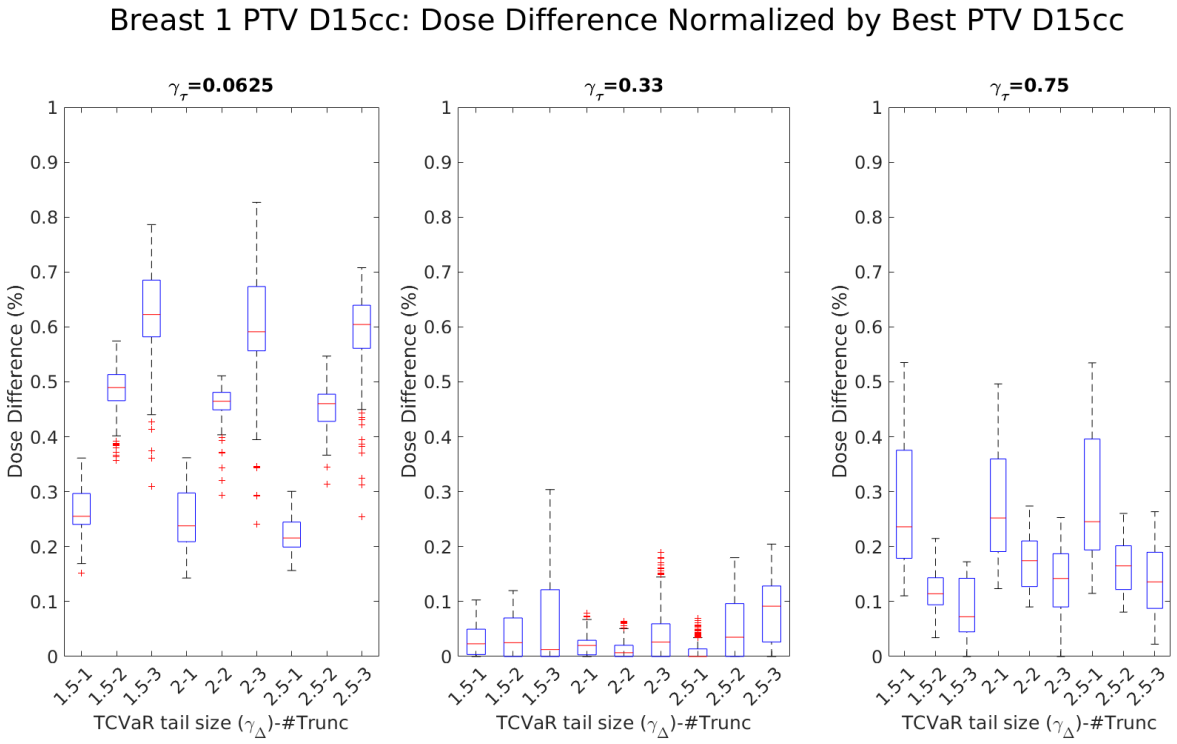


Figure 4.9: Breast 1. Effects of iterative truncation. For a set of partial candidate plans, the $K+1^{\text{st}}$ metric is computed using the library by metric interpolation. For a given truncation iteration (e.g., all plans obtained by truncating only twice), dose differences are computed using (4.19) and the distribution of percent differences are plotted as a box plot for each $\gamma_\Delta, \gamma_\tau$ combination (truncation iteration is clarified with the choice of γ_Δ).

4.6.5 Testing Parameter Choice

Based on Breast 1 and 2, we choose parameter $(\gamma_\Delta, \gamma_\tau) = (2, 0.0.33)$ and allow for 2 iterations of truncation (to save time) to generate a library. Figure 4.12 shows how the library we choose performs (in dose difference) compared to libraries generated by other TCVaR parameters. This library was used to normalize, so its corresponding box-plot has percent dose differences all at 0. Most plans from most other libraries perform up to 1% better, but many require 3 iterations of truncation (namely for $\gamma_\tau = 0.75$).

4.6.6 Comparison of Plan Quality with Commercial System

We compare the quality of plans we achieve with that of the commercial system's plan by fixing other metrics and comparing PTV D95, which is tumor coverage. In Figure 4.13,

Truncation Size (γ_τ)	Accuracy Increase Iter. 1 to 2	Accuracy Increase in Iter. 2-3
0.0625	100.0%	91.7%
0.33	100.0%	98.2%
0.75	100.0%	100.0%

Table 4.7: Breast 1. Percent of plans that led to improved truncation accuracy over 3 truncation iterations (100% means for all instances of feasible ϵ, truncation accuracy improved from one iteration to the next iteration.)

Truncation Size (γ_τ)	Accuracy Increase Iter. 1 to 2	Accuracy Increase in Iter. 2-3
0.0625	100.0%	97.3%
0.33	100.0%	81.8%
0.75	100.0%	91.2%

Table 4.8: Breast 2. Percent of plans that led to improved truncation accuracy over 3 truncation iterations (100% means for all instances of feasible ϵ, truncation accuracy improved from one iteration to the next iteration. See Figure 4.10 for trend in instances that improve in true positives.

we are competitive against the commercial system. In Figure 4.14, it is clear that with 3 truncations we perform better in PTV D95 than the commercial plans to.

4.7 Discussion

The results on TCVaR approximations of VaR as a function of $1-\Delta$ and τ suggest that for breast cases, good choices of parameters are independent of geometry. The rule for choosing γ_Δ to be close to 2 and γ_τ to be close to 0.5 seems to perform well as approximating VaR metrics. Furthermore, the choice of a warm-start plan to determine which voxels to truncate can be dynamically updated. It is possible that depending on the region of the trade-off surface, the set of hottest voxels may vary, and a plan that better captures the hottest voxels can be used to determine which voxels' doses to truncate, thereby improving the TCVaR approximation. A metric such as truncation accuracy (true positives on predicting hottest voxels) can be a proxy for determining when to update the plan with which we base truncation. Depending on the user's threshold for improvement in a metric such as PTV coverage (e.g., 1% vs 5%) there is a trade-off to consider between treatment plan quality

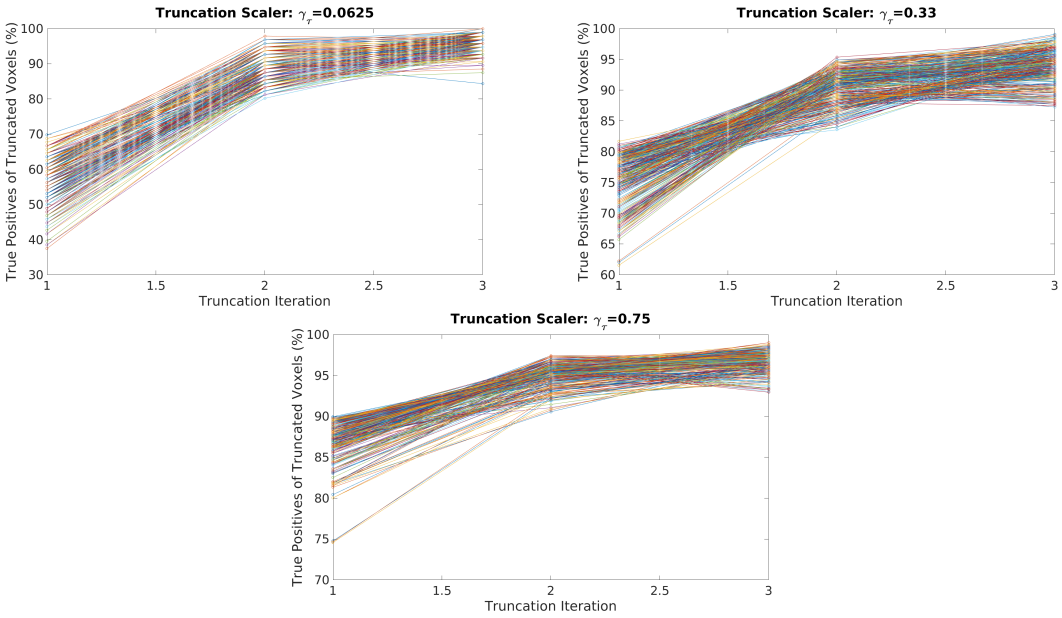


Figure 4.10: Breast 2. Truncation Accuracy by Iteration of Truncation (iteration 1 is truncating according to clinical plan); up to 3 iterations of truncation considered. $\gamma_\tau = 0.0625, 0.33, 0.75$.

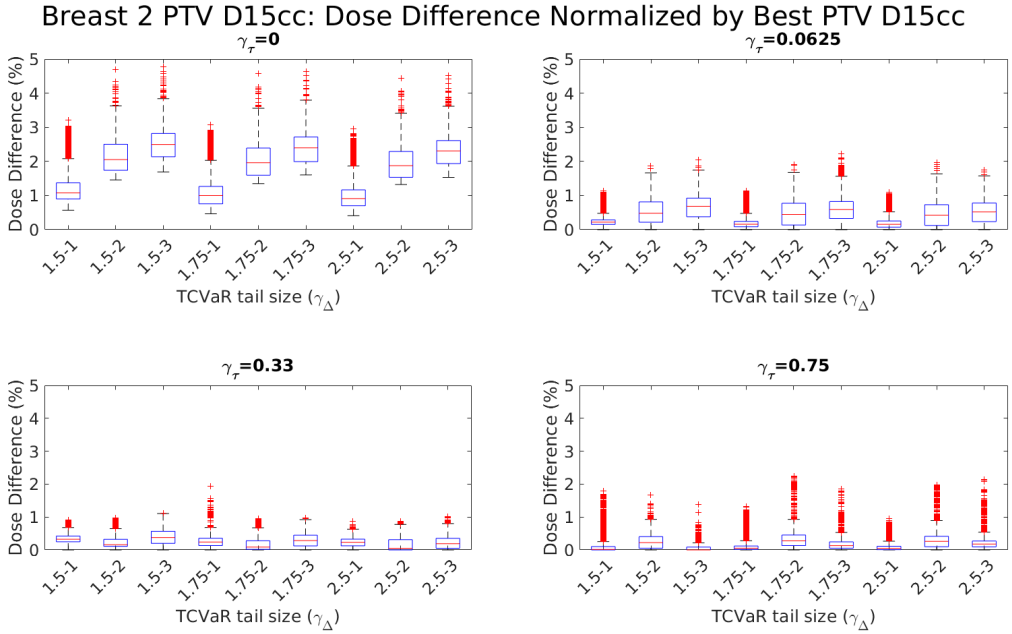


Figure 4.11: Breast 2. Effects of iterative truncation. For a given truncation iteration (e.g., all plans obtained by truncating only twice), dose differences are computed using (4.19) and the distribution of percent differences are plotted as a box plot for each $\gamma_\Delta, \gamma_\tau$ combination.

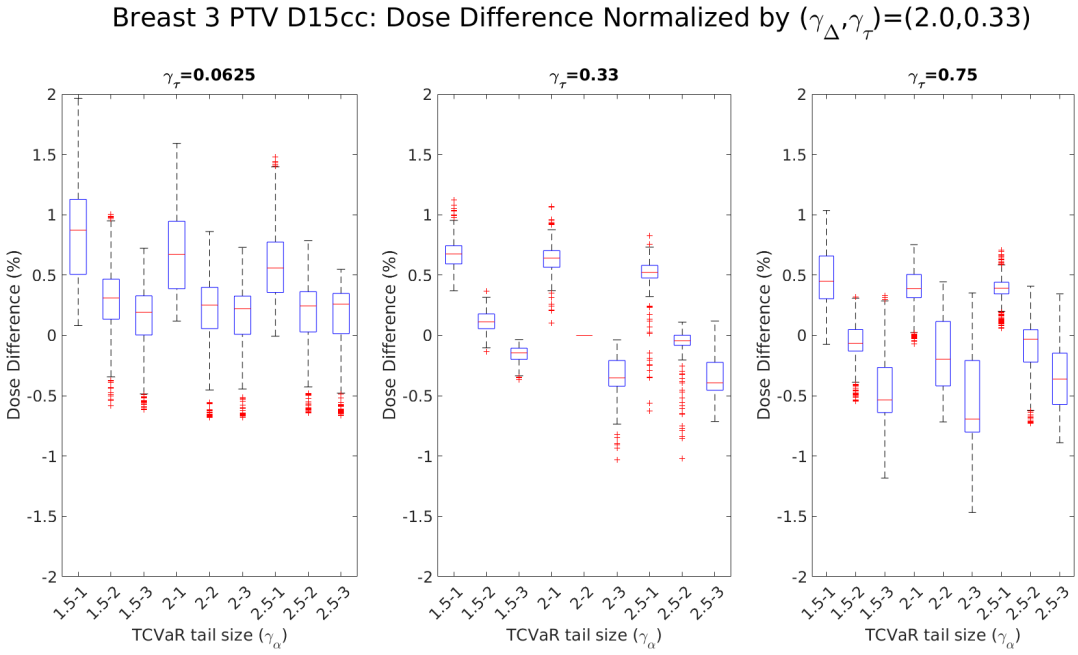


Figure 4.12: Breast 3. Choosing $(\gamma_\Delta, \gamma_\tau) = (2, 0.0.33)$ to generate a library based on Breast 1 and Breast 2.

and number of truncations. Our model produced competitive plans with respect to PTV coverage as the commercial planning system does, and with iterative truncation were able to gain up to 3% more coverage.

We include one empirical example each for cervix and prostate sites. In Figure 4.15, we consider a prostate case, but it seems that the choice of parameters do not perform well and suggests additional iterations of truncation is needed and/or choice of parameters are site-dependent. Figure 4.16 suggests that iterative truncation improves the accuracy in guessing the hottest voxels, but the range of values of PTV D1cc achieved can vary up to 10% from the best value among all combinations of TCVaRparameters. The cause of this larger variability (in comparison to the breast cases) is unclear and requires more investigation. One hypothesis for this is that although with iterative truncation, the accuracy increases, the small quantile considered makes the inclusion of a few hot voxels still skew the estimate, which could be remedied with an even larger quantile. This further suggests that the "ideal" parameters for TCVaRapproximation varies with quantile and/or treatment site.

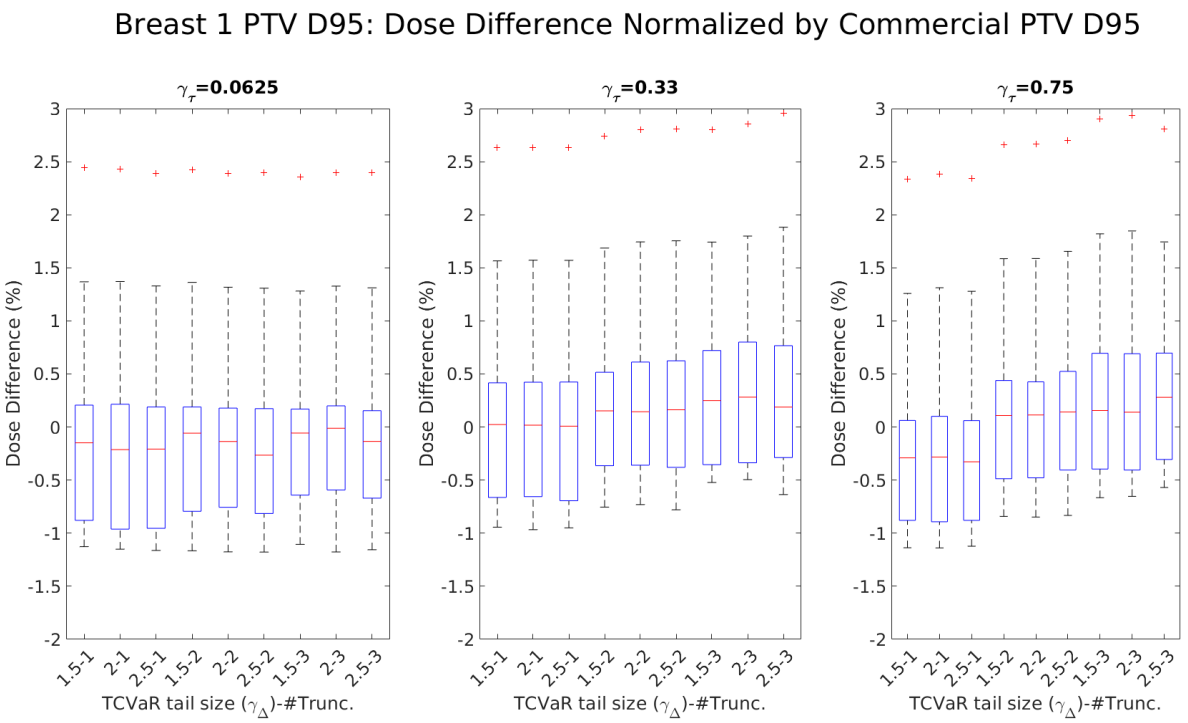


Figure 4.13: Breast 1 PTV Coverage Comparison with a Commercial System. A set of plans generated by an experienced medical physicist using the Eclipse Treatment Planning System. This set of plans was compared to the different libraries by fixing all other metrics and comparing PTV D95 achieved. Note these are sorted by TCVaR tail size and grouped by truncation size and truncation iteration.

We conduct similar analyses for a cervix case with Bladder D2cc truncation in Figures 4.17 and 4.18. It seems that for OARs without dwell locations do not require truncation, which is consistent with not having long DVH tails that the PTV would have.

This treatment planning approach can deal with any number of convex treatment criteria and can contain other treatment criteria such as those in *Romeijn et al.* (2004)). Moreover, as long as CVaR can capture the behavior of VaR, as we have done in this work with brachytherapy, this planning approach can be applied to other treatment modalities such as IMRT, VMAT, etc. With subsampling ($\leq$ 5000 voxels per structure) and using the primal simplex algorithm, instances to generate a library typically solved in less than 1 minute.

Our work also has limitations. Because the ground truth of optimal DVH values is unattainable, we have only relatively compared DVH measures achievable in parameterizations of our convex optimization model and the best among these are clinically acceptable.

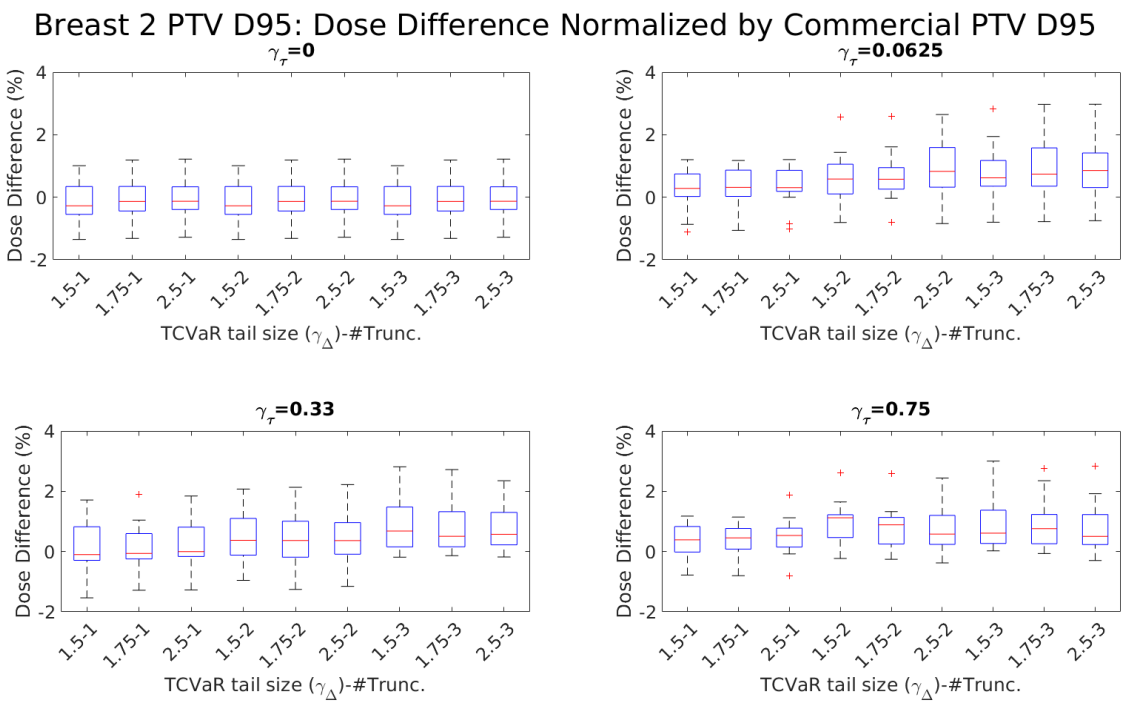


Figure 4.14: Breast 2 PTV Coverage Comparison with a Commercial System. A set of plans generated by an experienced medical physicist using the Eclipse Treatment Planning System. This set of plans was compared to the different libraries by fixing all other metrics and comparing PTV D95 achieved. Note these are sorted by TCVaR tail size and grouped by truncation size and truncation iteration.

The purpose of this work is to provide a proof-of-concept on how trade-off surfaces based on DVH metrics can be constructed. We also warn the reader that this work is only an empirical study. Validation of this method for clinical use will require QA on many more other patient cases and treatment sites.

4.7.1 Interpolation

4.5 and 4.6 suggest that interpolating by dwell time can be inaccurate, potentially due to choice of E (i.e., spacing between plans in the library should be smaller) or due to TCVaR being a poor approximation, which led to our investigation of iterative truncation. We note that though there are outliers in interpolating by dwell time, most are within 0.5% in replicating a plan in the library. To avoid significant loss in plan quality from dwell-time

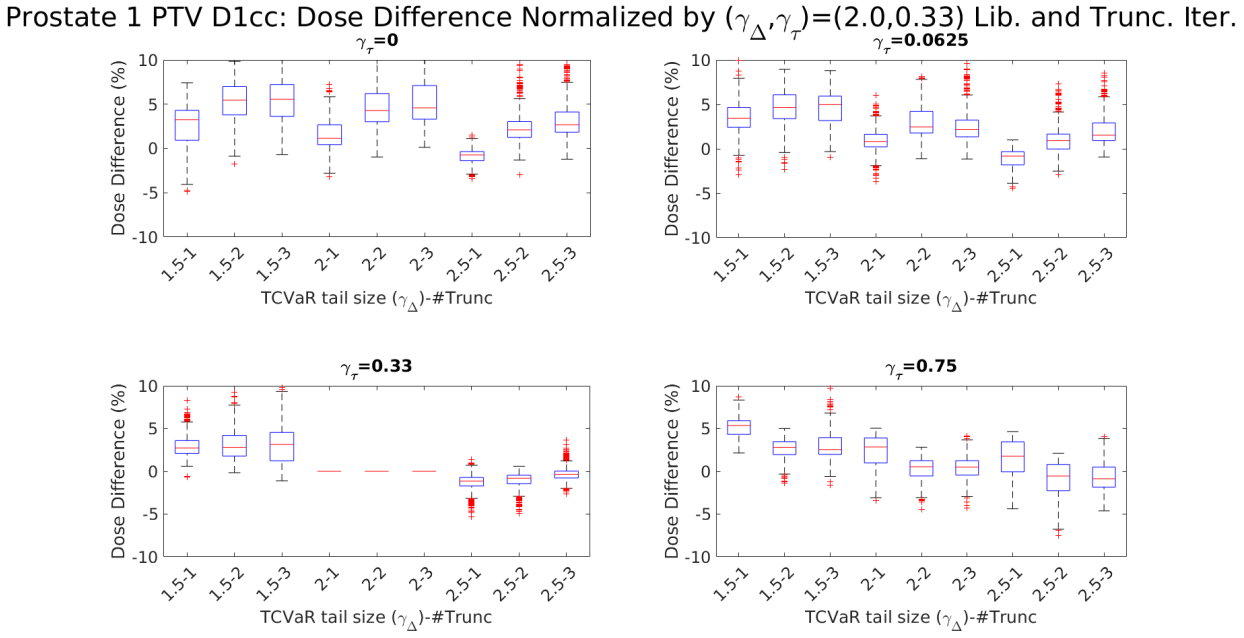


Figure 4.15: Example Prostate Case. Due to the wide range of values relative to the a particular library, either more truncations are needed or a better quality plan indicating hottest voxels is needed.

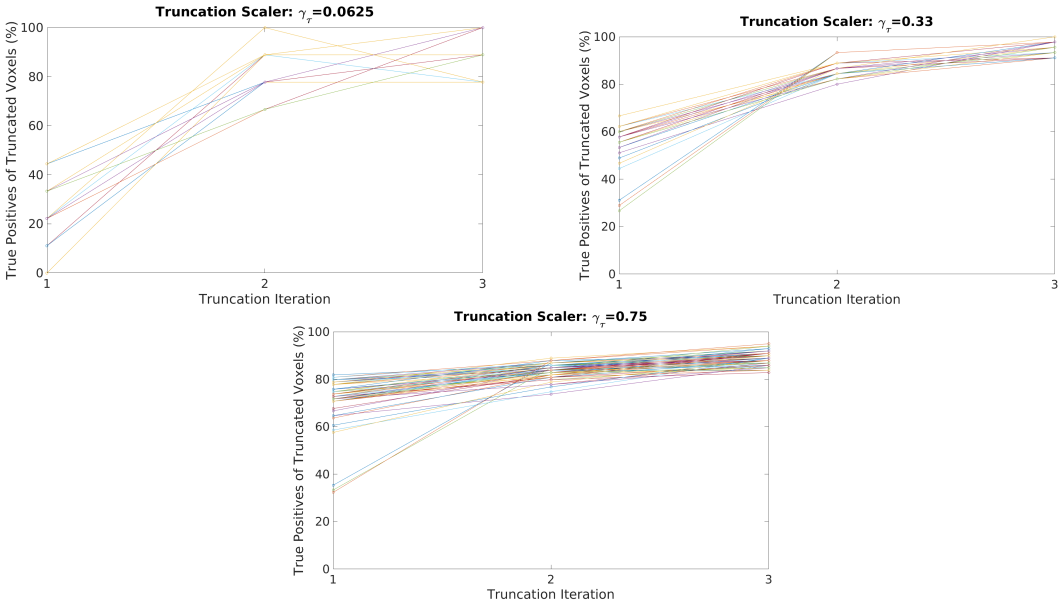


Figure 4.16: Prostate 1. Truncation Accuracy by Iteration of Truncation (iteration 1 is truncating according to clinical plan); up to 3 iterations of truncation considered. $\gamma_\tau = 0.0625, 0.33, 0.75$.

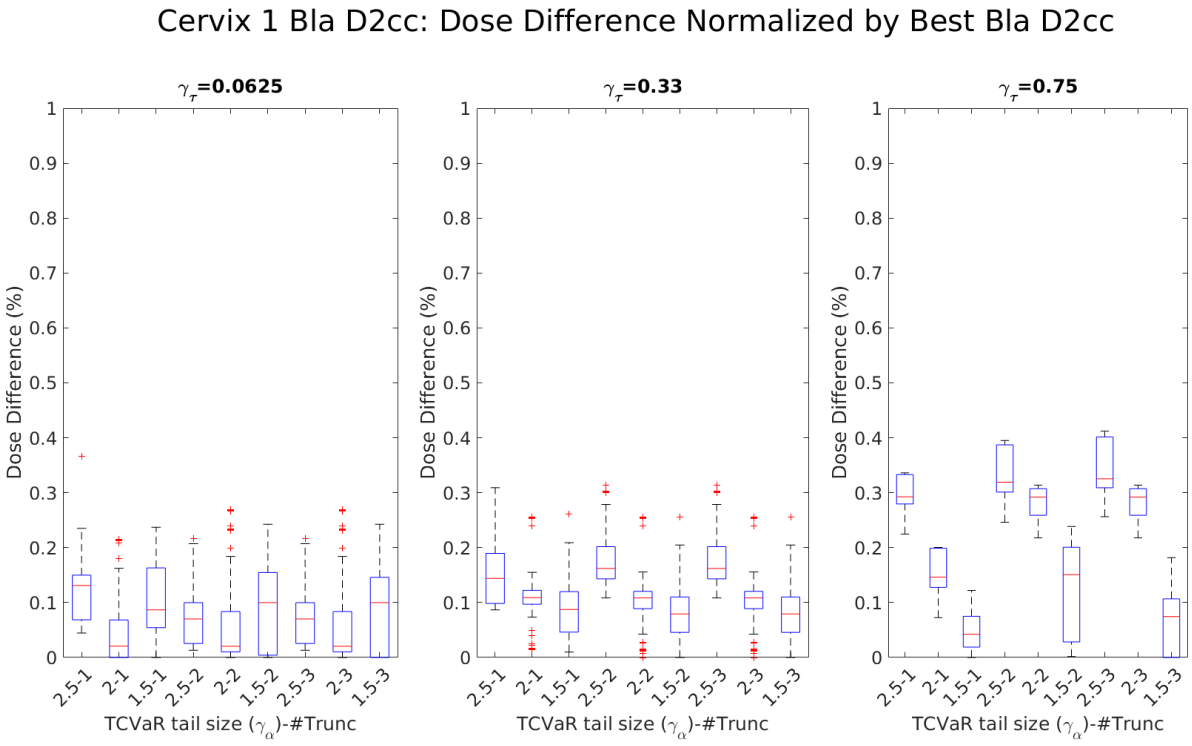


Figure 4.17: Example Cervix Case. Due to the wide range of values relative to the a particular library, either more truncations are needed or a better quality plan indicating hottest voxels is needed.

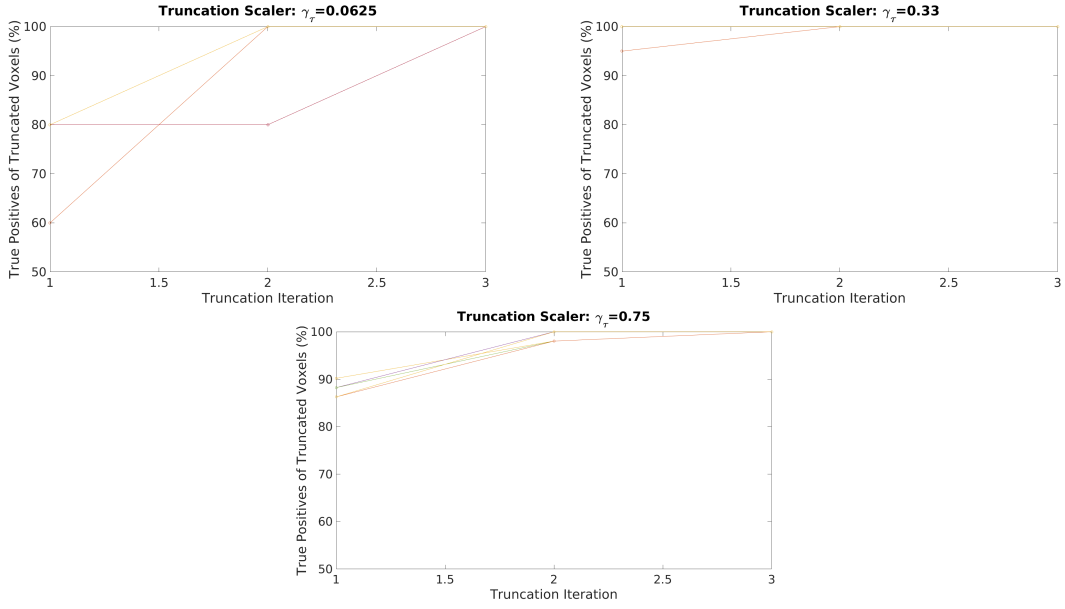


Figure 4.18: Cervix 1. Truncation Accuracy by Iteration of Truncation (iteration 1 is truncating according to clinical plan); up to 3 iterations of truncation considered. $\gamma_\tau = 0.0625, 0.33, 0.75$.

interpolation, once the final dose-volume metric values are decided, a final optimization can be performed to improve one of the metrics.

4.7.2 Computational Efforts

Library plan generation can be trivially parallelized given enough central processing unit cores (one instance on each core). Given enough cores, a library could then be generated within a few minutes. Moreover, using interpolation to increase density of plans in the neighborhood of a solution can also be easily parallelized, which allows for dynamic consideration of step sizes for trade-offs in the GUI. There are other ways to improve computation efficiency of our process: the library of plans may be dynamically generated by choice of ϵ and plan density can be locally increased via interpolation, which ultimately saves time from not generating plans in regions of the trade-off surface that are not considered.

Runtimes for each instance, depending on treatment site, varied from 3 seconds to 20 seconds. If increments of I are varied for K objectives as constraints and a range of R for each is considered, this means $(R/I)^K$ instances need to be ran to generate a library. If iterative truncation is considered, i.e., T iterative truncations (for one metric), then this would be $T(R/I)^K$ total instances. Our library generation is embarrassingly parallel and with C cores, this means up to a time of $20T(R/I)^K/C$. For $K = 5, R = 30\%, I = 6\%, T = 3, C = 500$, this results in a little over 6 minutes to generate a library.

4.8 Conclusions

Choosing the right set of parameters, truncated conditional value-at-risk in brachytherapy is a better approximation of value-at-risk than simply using conditional value-at-risk. If a good starting plan is provided to indicate hottest voxels, for breast, $\gamma_\Delta \in [1.5, 2.5]$ and $\gamma_\tau \in [0.33, 0.67]$) seem to be good choices. If an initial plan is poor in predicting hot voxels, iterative truncation should be applied. However, the reader should be cautious that the original library generation (at least) multiplies with the number of truncation iterations.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aleman, D. M., H. R. Ghaffari, V. V. Mišić, M. B. Sharpe, M. Ruschin, and D. A. Jaffray (2013), Optimization methods for large-scale radiotherapy problems, in *Systems analysis tools for better health care delivery*, pp. 1–20, Springer.

Barendsen, G. (1982), Dose fractionation, dose rate and iso-effect relationships for normal tissue responses, *International Journal of Radiation Oncology* Biology* Physics*, *8*(11), 1981–1997.

Bokrantz, R., and A. Forsgren (2013), An algorithm for approximating convex pareto surfaces based on dual techniques, *INFORMS Journal on Computing*, *25*(2), 377–393.

Bortfeld, T. (2006), Imrt: a review and preview, *Physics in medicine and biology*, *51*(13), R363.

Bowen, S. R., et al. (2015), Differential hepatic avoidance radiation therapy: Proof of concept in hepatocellular carcinoma patients, *Radiotherapy and Oncology*, *115*(2), 203–210.

Breedveld, S., P. R. Storchi, M. Keijzer, A. W. Heemink, and B. J. Heijmen (2007), A novel approach to multi-criteria inverse planning for imrt, *Physics in medicine and biology*, *52*(20), 6339.

Breedveld, S., P. R. Storchi, and B. J. Heijmen (2009), The equivalence of multi-criteria methods for radiotherapy plan optimization, *Physics in medicine and biology*, *54*(23), 7199.

Brock, K., M. Sharpe, L. Dawson, S. Kim, and D. Jaffray (2005), Accuracy of finite element model-based multi-organ deformable image registration, *Medical physics*, *32*(6), 1647–1659.

Burridge, N., A. Amer, T. Marchant, J. Sykes, J. Stratford, A. Henry, C. McBain, P. Price, and C. Moore (2006), Online adaptive radiotherapy of the bladder: small bowel irradiated-volume reduction, *International Journal of Radiation Oncology* Biology* Physics*, *66*(3), 892–897.

Cao, Y., J. Alspaugh, Z. Shen, J. M. Balter, T. S. Lawrence, and R. K. Ten Haken (2006), A practical approach for quantitative estimates of voxel-by-voxel liver perfusion using dce imaging and a compartmental model, *Medical physics*, *33*(8), 3057–3062.

Cao, Y., et al. (2008), Liver Function After Irradiation Based on Computed Tomographic Portal Vein Perfusion Imaging, *International Journal of Radiation Oncology* Biology* Physics*, *70*(1), 154–160.

Cao, Y., et al. (2013), Prediction of Liver Function by Using Magnetic Resonance-Based Portal Venous Perfusion Imaging, *International Journal of Radiation Oncology, Biology, Physics*, *85*(1), 258–263.

Chan, T. C., and V. V. Mišić (2013), Adaptive and robust radiation therapy optimization for lung cancer, *European Journal of Operational Research*, *231*(3), 745–756.

Chen, G.-P., E. Ahunbay, C. Schultz, and X. A. Li (2007), Development of an inverse optimization package to plan nonuniform dose distributions based on spatially inhomogeneous radiosensitivity extracted from biological images, *Medical physics*, *34*(4), 1198–1205.

Chen, M., W. Lu, Q. Chen, K. Ruchala, and G. Olivera (2008), Adaptive fractionation therapy: Ii. biological effective dose, *Physics in medicine and biology*, *53*(19), 5513.

Clark, V., Y. Chen, J. Wilkens, J. Alaly, K. Zakaryan, and J. Deasy (2008), Imrt treatment planning for prostate cancer using prioritized prescription optimization and mean-tail-dose functions, *Linear algebra and its applications*, *428*(5-6), 1345–1364.

Craft, D., T. Halabi, and T. Bortfeld (2005), Exploration of tradeoffs in intensity-modulated radiotherapy, *Physics in medicine and biology*, *50*(24), 5857.

Craft, D., T. Halabi, H. A. Shih, and T. Bortfeld (2007), An approach for practical multiobjective IMRT treatment planning, *International journal of radiation oncology* Biology* Physics*, *69*(5), 1600–1607.

Craft, D. L., T. S. Hong, H. A. Shih, and T. R. Bortfeld (2012), Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy, *International Journal of Radiation Oncology* Biology* Physics*, *82*(1), e83–e90.

Das, S., et al. (2004), Feasibility of optimizing the dose distribution in lung tumors using fluorine-18-fluorodeoxyglucose positron emission tomography and single photon emission computed tomography guided dose prescriptions, *Medical physics*, *31*(6), 1452–1461.

Deasy, J. (1997), Multiple local minima in radiotherapy optimization problems with dose–volume constraints, *Medical physics*, *24*(7), 1157–1161.

Deufel, C. L., and K. M. Furutani (2014), Quality assurance for high dose rate brachytherapy treatment planning optimization: using a simple optimization to verify a complex optimization, *Physics in medicine and biology*, *59*(3), 525.

Dinkla, A. M., R. Laarse, K. Koedooder, H. Petra Kok, N. Wieringen, B. R. Pieters, and A. Bel (2015), Novel tools for stepping source brachytherapy treatment planning: Enhanced geometrical optimization and interactive inverse planning, *Medical physics*, *42*(1), 348–353.

D'Souza, W. D., R. Meyer, B. R. Thomadsen, and M. Ferris (2001), An iterative sequential mixed-integer approach to automated prostate brachytherapy treatment plan optimization, *Physics in medicine and biology*, *46*(2), 297.

Engberg, L., A. Forsgren, K. Eriksson, and B. Hårdemark (2017), Explicit optimization of plan quality measures in intensity-modulated radiation therapy treatment planning, *Medical Physics*.

Feller, W. (1950), Probability theory and its applications, *New York*, pp. 118–123.

Feng, M., F.-M. Kong, M. Gross, S. Fernando, J. A. Hayman, and R. K. Ten Haken (2009), Using fluorodeoxyglucose positron emission tomography to assess tumor volume during radiotherapy for non–small-cell lung cancer and its potential impact on adaptive dose escalation and normal tissue sparing, *International Journal of Radiation Oncology* Biology* Physics*, *73*(4), 1228–1234.

Fertil, B., and E.-P. Malaise (1981), Inherent cellular radiosensitivity as a basic concept for human tumor radiotherapy, *International Journal of Radiation Oncology* Biology* Physics*, *7*(5), 621–629.

Foroudi, F., et al. (2009), Offline adaptive radiotherapy for bladder cancer using cone beam computed tomography, *Journal of medical imaging and radiation oncology*, *53*(2), 226–233.

Fowler, J. F. (1989), The linear-quadratic formula and progress in fractionated radiotherapy, *The British journal of radiology*, *62*(740), 679–694.

Giantsoudi, D., D. Baltas, A. Karabis, P. Mavroidis, N. Zamboglou, N. Tselis, C. Shi, and N. Papanikolaou (2013), A geud-based inverse planning technique for hdr prostate brachytherapy: Feasibility study, *Medical physics*, *40*(4).

Gill, S., D. Pham, K. Dang, M. Bressel, T. Kron, S. Siva, P. K. Tran, K. H. Tai, and F. Foroudi (2013), Plan of the day selection for online image-guided adaptive post-prostatectomy radiotherapy, *Radiotherapy and Oncology*, *107*(2), 165–170.

Haimes, Y. Y., L. Ladson, and D. A. Wismer (1971), Bicriterion formulation of problems of integrated system identification and system optimization, *IEEE Transactions on Systems Man and Cybernetics*, *1*(3), 296.

Holm, Å., T. Larsson, and Å. C. Tedgren (2013), A linear programming model for optimizing hdr brachytherapy dose distributions with respect to mean dose in the dvh-tail, *Medical physics*, *40*(8), 081,705.

HSL, A. (2013), collection of fortran codes for large-scale scientific computation, *See http://www. hsl. rl. ac. uk*.

Ireland, R. H., C. M. Bragg, M. McJury, N. Woodhouse, S. Fichele, E. J. Van Beek, J. M. Wild, and M. Q. Hatton (2007), Feasibility of Image Registration and Intensity-Modulated Radiotherapy Planning With Hyperpolarized Helium-3 Magnetic Resonance Imaging for Non–Small-Cell Lung Cancer, *International Journal of Radiation Oncology, Biology, Physics*, *68*(1), 273–281.

Jackson, A., R. Ten Haken, J. Robertson, M. Kessler, G. Kutcher, and T. Lawrence (1995), Analysis of Clinical Complication Data for Radiation Hepatitis Using a Parallel Architecture Model, *International Journal of Radiation Oncology, Biology, Physics*, *31*(4), 883–891.

Janssens, G., J. Orban de Xivry, S. Fekkes, A. Dekker, B. Macq, P. Lambin, and W. van Elmpt (2009), Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy, *Medical physics*, *36*(9), 4268–4276.

Jones, B., R. Dale, C. Deehan, K. Hopkins, and D. Morgan (2001), The role of biologically effective dose (bed) in clinical oncology, *Clinical oncology*, *13*(2), 71–81.

Karabis, A., S. Giannouli, and D. Baltas (2005), 40 hipo: A hybrid inverse treatment planning optimization algorithm in hdr brachytherapy, *Radiotherapy and Oncology*, *76*, S29.

Karabis, A., P. Belotti, and D. Baltas (2009), Optimization of catheter position and dwell time in prostate hdr brachytherapy using hipo and linear programming, in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pp. 612–615, Springer.

Kessler, M. L., D. L. Mcshan, M. A. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass (2005), Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans, *Optimization and Engineering*, *6*(4), 421–448.

Kim, J., N. Dogan, D. McShan, and M. Kessler (1995), An avs-based system for optimization of conformal radiotherapy treatment plans, *Proceedings of AVS 95*, pp. 417–431.

Kim, M., A. Ghate, and M. Phillips (2009), A markov decision process approach to temporal modulation of dose fractions in radiation therapy planning, *Physics in Medicine and Biology*, *54*(14), 4455.

Kim, M., A. Ghate, and M. H. Phillips (2012), A stochastic control formalism for dynamic biologically conformal radiation therapy, *European Journal of Operational Research*, *219*(3), 541–556.

Küfer, K.-H., M. Monz, A. Scherrer, P. Süss, F. Alonso, A. S. A. Sultan, T. Bortfeld, and C. Thieke (2009), Multicriteria optimization in intensity modulated rediotherapy planning, in *Handbook of optimization in medicine*, pp. 1–45, Springer.

Kuyumcian, A., D. Pham, J. M. Thomas, A. Law, D. Willis, T. Kron, and F. Foroudi (2012), Adaptive radiotherapy for muscle-invasive bladder cancer: Optimisation of plan sizes, *Journal of medical imaging and radiation oncology*, *56*(6), 661–667.

Lahanas, M., D. Baltas, and N. Zamboglou (1999), Anatomy-based three-dimensional dose optimization in brachytherapy using multiobjective genetic algorithms, *Medical Physics*, *26*(9), 1904–1918.

Lahanas, M., D. Baltas, and S. Giannouli (2003), Global convergence analysis of fast multiobjective gradient-based dose optimization algorithms for high-dose-rate brachytherapy, *Physics in medicine and biology*, *48*(5), 599.

Langen, K., and D. Jones (2001), Organ motion and its management, *International Journal of Radiation Oncology* Biology* Physics*, *50*(1), 265–278.

Langer, M., V. Thai, and L. Papiez (2001), Improved leaf sequencing reduces segments or monitor units needed to deliver IMRT using multileaf collimators, *Medical Physics*, *28*(12), 2450–2458.

Lee, E. K., and M. Zaider (2003), Mixed integer programming approaches to treatment planning for brachytherapy–application to permanent prostate implants, *Annals of operations research*, *119*(1-4), 147–163.

Lee, E. K., R. J. Gallagher, D. Silvern, C.-S. Wuu, and M. Zaider (1999), Treatment planning for brachytherapy: an integer programming model, two computational approaches and experiments with permanent prostate implant planning, *Physics in Medicine and Biology*, *44*(1), 145.

Lee, E. K., T. Fox, and I. Crocker (2003), Integer programming applied to intensity-modulated radiation therapy treatment planning, *Annals of Operations Research*, *119*(1-4), 165–181.

Lei, Y., and Q. Wu (2010), A hybrid strategy of offline adaptive planning and online image guidance for prostate cancer radiotherapy, *Physics in medicine and biology*, *55*(8), 2221.

Li, X. A., et al. (2012), The use and QA of biologically related models for treatment planning: Short report of the tg-166 of the therapy physics committee of the aapm a), *Medical Physics*, *39*(3), 1386–1409.

Ling, C. C., J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, and J. A. Koutcher (2000), Towards multidimensional radiotherapy (md-crt): biological imaging and biological conformality, *International Journal of Radiation Oncology* Biology* Physics*, *47*(3), 551–560.

Liu, E., et al. (2013), Stereotactic body radiation therapy for primary and metastatic liver tumors, *Translational oncology*, *6*(4), 442–446.

Long, T., M. Matuszak, M. Feng, B. Fraass, R. Ten Haken, and H. Romeijn (2012), Sensitivity analysis for lexicographic ordering in radiation therapy treatment planning, *Medical physics*, *39*(6), 3445–3455.

Long, T. C. (2015), Optimization problems in radiation therapy treatment planning, Ph.D. thesis, The University of Michigan.

McShan, D., M. Kessler, K. Vineberg, and B. Fraass (2006), Inverse plan optimization accounting for random geometric uncertainties with a multiple instance geometry approximation (miga), *Medical physics*, *33*(5), 1510–1521.

Men, C., H. E. Romeijn, A. Saito, and J. F. Dempsey (2012), An efficient approach to incorporating interfraction motion in IMRT treatment planning, *Computers & Operations Research*, *39*(7), 1779–1789.

Miften, M. M., S. K. Das, M. Su, and L. B. Marks (2004), Incorporation of Functional Imaging Data in the Evaluation of Dose Distributions Using the Generalized Concept of Equivalent Uniform Dose, *Physics in Medicine and Biology*, *49*(9), 1711.

Milickovic, N., M. Lahanas, M. Papagiannopoulou, N. Zamboglou, and D. Baltas (2002), Multiobjective anatomy-based dose optimization for hdr-brachytherapy with constraint free deterministic algorithms, *Physics in medicine and biology*, *47*(13), 2263.

Monz, M., K. Küfer, T. Bortfeld, and C. Thieke (2008), Pareto navigationalgorithmic foundation of interactive multi-criteria IMRT planning, *Physics in medicine and biology*, *53*(4), 985.

Mosek, A. (2010), The mosek optimization software, *Online at http://www. mosek. com*, *54*, 2–1.

Murthy, V., Z. Master, P. Adurkar, I. Mallick, U. Mahantshetty, G. Bakshi, H. Tongaonkar, and S. Shrivastava (2011), plan of the dayadaptive radiotherapy for bladder cancer using helical tomotherapy, *Radiotherapy and Oncology*, *99*(1), 55–60.

Mutic, S., and J. F. Dempsey (2014), The viewray system: Magnetic resonance–guided and controlled radiotherapy, in *Seminars in radiation oncology*, vol. 24, pp. 196–199, Elsevier.

Niemierko, A. (1999), A Generalized Concept of Equivalent Uniform Dose (EUD), *Medical Physics*, *26*(6), 1100.

Pandharipande, P. V., G. A. Krinsky, H. Rusinek, and V. S. Lee (2005), Perfusion imaging of the liver: current challenges and future goals 1, *Radiology*, *234*(3), 661–673.

Pflug, G. C. (2000), Some remarks on the value-at-risk and the conditional value-at-risk, in *Probabilistic constrained optimization*, pp. 272–281, Springer.

Ripsman, D. A., D. M. Aleman, and K. Ghobadi (2015), Interactive visual guidance for automated stereotactic radiosurgery treatment planning, *Expert Systems with Applications*, *42*(21), 8337–8348.

Rockafellar, R. T., and R. J.-B. Wets (1991), Scenarios and policy aggregation in optimization under uncertainty, *Mathematics of operations research*, *16*(1), 119–147.

Romeijn, H. E., and J. F. Dempsey (2008), Intensity modulated radiation therapy treatment plan optimization, *Top*, *16*(2), 215.

Romeijn, H. E., R. K. Ahuja, J. F. Dempsey, A. Kumar, and J. G. Li (2003), A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning, *Physics in Medicine and Biology*, *48*(21), 3521.

Romeijn, H. E., J. F. Dempsey, and J. G. Li (2004), A unifying framework for multi-criteria fluence map optimization models, *Physics in Medicine and Biology*, *49*(10), 1991.

Romeijn, H. E., R. K. Ahuja, J. F. Dempsey, and A. Kumar (2006), A new linear programming approach to radiation therapy treatment planning problems, *Operations Research*, *54*(2), 201–216.

Ruotsalainen, H., K. Miettinen, J.-E. Palmgren, and T. Lahtinen (2010), Interactive multiobjective optimization for anatomy-based three-dimensional hdr brachytherapy, *Physics in medicine and biology*, *55*(16), 4703.

Saberian, F., A. Ghate, and M. Kim (2015), Spatiotemporally integrated fractionation in radiotherapy, *INFORMS J. Comput.*

Samavati, N., M. Velec, and K. Brock (2015), A hybrid biomechanical intensity based deformable image registration of lung 4dct, *Physics in medicine and biology*, *60*(8), 3359.

Seppenwoolde, Y., et al. (2002), Optimizing Radiation Treatment Plans for Lung Cancer Using Lung Perfusion Information, *Radiotherapy and Oncology*, *63*(2), 165–177.

Siebers, J. V., M. Lauterbach, P. J. Keall, and R. Mohan (2002), Incorporating multi-leaf collimator leaf sequencing into iterative IMRT optimization, *Medical physics*, *29*(6), 952–959.

Sir, M. Y., M. A. Epelman, and S. M. Pollock (2012), Stochastic programming for off-line adaptive radiotherapy, *Annals of Operations Research*, *196*(1), 767–797.

Stroom, J. C., and B. J. Heijmen (2002), Geometrical uncertainties, radiotherapy planning margins, and the icru-62 report, *Radiotherapy and oncology*, *64*(1), 75–83.

Thieke, C., T. Bortfeld, and K.-H. Küfer (2002), Characterization of Dose Distributions Through the Max and Mean Dose Concept, *Acta Oncologica*, *41*(2), 158–161.

Thieke, C., K.-H. Küfer, M. Monz, A. Scherrer, F. Alonso, U. Oelfke, P. E. Huber, J. Debus, and T. Bortfeld (2007), A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation, *Radiotherapy and Oncology*, *85*(2), 292–298.

Unkelbach, J., and U. Oelfke (2005), Incorporating organ movements in IMRT treatment planning for prostate cancer: minimizing uncertainties in the inverse planning process, *Medical physics*, *32*(8), 2471–2483.

Unkelbach, J., and D. Papp (2015), The emergence of nonuniform spatiotemporal fractionation schemes within the standard bed model, *Medical physics*, *42*(5), 2234–2241.

Unkelbach, J., D. Craft, E. Salari, J. Ramakrishnan, and T. Bortfeld (2012), The dependence of optimal fractionation schemes on the spatial dose distribution, *Physics in medicine and biology*, *58*(1), 159.

Van Herk, M. (2004), Errors and margins in radiotherapy, in *Seminars in radiation oncology*, vol. 14, pp. 52–64, Elsevier.

Wächter, A., and L. T. Biegler (2006), On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Mathematical programming*, *106*(1), 25–57.

Wang, H., M. Feng, A. Jackson, R. K. Ten Haken, T. S. Lawrence, and Y. Cao (2016), Local and global function model of the liver, *International Journal of Radiation Oncology* Biology* Physics*, *94*(1), 181–188.

Wu, Q. J., D. Thongphiew, Z. Wang, B. Mathayomchan, V. Chankong, S. Yoo, W. R. Lee, and F.-F. Yin (2008), On-line re-optimization of prostate IMRT plans for adaptive radiation therapy, *Physics in medicine and biology*, *53*(3), 673.

Xia, P., and L. J. Verhey (1998), Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments, *Medical Physics*, *25*(8), 1424–1434.

Yang, C., et al. (2014), Combined online and offline adaptive radiation therapy: A dosimetric feasibility study, *Practical radiation oncology*, *4*(1), e75–e83.

Zhang, J., et al. (2010), Radiation-induced reductions in regional lung perfusion: 0.1–12 year data from a prospective clinical study, *International Journal of Radiation Oncology* Biology* Physics*, *76*(2), 425–432.

Zhong, H., and I. J. Chetty (2017), Caution must be exercised when performing deformable dose accumulation for tumors undergoing mass changes during fractionated radiation therapy, *International Journal of Radiation Oncology Biology Physics*, *97*(1), 182–183.