Distributionally Robust Chance Constraints for Radiotherapy Treatment Planning under Geometric Uncertainty

by

Stoyan Hristov Hristov

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2025

© Stoyan Hristov Hristov 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Radiation therapy is a popular treatment modality for cancer which benefits from optimization. Mainstream approaches define a margin around the tumor called the planning target volume (PTV) to account for changes in patient anatomy throughout the course of treatment. However, the PTV concept does not adequately address uncertainty in the position of the tumor or in the probability distribution of its shifts.

In this thesis, we develop a distributionally robust optimization framework for radiation therapy treatment planning under inter-fraction geometric uncertainty of the clinical target volume (CTV) where the underlying probability distribution of rigid shifts is not well known. To achieve this, we begin with a deterministic model, then augment it using chance-constrained optimization. Next, we consider an ambiguous chance constraint and apply its robust reformulation to treatment planning models.

We first apply these three frameworks to a basic fluence map optimization model. Then, we include machine deliverability constraints and formulate a direct aperture optimization model. Due to the large-scale combinatorial structure, these models cannot be solved efficiently at a clinical scale, making them unusable for modern treatment approaches such as volumetric-modulated arc therapy. Therefore, we extend a sequential convex programming algorithm to include robust and distributionally robust optimization.

Through experiments on two lung cancer patients, we show that the novel distributionally robust treatment planning model can achieve better tumor coverage than the robust model as well as better healthy tissue sparing than the standard PTV approach. We also show that adding robust and distributionally robust optimization does not degrade the quality of the convex approximation. We conclude with a discussion on limitations and present ideas for future research projects.

Acknowledgments

Accepting a business student to an engineering research lab is probably not an intuitive decision. Consequently, Prof. Mahmoudzadeh took a bit of a risk in hiring me. Over the last two years, her patient mentorship shaped me both as an aspiring researcher and as a person. Houra, thank you so much for your support and for giving me every opportunity that came your way. Your dedication to excellence in each facet of being a professor, from research to teaching to mentorship, has been so inspiring. As I grow in my career, I truly hope to strike a similar balance of sincerity and skill.

As operations research students, we often work towards becoming experts in optimization and modeling. I believe, however, that without the potential for societal impact, these models lose their meaning. To that end, I would like to appreciate our partners Dr. Osei and Dr. Darko for their meaningful collaboration. Ernest and Johnson, thank you for sharing your ideas and passion. In addition to shaping the research questions in this thesis, your coaching showed me so many interesting things about medicine and cancer treatment.

I also wanted to give thanks to my readers, Prof. Alumur Alev and Prof. Mehrizi for their willingness and time in reading my thesis. Sibel, your enthusiasm is infectious; thank you for all your encouragement and kind words. Hossein, thank you for the insightful questions and for pushing me to see the bigger picture beyond the specific methodologies in this thesis.

Many times, I worked at the office until late at night. Some of those times, I questioned why I was doing this to myself. Every time those questions popped up, I bumped into one of the janitors who always struck up conversation. Sinisa, thank you for your effort to engage with students, it makes more of an impact than you think! Thank you also for the encouragement, pride, and for empathizing with the European immigrant mentality. I wish you and your family all the best.

As we have seen, research seldom happens in isolation, yet the process itself can be lonesome. I owe much of my remaining sanity to the friends I made here at Waterloo. Bradley and Gatik, thank you for the camaraderie, the fun chats, and for always having

my back. Dani, thank you for the bad jokes and good advice, as well as the occasional good jokes and bad advice. I really enjoyed our chaotic sibling-esque exchanges as well as the deeper conversations. Thank you, especially, for always challenging my imposter syndrome and for helping me learn more about how my brain works.

I would also like to thank Fuat, who deserves his own paragraph, but not much more. The first time we met virtually, he recognized my very niche band shirt and then questioned my math skills. Neither had ever happened to me before, and I quickly came to think of him as an older brother. Fuat, thank you for being the strongest force against the infamous Waterloo effect. Thank you also for the unconditional kindness. From boxing nights to unhinged conversations to optimal vitamin D intake to haircut critiques to philosophical movies, our friendship is one of the best highlights of my time here. We also make an unmatched teaching team; I'll really miss cooking up tutorials together and discussing the progression of education.

As a student, I made the most of every break between terms. Every time I visited home, my inner circle would find ways to remind me that there is more to life than this degree. Thank you especially to my best friends: Winson, Martin, Powys, and Crystal. I am so lucky to have you in my life. Of course, each visit home had to be accompanied by burger nights. For those memories, as well as the banter that ensued, I would like to thank Jack, Austin, and Marlon.

There is some excellent food here in Waterloo, too. QuickSandwiches, for instance, was my favorite place to eat. Thank you to Ishan and his parents, Rama and Gurinder, for hosting this lovely shop. Every time I walked in, the only thing that mattered was leaving happy and with a full stomach. The love and care they pour into their food shows with every bite.

Switching gears, I feel I owe a very special thank you to Forough, whose advice has been consistently shaping my career trajectory. We met back in 2018 when I was a first-year undergraduate student slightly confused by linear programming but very curious about geometry. Luckily, Forough's office hours answered all of my questions and started a meaningful friendship. Forough, thank you for encouraging me to pursue research and for pointing me in Houra's direction. Thank you, also, for empathetically listening to my

many rambles about PhD programs and for providing so much support. I wouldn't be here without you!

One thing I noticed about engineering education is that there is often a disregard for creativity. If unchecked, consistent exposure can dull the human parts of our brains. As such, I would like to thank some of the communities and artists that reignited that part of my brain. Thank you to Prof. Cameron Shelley and the Centre for Society, Technology & Values for hosting intriguing movie nights and introducing me to Plato's allegory of the cave. I wanted to express appreciation for π (1998) and The Prestige for capturing some of what it means to be human. Thank you also to LP, Tesseract, RATM, ETC, LOG, SOAD, Ер малък, and FYRE for being the soundtrack of my last two years. Last, I would like to commemorate Ozzy Osbourne, who passed away the week of my defense. I'm thankful for his lasting influence on my music taste and for how he modeled resilience.

Everyone, from my high school math teacher to my Taekwondo instructor, warned me about how depressing Waterloo can get. They were all right. Spending most of my time in an office with no windows, working deeply on a project that only one other person really cares about, and being far away from everyone I know was not easy. I stumbled a lot, but, thankfully, I found some incredible support systems. I would like to extend a very heartfelt thank you to Brenda, Jess, Sean, and everyone from the fellowship meetings for helping with recovery.

Words can't fully express how grateful I am to my partner for the unconditional love, support, and patience throughout the last two years. Matina, thank you for being such a nurturing presence as we uproot our bitter trees and cultivate our garden. ♡

Finally, I would like to thank my parents, Hristo Stoyanov Hristov and Tanya Kirova Hristova, for everything they sacrificed so that I could pursue these opportunities. Thank you for exemplifying diligence, for pushing me hard, and for trusting me to move across the country and study something so seemingly foreign.

In Memoriam

In loving memory of Professor Emeritus David Gillen. Thank you for your kindness, warmth, and encouragement to pursue graduate studies.

Dedication

I would like to dedicate this thesis to my little brother Alex. May you always have the courage and security to follow your dreams.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgments iv

In Memoriam vii

Dedication viii

List of Figures xiii

List of Tables xvi

List of Abbreviations xviii

1 Introduction 1

2 Literature Review 4

2.1 Optimization in Intensity-Modulated Radiation Therapy 4

2.2 Optimization in Volumetric-Modulated Arc Therapy . . . 5
2.3 Uncertainty Mitigation in Radiation Therapy . . . 7
2.4 Gap, Contributions, and Scope . . . 10

3 Problem Formulation **14**

3.1 Deterministic, Robust, and Distributionally Robust Chance Constraint Derivations . . . 14
3.1.1 Deterministic Model . . . 14
3.1.2 Robust Chance-Constrained Model . . . 15
3.1.3 Distributionally Robust Chance-Constrained Model . . . 16
3.2 Radiotherapy Problem Setting . . . 19
3.3 Fluence Map optimization (FMO) Models . . . 24
3.3.1 Deterministic FMO Model . . . 24
3.3.2 Robust Chance-Constrained FMO Model . . . 25
3.3.3 Distributionally Robust Chance-Constrained FMO Model . . . 26
3.4 Direct Aperture optimization (DAO) Models . . . 28
3.4.1 Model Preliminaries . . . 28
3.4.2 Deterministic DAO Model . . . 30
3.4.3 Robust Chance-Constrained DAO Model . . . 32
3.4.4 Distributionally Robust Chance-Constrained DAO Model . . . 34
3.4.5 Model Size . . . 35

4 Solution Methodology **37**

4.1 Deterministic Sequential Convex Programming Algorithm . . . 37

4.2 Incorporating Robustness in the Intermediate and Correction Problems . . 42
4.3 Incorporating Robustness in the Feasibility Check 43
4.4 Implementation of the Robust Sequential Convex Programming Algorithm 44

5 Results **46**

5.1 Fluence Map Optimization 46
5.1.1 Experimental Design 47
5.1.2 Dose-Volume Histograms 48
5.1.3 Sensitivity Analysis and Out-of-sample Performance 50
5.2 Direct Aperture optimization 54
5.2.1 Experimental Design 54
5.2.2 Dose-Volume Histograms 55
5.2.3 Sensitivity Analysis and Out-of-sample Performance 60
5.3 Sequential Convex Programming 63
5.3.1 Experimental Design 64
5.3.2 Dose-Volume Histograms 65
5.3.3 Sensitivity Analysis and Out-of-sample Performance 67
5.3.4 Quality of Convex Approximation 71
5.3.5 Computational Times 73

6 Extensions **77**

6.1 Smooth Fluence Map Optimization 78
6.2 Neighboring Aperture Similarity 79
6.3 Towards Clinical Realism 82

7 Conclusion **84**

References **86**

APPENDICES **93**

A Notations **94**

A.1 Sets 94

A.2 Parameters 95

A.3 Variables 96

List of Figures

2.1 The main approaches to IMRT planning with red indicating the scope of this thesis. 12

2.2 A visualization of the optimization outputs in each step of each approach. . 13

3.1 A comparison of the clinical setting and various treatment planning approaches. 22

3.2 An example of the DRO set-up with five possible realizations in the CTV position. 23

3.3 Visual schematic of mapping aperture intensity and leaf positions to beamlet intensities. 29

3.4 An example of a feasible aperture. Orange beamlets are on, blue beamlets are off, and the grey rectangles represent the MLC leaves. Note: rows 7 and 8 present an effect called *interdigitation*. If interdigitation occurs excessively, it may damage the MLC leaves. This effect can be discouraged through feasibility constraints. 30

4.1 An example aperture using the convex approximation formulation. Here, orange beamlets are combined to represents the interior beamlet, which remains on throughout subsequent iterations. Navy beamlets can be combined to represent the exterior beamlet, which remains off throughout subsequent iterations. Green beamlets can be combined to represent the boundary beamlets. 39

5.1 A visual depiction of each model. 48
5.2 Depiction of the dose distribution received by the CTV and all OARs if the nominal position of the ctv is realized. 49
5.3 Full dose distribution resulting from FMO models in the nominal setting, in-sample shifts, and out-of-sample shifts. 51
5.4 A summary of the mean dose and bottom CVaR of the CTV under each model. 52
5.5 Box-plots representing the dose distribution to each OAR when treated by each model. 55
5.6 Depiction of the dose distribution received by the CTV and all OARs if the nominal position of the CTV is realized. 57
5.7 Full dose distribution resulting from FMO models in the nominal setting, in-sample shifts, and out-of-sample shifts. 59
5.8 A summary of the mean dose and bottom CVaR of the CTV under each model. 61
5.9 Box-plots representing the dose distribution to each OAR when treated by each model. 63
5.10 Depiction of the dose distribution received by the CTV and all OARs if the nominal position of the CTV is realized. 66
5.11 Full dose distribution resulting from SCP models in the nominal setting, in-sample shifts, and out-of-sample shifts. 68
5.12 A summary of the mean dose and bottom CVaR of the CTV under each model. 69
5.13 Box-plots representing the dose distribution to each OAR when treated by each model. 72
5.14 Comparison of the dose distributions resulting from the ground-truth and the convex approximation models at each level of uncertainty. 74

5.15 Convergence behavior for N-SCP, R-SCP, and DR-SCP variants. 75

6.1 Side-by-side comparison of an example fluence map resulting from FMO models that do (left) and do not (right) consider smoothness. 78

List of Tables

3.1 Problem size of each MILP instance denoting the number of individual continuous, binary, and integer variables as well as the number of individual constraints. 36

5.1 FMO models tested in clinical-scale experiment. 47

5.2 The mean dose received by the 5% least-dosed voxels in the CTV under each shift when treated by each model. Results are separated by nominal, in-sample, and out-of-sample realizations. Bold indicates that the CVaR requirement is satisfied in that scenario. 53

5.3 DAO models tested in clinical-scale experiment. 56

5.4 Problem size of each MILP instance, split into full and small scales for continuous, binary, and integer variables, as well as constraints, after compilation in CVXPy. 56

5.5 The mean dose received by the 10% least-dosed voxels in the CTV under each shift when treated by each model. Results are separated by nominal, in-sample, and out-of-sample realizations. 62

5.6 Models tested in clinical-scale experiment. 64

5.7 The mean dose received by the 10% least-dosed voxels in the CTV under each shift when treated by each model. Results are separated by nominal, in-sample, and out-of-sample realizations. Bold indicates that the CVaR requirement is satisfied. 70

5.8 Comparison of optimality gaps between MILP and SCP approaches. 73

5.9 Solution times in seconds for each model at both scales explored. Entries marked with OOM indicate an out-of-memory error when attempting to solve the model with 512 GB of RAM. 75

A.1 Notations used for sets. 94

A.2 Notations used for parameters. 95

A.3 Notations used for decision variables. 96

List of Abbreviations

CC chance constraint 9, 70

CCP chance-constrained programming 3, 34, 84

CP control point 2, 6, 11, 40, 80, 81

CT computed tomography 2, 3, 7

CTV clinical target volume 7, 9–11, 21, 22, 24, 25, 47, 49, 50, 53, 56–58, 60, 62, 64, 65, 67, 69, 71, 82, 83

CVaR conditional value at risk 5, 9, 21, 25, 26, 43, 44, 47, 49–51, 53, 54, 57, 58, 60, 62, 64, 65, 67, 69, 70, 96

DAO direct aperture optimization 5, 6, 10, 14, 23, 27, 28, 31, 32, 46, 54, 57, 71, 73, 84

DRCCP distributionally robust chance-constrained program 9, 10, 45

DRO distributionally robust optimization 11, 40, 41, 45, 48, 55, 65, 73

DVH dose-volume histogram 5, 21, 46, 48–50, 55, 57, 58, 63, 65, 67

FMO fluence map optimization 4, 5, 10, 11, 14, 22, 24, 26, 28, 32, 46, 47, 49, 50, 54, 77–79, 84

GTV gross target volume 7, 47, 64

IMRT intensity-modulated radiation therapy 2, 4, 5, 7, 11, 47

LS leaf sequencing 4, 5, 11, 24, 27, 28, 47, 77–79

MILP mixed-integer linear program 37, 38, 64, 71, 80

MIP mixed-integer programming 6, 9, 77, 80, 84

MIQP mixed-integer quadratic program 82

MLC multi-leaf collimator 4, 5, 28, 30, 31, 79, 80, 82

MSP minimax stochastic programming 9

OAR organ at risk 2, 3, 7, 8, 19–21, 49, 54, 57, 58, 62, 63, 65, 71, 84, 85

PTV planning target volume 7–11, 22, 25, 49, 50, 53, 57, 58, 60, 62, 64, 65, 67, 69–71

RO robust optimization 3, 8, 9, 11, 40, 41, 48, 55, 65, 73

RT radiation therapy 1, 2, 4, 10, 19–22, 24, 77, 82

SCP sequential convex programming 6, 23, 37, 41, 45, 46, 63, 64, 67, 71, 73, 77, 80–82, 85

SP stochastic programming 3, 8, 9

SW sliding windows 5, 6

VMAT volumetric-modulate radiation therapy 2, 4–6, 9–11, 14, 23, 28, 31, 45, 64, 67, 76, 77, 79, 81, 84

Chapter 1

Introduction

With approximately one-fifth of people developing cancer in their lifetime (World Health Organization and International Agency for Research on Cancer, 2024), cancer is a leading cause of mortality worldwide. The number of new diagnoses is predicted to increase by 77% from 2022 to 2050, largely due to the growing and aging population as well as environmental factors such as air pollution (World Health Organization and International Agency for Research on Cancer, 2024).

Cancer is a pressing societal problem in Canada too, and an increase in treatment quality can have direct benefits to the Canadian economy. More than one third of Canadians develop cancer in their lifetime (National Research Council Canada, 2023), and the number of new diagnoses was nearly doubled between 2000 and 2024 (Brenner et al., 2024). According to the Canadian Cancer Society (2024), the projected total societal cost of cancer in Canada is $37.7B. Approximately 80% of the societal cost is a direct cost to the Canadian healthcare system, whereas the remaining 20% is a mixture of direct, indirect, and time costs to people with cancer and their caregivers. This motivates the higher quality treatments to improve efficiency.

Approximately half of all patients diagnosed with cancer undergo radiation therapy (RT) treatment, either alone or in conjunction with other treatment methods such as chemotherapy or hormone therapy (National Research Council Canada, 2023). In RT, a beam of high-energy radiation is used to irradiate the cancerous cells to eliminate these

cells and stop the spread of cancer, however, the radiation can also damage any healthy cells on its path. The goal of RT is to deliver a sufficient radiation dose to one or multiple tumors (cancerous targets) while sufficiently sparing the nearby organ at risks (OARs) from excessive radiation. As such, optimizing RT treatment plans is important for achieving personalized high-quality treatments.

A type of radiation therapy treatment modality known as intensity-modulated radiation therapy (IMRT) uses a rotating gantry around the patient to deliver radiation through a photon beam. The shape of each beam is tailored to the shape of the target as captured on a computed tomography (CT) scan at that angle (Mayo Clinic, 2025). In IMRT planning, the radiation beam is discretized into a two-dimensional grid of beamlets, where the intensity of each beamlet can be controlled independently. The organs of interest are delineated on the CT scan, and each organ is then discretized into three-dimensional cubes, called voxels. Subsequently, a dose deposition matrix is calculated that details the amount of radiation each beamlet delivers to each voxel at unit intensity. Then, an optimization model is solved to prescribe the optimal intensities of each beamlet at each angle. Once a treatment plan is designed, it is divided into fractions and delivered to the patient over multiple daily consecutive treatment sessions that often span weeks (Bortfeld, 2006).

There are multiple variations of the IMRT approach, as outlined in Bortfeld (2006). The most popular IMRT modality is the step-and-shoot IMRT, in which a set of beam angles is chosen in advance, and the radiation is only delivered to the patient when the beam is positioned at those angles. The beam is then turned off while it moves to the next angle chosen in the treatment plan. The beamlet intensity is linked to a multileaf collimator (MLC), which modulates individual leaves to define apertures that tailor the shape of the radiation beam. In contrast, volumetric-modulate radiation therapy (VMAT) is a more recent form of IMRT where the gantry continuously delivers radiation while traversing an arc around the patient. This reduces treatment delivery times and improves OAR sparing (Zhang et al., 2010). The continuous rotation is approximated by a discrete set of equidistant angles called control points (CPs), each corresponding to a single aperture. This adds a layer of complexity by posing a larger-scale optimization problem.

Ignoring changes in a patient's anatomy throughout the course of a treatment may result in delivering radiation that largely differs from the planned treatment. However,

most of the literature focuses on deterministic models that do not generalize to clinical settings with myriad uncertainties. For example, when the tumor is positioned in the chest area, breathing motion can shift the position of the tumor while the treatment is being delivered (called intra-fraction motion). Similarly, when the tumor is in the abdomen, the shape of the rectum, bladder, or stomach can deform and change between treatments, causing the position of the tumor to differ from what was captured on the initial CT scan (referred to as inter-fraction motion). To mitigate uncertainty, treatment planners use margin-based approaches to target an area larger than the tumor. However, the increased tumor control comes at the expense of excessive radiation to the surrounding healthy tissue (Unkelbach et al., 2018). Consequently, stochastic programming (SP) and robust optimization (RO) approaches have been developed to balance tradeoffs between tumor coverage and OAR sparing. However, each approach heavily favored one of these goals. Consequently, probabilistic approaches were developed to serve as a middle ground. For example, chance-constrained programming (CCP) aims to give up a small amount of robustness in favor of a larger amount of effectiveness. However, these approaches assume that the underlying probability distribution of the movements is known, which is not realistic. If these uncertainties are not considered, then the radiation delivered may not sufficiently cover the tumor and may result in high levels of toxicity to the OARs. This could lead to unsafe outcomes for the patient's health in addition to wasted resources for the country and other patients.

The purpose of this study is to address inter-fraction motion of the tumor's position throughout the course of a radiation therapy treatment when the probability of each shift is not well known. To accomplish this, we integrate distributionally robust chance-constrained programming with sequential convex programming.

Chapter 2

Literature Review

In this section, we begin with a brief history of IMRT optimization (Section 2.1) followed by a review of the literature on optimization approaches to VMAT treatment planning (Section 2.2). Then, we discuss uncertainty mitigation in RT in Section 2.3. Last, in Section 2.4, we emphasize the research gap, outline the novel contributions of this thesis, and define its scope.

2.1 Optimization in Intensity-Modulated Radiation Therapy

In IMRT delivery, a multi-leaf collimator (MLC) is used to modulate the shape of the beam at each angle, in which a set of metal leaves cover parts of the beam to create a custom-shaped aperture that the radiation passes through, while blocking the radiation elsewhere. Therefore, in the treatment planning process, the collection of individual varying beamlet intensities must be translated into a set of apertures with uniform intensities. As outlined in Bortfeld (2006), there are two approaches to ensure IMRT treatment plans are deliverable. The multi-step approach involves solving a fluence map optimization (FMO) problem to obtain the required beamlet intensities at each angle, followed by solving a leaf sequencing (LS) problem to convert the fluence map of each angle into a set of deliverable

apertures (Romeijn et al., 2003; Kamath et al., 2003). Alternatively, the direct aperture optimization (DAO) approach is used to combine the FMO and LS steps and directly yield deliverable apertures (Shepard et al., 2002). Optimization models for DAO often use binary variables to include or exclude specific beamlets in apertures, which results in large-scale combinatorial optimization problems that are NP-hard (Baatar et al., 2005).

Clinical criteria can be considered in both FMO and DAO formulations. For example, clinicians can specify dose-volume histogram (DVH) criteria such as "the dose to the 10% least dosed voxels in the tumor should be at least 90% of the prescription". Romeijn et al. (2003) show that DVH criteria can be conservatively approximated by conditional value at risk (CVaR), which is a risk measure that can control the average above or below a specified percentile in a given distribution (Rockafellar and Uryasev, 2002; Uryasev and Rockafellar, 2001).

2.2 Optimization in Volumetric-Modulated Arc Therapy

Introduced by Yu (1995), VMAT has been gaining popularity due to its shorter delivery time and improved OAR sparing. However, it is a much larger problem which results in optimization models that are more challenging to solve (Otto, 2008). Similarly to IMRT, the two main categories for VMAT optimization models are multi-step and DAO (Unkelbach et al., 2015).

Multi-step optimization approaches that involve optimizing for fluence maps and converting them into MLC leaf segments are explored in Bzdusek et al. (2009), Cameron (2005), and Wang et al. (2008). Moreover, Craft et al. (2014), Unkelbach et al. (2015), Luan et al. (2008), and Cameron (2005) review sliding windows (SW) approaches for leaf sequencing. Further formulations are explored by Sun et al. (2013), who propose a nonlinear model, and Craft et al. (2014), who propose a multi-criteria optimization model. These approaches all introduce error when translating fluence maps from a control point into a deliverable aperture shape. Since the VMAT treatment modality involves many con-

trol points, this error is amplified and the effectiveness of the treatment, which is supposed to be its main advantage, degrades (Wang et al., 2022).

Unkelbach et al. (2015) critique the above approaches and propose a DAO-based formulation to be solved in one step. They argue that closely reproducing a large number of fluence maps is challenging to do efficiently and within the clinically acceptable time frame. In contrast, the DAO formulation reflects the exact DICOM specifications of a treatment plan that is communicated to the machine. Akartunalı et al. (2015) propose a mixed-integer programming (MIP) formulation of VMAT that simultaneously optimizes fluence maps and aperture shapes for a given set of CPs. Dursun et al. (2019a) formulate a MIP to minimize the total radiation dose and embed clinical requirements in the constraints.

The VMAT DAO formulations involve the use of one binary variable per beamlet, which results in a large-scale NP-hard combinatorial optimization problem (Baatar et al., 2005). As such, there is a body of research focused on developing more efficient solution schemes to achieve better solutions quicker at a more realistic scale. There are four main strategies to solve the DAO formulation: gradient-based with FMO-informed segment initialization, probabilistic search using heuristics, direct SW, and iterative aperture generation. Gradient-based optimization strategies (Bzdusek et al., 2009; MacFarlane et al., 2020) can only guarantee local optima; a near-(global)-optimal plan is not guaranteed. Alternatively, heuristics such as simulated annealing (Otto, 2008) and tabu search (Ulrich et al., 2007) can be used. However, due to the problem size, it is time consuming to fully explore the feasible region and obtain high quality solutions. Approaches that use SW, such as Papp and Unkelbach (2014), model continuous leaf trajectories and are hence more sensitive to discretization in the dose-deposition matrix.

Iterative aperture generation gradually refines aperture shapes until a stopping criterion is fulfilled. Peng et al. (2012), Dursun et al. (2019b), and Wang et al. (2022) develop column generation algorithms while Sun et al. (2013) and Dursun et al. (2019a) use Bender's decomposition. Akartunalı et al. (2015) provide several heuristic methods and a meta-heuristic approach. Moreover, they argue that, since treatment planning often happens up to three weeks before the treatment is carried out, it is more valuable to search for a solution that satisfies all requirements than to provide a quick solution that may not comply. Dursun et al. (2021) develop a sequential convex programming (SCP) scheme

which iteratively solves convex approximations of the original problem. Although this approach does not have global guarantees, they show that the dose distributions are near-optimal. To the best of our knowledge, this approach is the state of the art in terms of modeling and solution schemes, trade-offs between accuracy and speed, and clinical implementation.

2.3 Uncertainty Mitigation in Radiation Therapy

There are different sources of uncertainty in radiation therapy planning. For example, uncertainty in the fluence realized by the machine is addressed in Bertsimas et al. (2010). This formulation, however, shifts the question to whether the machine is able to accurately execute the instructions given. In contrast, this thesis focuses on whether the specified plan effectively treats the patient. The consensus in the literature is that mitigating uncertainty through optimization can yield better target coverage and/or better OAR sparing when compared to margin-based methods (Unkelbach et al., 2018).

To address potential uncertainties in practice, treatment planners often define several target volumes that capture different extents of uncertainty. The gross target volume (GTV) is the (deterministic) set of all cancerous voxels as identified on the CT scan. There may, however, be microscopic evidences of disease around the GTV which a CT scan fails to accurately capture. Clinicians consider this uncertainty by defining a larger area, denoted as the clinical target volume (CTV), that encompasses the GTV (Yahalom et al., 2015). Finally, to account for motion and setup uncertainties, a margin is placed around the CTV to create the planning target volume (PTV). In clinics, the PTV is typically used as the area to be irradiated in IMRT treatment planning.

Although the PTVs approach is often used to address uncertainties in clinical settings, Unkelbach et al. (2018) outline the limitations of this approach in accurately capturing uncertainties while avoiding over-conservatism. The PTV concept aims to deliver the prescribed dose to all the possible regions where the tumor may be. This does not differentiate between the regions where the tumor is less likely to be and other regions that the tumor surely occupies, hence, taking a protective approach to tumor control at the expense of

excessive radiation to the surrounding healthy tissue.

Although inter-fraction motion is complex to model, it can be approximated as a rigid shift of the tumor with respect to the isocenter (Unkelbach et al., 2018). This is typically interpreted as uncertainty in the dose-influence matrix and can be represented with a finite set of scenarios. It can then be combined with optimization approaches that consider uncertainty, such as SP and RO, as an alternative to the PTV concept for a better trade-off between tumor coverage and OAR sparing.

Optimization problems involving uncertainty in the parameters may be solved using SP (Dantzig, 1955). This involves defining a set of scenarios where each scenario either occurs with a measured probability or receives a weight. Then, SP optimizes for the expected value or weighted sum of a given function of the scenario set. This approach is effective because each scenario is considered to the extent that it is important. However, this scenario is not robust because less-likely scenarios are emphasised less; hence, should one occur, it is less likely to be protected.

Several clinical studies have been conducted to compare the performance of SP-based plans to the conventional PTV-based plans. Bohoslavsky et al. (2013) use a probabilistic objective function to create dose distributions similar to or better than PTV-based plans. They also argue that the value added from increasing planners' flexibility exceeds the additional computational burden. Witte et al. (2007) use SP-based plans to reduce the expected occurrence of severe healthy organ toxicity by half without compromising the expected tumor coverage. They also find that probabilistic treatment plans are robust against parameter variations but degrade rapidly when geometric uncertainties are underestimated. This motivates the use of an approach that is more robust.

Similarly to SP, RO (Bertsimas et al., 2011) considers a scenario set. In contrast, RO does not use probabilities or weights. Instead, RO optimizes for the worst-case scenario. This involves the decision maker developing a solution which is feasible for any realization in the uncertainty set and, hence, is both effective and robust. However, it may be overly conservative. For example, a decision maker may be willing to risk an unlikely scenario being realized in order to improve his objective function in more likely scenarios, which RO does not address.

Several clinical studies have been conducted to compare the performance of RO-based plans to the conventional PTV-based plans. Fredriksson et al. (2011) and Liu et al. (2012a) compare worst-case optimization with margin-based approaches using discrete scenario sets. They both find that RO results in robust tumor coverage and better healthy tissue sparing than margin-based approaches. Liu et al. (2012b) argue that all plans should be robustly optimized. Wada et al. (2022) conduct a clinical study to compare the robustness with setup error to the robustness of the conventional PTV-based plans in VMAT for prostate cancer. They used Raystation to simulate setup errors within the PTV margin, and found that the robust plan can reduce dose uncertainty due to setup errors. Fredriksson et al. (2015) let the size of the uncertainty set vary and optimize for the probability of satisfying clinical goals in the presence of uncertainty. Through a proof-of-concept considering one-dimensional uncertainty, they show that plans with an uncertainty set that is too large can decrease the plan quality; the best plans considered uncertainty without turning overly robust. This motivates the use of an approach that is less robust than RO but still effective.

In contrast to RO and SP, minimax stochastic programming (MSP) ranges between optimization for the expected and the worst case scenario (Fredriksson, 2012). For example, chance constraints (CCs) (Charnes and Cooper, 1959) represent the probabilistic inclusion of a constraint in an optimization model; rather than enforcing that a constraint holds in all scenarios, one can enforce that it holds with a high enough probability to be robust but not overly conservative. Therefore, chance constraints can be used to ensure an adequately high level of robustness without being overly conservative. An et al. (2017) impose CCs on dose-volume criteria using CVaR with the CTV as the target. They assume the inter-fraction motion can be modeled using a discrete scenario set. Then, they reformulate the chance constraint as a MIP. Their results show that using chance constraints with the CTV as the target can outperform PTV-based plans. To the best of our knowledge, using chance constraints to hedge against inter-fraction motion has yet to be explored in VMAT treatment planning.

A limitation of chance constraints is the need to estimate the probability of each scenario occurring. These estimates may be erroneous, which motivates a second level of robustness. Distributionally robust chance-constrained programs (DRCCPs) can hedge

against uncertainty in the underlying probability distribution. Recent works (Xie, 2021; Ho-Nguyen et al., 2023; Küçükyavuz and Jiang, 2022) assume the planner has a large number of samples that can be used to estimate a distribution. In practice, this is seldom possible. Yanıkoğlu (2019) considers a chance constrained program with a discrete scenario set and an estimated but uncertain nominal probability distribution. They hedge against uncertainty in the probability distribution by defining an ambiguity set centered around the nominal distribution and derive a robust reformulation of the DRCCP. Then, they apply the methods to a small-scale knapsack problem. To the best of our knowledge, no work in radiation therapy has addressed uncertainty in the discrete distribution of a chance constraint.

2.4 Gap, Contributions, and Scope

We emphasise the following gaps in the literature:

1. No research in VMAT has explored chance-constrained programming to model rigid shifts of the CTV and compared its performance to PTV-based approachs.
2. No research in RT has considered uncertainty in the underlying probability distribution of a chance constraint when accounting for geometric uncertainty.

This thesis aims to fill these gaps with the following specific contributions:

1. We extend an existing chance-constrained framework to the case where the underlying probability distribution of the rigid shifts is also uncertain by integrating the robust reformulation of an ambiguous constraint in FMO and DAO formulations (Gap 2).
2. We modify a sequential convex programming algorithm to include robust and distributionally robust optimization and show that the quality of the convex approximation does not degrade with robustness (Gaps 1 and 2).

3. As a proof-of-concept, we apply these models to open-sourced lung cancer datasets with one-dimensional uncertainty in the position of the CTV and compare the performance of nominal, robust, and distributionally robust models against PTV-based plans (Gaps 1 and 2).

Figures 2.1 and 2.2 visualize the comparison between multi-step, direct, and VMAT optimization approaches. Figure 2.1 depicts the process for multi-step and direct approaches when applied to standard IMRT and VMAT. It outlines our focus on the optimization steps. For the scope of this thesis, we assume beam angle selection and leaf sequencing protocols (where relevant) are given. Consequently, we focus on the optimization step related to fluence. Note, however, that the FMO step in VMAT is out of scope. In Chapter 5, we show that the flexibility of FMO in a small-scale IMRT case dampens the effects of RO and distributionally robust optimization (DRO). As such, considering FMO in VMAT, which includes many CPs and significantly increases flexibility, is not worth exploring at this stage. Figure 2.2 depicts example outputs of each approach. Most notably, the multi-step IMRT approach takes a selected set of beam angles, solves for the fluence of each beamlet individually, then converts it to a deliverable aperture using LS. In contrast, the direct IMRT pipeline obtains the apertures in one step. Finally, the direct VMAT pipeline uses a discretized arc instead of a preselected set of CPs and solves for the apertures directly.

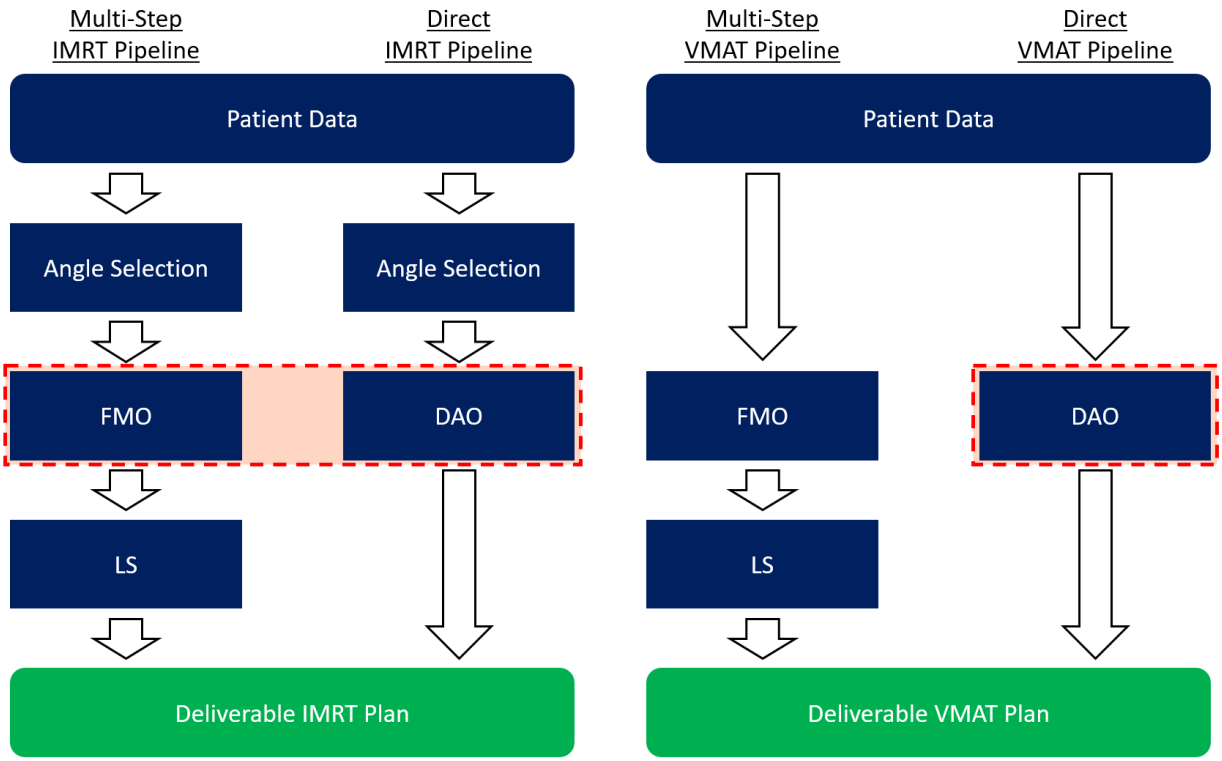


Figure 2.1: The main approaches to IMRT planning with red indicating the scope of this thesis.

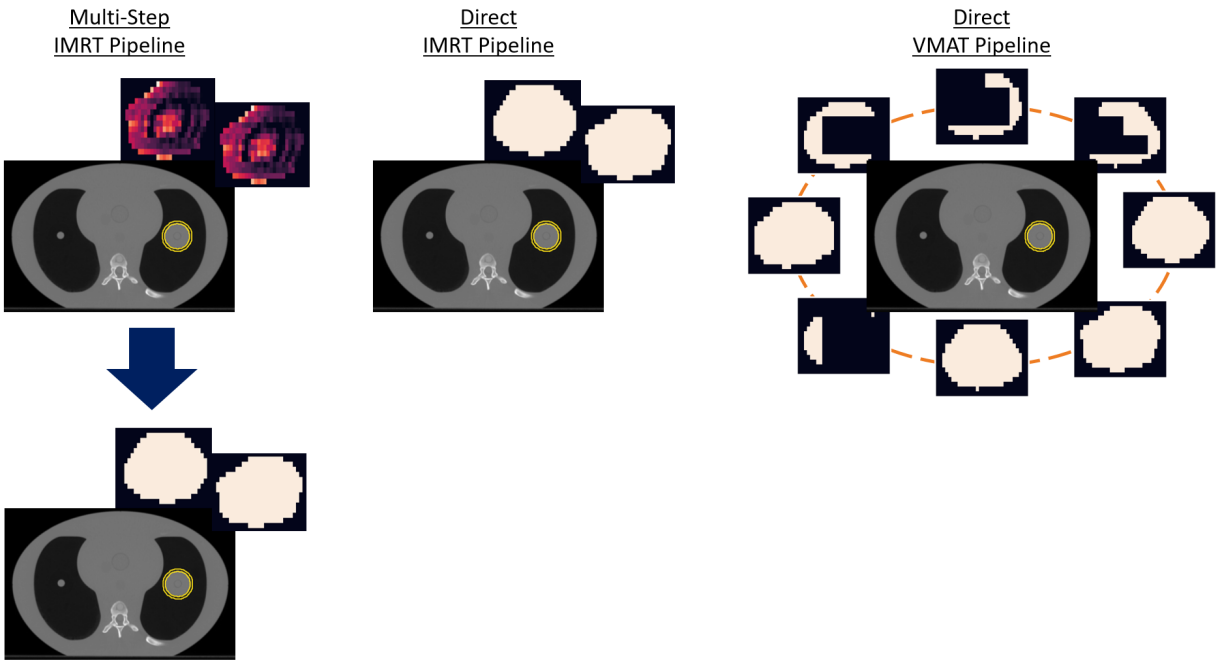


Figure 2.2: A visualization of the optimization outputs in each step of each approach.

Chapter 3

Problem Formulation

This chapter begins by presenting a general deterministic constrained optimization model and deriving the robust and distributionally robust counterparts in Section 3.1. Then, in Section 3.2, we describe how they relate to radiation therapy. Next, we apply the deterministic, robust, and distributionally robust models to FMO (Section 3.3), and DAO (Section 3.4) for VMAT.

3.1 Deterministic, Robust, and Distributionally Robust Chance Constraint Derivations

In this section, we begin by introducing a general deterministic constrained optimization model in 3.1.1. Then, in 3.1.2, we present a chance-constrained setting and derive the robust counterpart. Finally, we present a case with an ambiguous chance constraint and derive the distributionally robust counterpart in 3.1.3.

3.1.1 Deterministic Model

Consider the general deterministic constrained optimization model in (3.1). Here, $\mathbf{x} \in \mathbb{R}^n$ is the decision variable and we want to minimize any function $f : \mathbb{R}^n \to \mathbb{R}$. Constraint (3.1b)

is a particular constraint of interest where $g : \mathbb{R}^n \to \mathbb{R}$ is a linear function in $\mathbf{x}$. Set $\mathcal{X}$ represents any additional feasible set that may be specified.

$$\begin{aligned} \min \quad & f(\mathbf{x}) && \text{(3.1a)} \\ \text{s.t.} \quad & g(\mathbf{x}) \leq 0 && \text{(3.1b)} \\ & \mathbf{x} \in \mathcal{X} && \text{(3.1c)} \end{aligned}$$

This model enforces constraint (3.1b) to always be met and assumes its parameters are always certain. Strict enforcement may not always be desirable, as there may be some uncertainty in the parameters which could lead to leeway in violating the constraint in some scenarios to improve the objective function. As such, in the following section, we present a chance-constrained formulation and its robust counterpart.

3.1.2 Robust Chance-Constrained Model

Consider the general stochastic chance-constrained optimization model in (3.2).

$$\begin{aligned} \min \quad & f(\mathbf{x}) && \text{(3.2a)} \\ \text{s.t.} \quad & \mathbb{P}\Big(g(\mathbf{x}; \tilde{\xi}) \leq 0\Big) \geq 1 - \epsilon && \text{(3.2b)} \\ & \mathbf{x} \in \mathcal{X} && \text{(3.2c)} \end{aligned}$$

Here, we consider the case where constraint (3.1b) has uncertain parameters and the planner can tolerate it being violated with probability ϵ. Suppose $\mathbb{P}$ is a known discrete probability distribution and let $\mathcal{S}$ be a set of scenarios such that $\mathbb{P}(s : s \in \mathcal{S}) = p^s$. Let $\tilde{\xi}$ be a random vector that influences the outcome of the uncertain parameters of g, where $\tilde{\xi} \in \{\xi^s\}_{s \in \mathcal{S}}$ Since there are multiple scenarios, the value of $g(\mathbf{x})$ may change. As such, we define $g(\mathbf{x}; \xi^s)$ to be the value of $g(\mathbf{x})$ given the random vector $\tilde{\xi}$ when scenario $s \in \mathcal{S}$ is realized. Constraint (3.2b) is a chance constraint with known probability distribution $\mathbb{P}$ to allow a specified tolerance of violations, ϵ.

Let $\mathbf{z} \in \mathbb{R}^{|\mathcal{S}|}$ be a binary vector such that $z^s = 1$ if scenario $s \in \mathcal{S}$ is protected by the chance constraint, and 0 otherwise. Now, model (3.2) admits the big-M formulation in (3.3).

$$\begin{aligned}
\min \quad & f(\mathbf{x}) && \text{(3.3a)}\\
\text{s.t.} \quad & g(\mathbf{x}; \xi^s) \leq M(1 - z^s) \quad \forall s \in \mathcal{S} && \text{(3.3b)}\\
& \sum_{s \in \mathcal{S}} p^s z^s \geq 1 - \epsilon && \text{(3.3c)}\\
& \mathbf{x} \in \mathcal{X}, \quad \mathbf{z} \in \{0,1\}^{|\mathcal{S}|} && \text{(3.3d)}
\end{aligned}$$

This model enforces constraint (3.2b) to be met with a specified probability given a known probability distribution and can be solved efficiently for reasonably sized uncertainty sets (An et al., 2017). An accurate probability distribution may not always be obtainable, as there may be some uncertainty in $\mathbb{P}$. As such, in the following section, we present an ambiguous chance-constrained formulation and its distributionally robust counterpart.

3.1.3 Distributionally Robust Chance-Constrained Model

In this section, we will consider uncertainty in the underlying probability distribution of the chance constraint and derive a distributionally robust counterpart following Yanıkoğlu (2019). In particular, we consider the case where a planner can tolerate constraint (3.1b) being violated with probability ϵ but the underlying probability distribution is not fully known. To model this, we assume that the nominal distribution, $\mathbf{p}$, can be reasonably estimated and define the following polyhedral ambiguity set:

$$\mathcal{P} := \left\{ \tilde{\mathbf{p}} = \mathbf{p} + \hat{\mathbf{p}} : \underline{\mathbf{p}} \leq \hat{\mathbf{p}} \leq \overline{\mathbf{p}}, \mathbf{e}'\hat{\mathbf{p}} = 0 \right\}.$$

Let $\mathbb{Q}$ be a probability measure induced by any probability distribution that $\tilde{\mathbf{p}}$ lies in the ambiguity set, $\mathcal{P}$. $\mathcal{P}$ is the set of possible probability distributions that the chance constraint may follow. The uncertain probability, $\tilde{\mathbf{p}}$, is depicted by the nominal probability

vector $\mathbf{p}$ plus some error term $\hat{\mathbf{p}}$. The error term may lie between lower and upper bounds $\underline{\mathbf{p}}$ and $\overline{\mathbf{p}}$, respectively. Regardless of the error term, $\tilde{\mathbf{p}}$ must remain a valid probability distribution. As such, the all atoms in the error term must sum to 0. With this definition, we can now present the ambiguous chance-constrained program in (3.4).

$$\begin{aligned} \min \quad & f(\mathbf{x}) && \text{(3.4a)} \\ \text{s.t.} \quad & \min_{\mathbb{Q}\in\mathcal{P}} \mathbb{Q}\Big(g(\mathbf{x};\tilde{\xi}) \leq 0\Big) \geq 1-\epsilon && \text{(3.4b)} \\ & \mathbf{x} \in \mathcal{X} && \text{(3.4c)} \end{aligned}$$

The difference between models (3.2) and (3.4) is that (3.4) considers the worst case probability distribution within the ambiguity set. This results in robustness to the distribution $\mathbb{Q}$ as well as the realizations of $\tilde{\xi}$. Through big-M notation, model (3.4) can be reformulated as the bilevel program in (3.5).

$$\begin{aligned} \min \quad & f(\mathbf{x}) && \text{(3.5a)} \\ \text{s.t.} \quad & g(\mathbf{x};\xi^s) \leq M(1-z^s) \quad \forall s \in \mathcal{S} && \text{(3.5b)} \\ & \min_{\tilde{\mathbf{p}}\in\mathcal{P}} \sum_{s\in\mathcal{S}} \tilde{p}^s z^s \geq 1-\epsilon && \text{(3.5c)} \\ & \mathbf{x} \in \mathcal{X} && \text{(3.5d)} \\ & \mathbf{z} \in \{0,1\}^{|\mathcal{S}|} && \text{(3.5e)} \end{aligned}$$

Consider constraint (3.5c). This is an inner optimization problem which can be reformulated using the definition from the ambiguity set:

$$\min_{\hat{\mathbf{p}}\in\mathcal{P}} \Big\{ \sum_{s\in\mathcal{S}} (p^s + \hat{p}^s) z^s : \underline{\mathbf{p}} \leq \hat{\mathbf{p}} \leq \overline{\mathbf{p}}, \mathbf{e}'\mathbf{p} = 0 \Big\} \geq 1-\epsilon.$$

Note that in this inner optimization problem, all atoms of $\mathbf{z}$ are constant, and so are the atoms of $\mathbf{p}$. As such, the above can be rearranged as follows:

$$\min_{\hat{\mathbf{p}}\in\mathcal{P}}\Big\{\sum_{s\in\mathcal{S}}\hat{p}^s z^s : \underline{\mathbf{p}} \le \hat{\mathbf{p}} \le \overline{\mathbf{p}}, \mathbf{e}'\mathbf{p} = 0\Big\} \ge 1 - \epsilon - \mathbf{z}'\mathbf{p}.$$

Focusing on the left-hand-side of the above expression, we obtain a linear programming formulation of the inner problem in (3.6).

$$\min_{\hat{\mathbf{p}}} \quad \hat{\mathbf{p}}'\mathbf{z} \tag{3.6a}$$
$$\text{s.t.} \quad \mathbf{e}'\hat{\mathbf{p}} = 0 \tag{3.6b}$$
$$\hat{\mathbf{p}} \le \overline{\mathbf{p}} \tag{3.6c}$$
$$\hat{\mathbf{p}} \ge \underline{\mathbf{p}} \tag{3.6d}$$

Let $\lambda \in \mathbb{R}$, $\boldsymbol{\theta} \in \mathbb{R}_+^{|\mathcal{S}|}$, and $\boldsymbol{\eta} \in \mathbb{R}_-^{|\mathcal{S}|}$ be the dual variables associated with constraints (3.6b), (3.6c), (3.6d), respectively. It follows that the dual of (3.6) is:

$$\max_{\lambda,\boldsymbol{\eta}\le\mathbf{0},\boldsymbol{\theta}\ge\mathbf{0}} \quad \underline{\mathbf{p}}'\boldsymbol{\eta} + \overline{\mathbf{p}}'\boldsymbol{\theta}, \tag{3.7a}$$
$$\text{s.t.} \quad \lambda + \eta^s + \theta^s = z^s, \quad \forall s \in \mathcal{S}. \tag{3.7b}$$

Then, observe that by strong duality the following holds:

$$\max_{\lambda,\boldsymbol{\eta}\le\mathbf{0},\boldsymbol{\theta}\ge\mathbf{0}}\Big\{\underline{\mathbf{p}}'\boldsymbol{\eta}+\overline{\mathbf{p}}'\boldsymbol{\theta} : \lambda\mathbf{e}+\boldsymbol{\eta}+\boldsymbol{\theta} = \mathbf{z}\Big\} = \min_{\hat{\mathbf{p}}\in\mathcal{P}}\Big\{\sum_{s\in\mathcal{S}}\hat{p}^s z^s : \mathbf{e}'\hat{\mathbf{p}} = 0, \underline{\mathbf{p}} \le \hat{\mathbf{p}} \le \overline{\mathbf{p}}\Big\} \ge 1-\epsilon-\mathbf{z}'\mathbf{p}.$$

Thus (3.5c) becomes:

$$\mathbf{z}'\mathbf{p} + \max_{\lambda,\boldsymbol{\eta}\le\mathbf{0},\boldsymbol{\theta}\ge\mathbf{0}}\Big\{\underline{\mathbf{p}}'\boldsymbol{\eta} + \overline{\mathbf{p}}'\boldsymbol{\theta} : \lambda\mathbf{e} + \boldsymbol{\eta} + \boldsymbol{\theta} = \mathbf{z}\Big\} \ge 1 - \epsilon.$$

We can then remove the maximization term because any feasible solution to the maximization problem will be a lower bound to the original minimization problem by weak duality. It follows that model (3.5) admits the equivalent reformulation depicted in (3.8).

$$
\begin{aligned}
\min \quad & f(\mathbf{x}) && && \text{(3.8a)}\\
\text{s.t.} \quad & g(\mathbf{x};\xi^s) \leq M(1-z^s) && \forall s \in \mathcal{S} && \text{(3.8b)}\\
& \mathbf{z}'\mathbf{p} + \underline{\mathbf{p}}'\boldsymbol{\eta} + \overline{\mathbf{p}}\boldsymbol{\theta} \geq 1-\epsilon && && \text{(3.8c)}\\
& \lambda\mathbf{e} + \boldsymbol{\eta} + \boldsymbol{\theta} = \mathbf{z} && && \text{(3.8d)}\\
& \mathbf{x} \in \mathcal{X} && && \text{(3.8e)}\\
& \mathbf{z} \in \{0,1\}^{|\mathcal{S}|} && && \text{(3.8f)}\\
& \lambda \in \mathbb{R} && && \text{(3.8g)}\\
& \boldsymbol{\eta} \leq \mathbf{0}, \quad \boldsymbol{\theta} \geq \mathbf{0} && && \text{(3.8h)}
\end{aligned}
$$

This model enforces constraint (3.4b) to be met with a specified probability if the worst-case probability distribution within the ambiguity set is realized. This now accounts for both uncertainty in g and uncertainty in the underlying distribution. Going forward, all constraints that model robustness remain the same. To apply these models to the radiation therapy setting, all that must be specified is f, g, and $\mathcal{X}$.

3.2 Radiotherapy Problem Setting

The focus of this section is to relate the deterministic, robust, and distributionally robust models to a radiation therapy setting and provide an intuitive primer before presenting the mathematical formulations.

In our setting, the objective of a RT model is to deliver an appropriate dose of radiation to the patient. To achieve this, we consider an objective function similar to An et al. (2017) which penalizes the linear deviation from the tumor's prescribed dose as well as the total radiation delivered to the OARs to create conformal plans. Although the clinical literature often uses quadratic penalties, we consider a linear one for computational tractability. Nevertheless, switching to any other objective, such as a quadratic penalty, would not affect the proposed methodology. Note that the objective function focuses on the nominal

scenario, which is that the tumor is exactly where the images depict it to be. This is not always the case, so we account for the other scenarios using constraints.

Let Υ be the set of OARs. For each OAR $\upsilon \in \Upsilon$, let $\mathcal{H}_\upsilon$ be the set of voxels it contains. Then, let $\mathcal{T}$ be the set of voxels in the target. Let $\mathbf{d}_i'$ be a row in the dose deposition matrix associated with voxel $i \in \mathcal{T} \cup \{\mathcal{H}_\upsilon\}_{\upsilon\in\Upsilon}$ and let $\bar{d}$ be the prescribed dose to the tumor. Finally, our decision variable is the intensity of each beamlet, which we denote in vector $\mathbf{x}$. In the RT problem setting, the objective function follows equation (3.9a).

$$f(\mathbf{x}) = \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} |\mathbf{d}_i'\mathbf{x} - \bar{d}| + \sum_{\upsilon\in\Upsilon}\frac{1}{|\mathcal{H}_\upsilon|}\sum_{i\in\mathcal{H}_\upsilon}\mathbf{d}_i'\mathbf{x} \tag{3.9a}$$

The first term in the function $f(x)$ is the linear deviation between the dose delivered to the tumor and the dose prescribed. We penalize both over- and underdoses to achieve conformal plans. The second term represents the linear deviation between the healthy tissue's dose received and prescribed. This term is equivalent to minimizing the mean dose to OARs because the prescription is 0 Gy.

Given that the objective function has absolute values, it is non-linear but can easily be linearized. Let ψ_i^{over} be an auxiliary variable for the overdose on voxel $i \in \mathcal{T}$ and let ψ_i^{under} be an auxiliary variable for the underdose on voxel $i \in \mathcal{T}$. The objective function can be linearized as follows:

$$\text{Min} \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\psi_i^{\text{over}} + \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\psi_i^{\text{under}} + \sum_{\upsilon\in\Upsilon}\frac{1}{|\mathcal{H}_\upsilon|}\sum_{i\in\mathcal{H}_\upsilon}\mathbf{d}_i'\mathbf{x} \tag{3.10a}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{T}}\mathbf{d}_i'\mathbf{x} \geq \bar{d} - \psi_i^{\text{under}} \qquad \forall i \in \mathcal{T} \tag{3.10b}$$

$$\sum_{i\in\mathcal{T}}\mathbf{d}_i'\mathbf{x} \leq \bar{d} + \psi_i^{\text{over}} \qquad \forall i \in \mathcal{T} \tag{3.10c}$$

$$\psi_i^{\text{over}} \geq 0, \quad \psi_i^{\text{under}} \geq 0 \qquad \forall i \in \mathcal{T} \tag{3.10d}$$

We use this linearization in our implementations. However, to preserve readability and intuition, we continue our formulations with the original compact formulation in (3.9a).

As our focus is on robust tumor coverage, the main constraint of the RT models in this problem setting will relate to tumor coverage. Specifically, we use a CVaR-based coverage constraint to approximate DVH criteria, similarly to Romeijn et al. (2003) and An et al. (2017). This enforces that the mean dose to the lower tail of the target's dose distribution be above a specified threshold. For demonstration purposes, we consider the case where there is only one target; however, DVH criteria on multiple targets can easily be integrated in the models.

Let ρ be an auxiliary variable used in the CVaR formulation. Further, let β be a specified percentile and let $\underline{d}$ be a lower bound for the dose. Finally, let $(\cdot)^+ = \max\{0, \cdot\}$. It follows that we can replace $g(x) \leq 0$ with the following constraint.

$$\rho - \frac{1}{(1-\beta)|\mathcal{T}|} \sum_{i \in \mathcal{T}} (\rho - \mathbf{d}_i' \mathbf{x})^+ \geq \underline{d} \tag{3.11a}$$

This constraint specifies that the mean dose to the $100 * (1-\beta)\%$ least-dose voxels of the CTV must be at least $\underline{d}$. It is non-linear but can easily be linearized. Let ω_i be an auxiliary variable to represent the underdose to voxel i of the tumor relative to the value at risk. Then, the CVaR constraint can be replaced with (3.12a)-(3.12c). The same changes can be made in the robust and distributionally robust models. As such, to preserve readability and intuition, we will continue discussion with constraint (3.11a).

$$\rho - \frac{1}{(1-\beta)|\mathcal{T}|} \sum_{i \in \mathcal{T}} \omega_i \geq \underline{d} \tag{3.12a}$$

$$\omega_i \geq \rho - \mathbf{d}_i' \mathbf{x} \qquad \forall i \in \mathcal{T} \tag{3.12b}$$

$$\omega_i \geq 0 \qquad \forall i \in \mathcal{T} \tag{3.12c}$$

In a clinical setting, the objective functions and constraints are more complex. For example, constraints on the mean and maximum dose can be enforced on multiple OARs. Further, in the case where constraints clash and lead to infeasibilities, multi-criteria optimization can be used to incorporate soft constraints. Multi-criteria optimization can also be used to incentivise "nicer" aperture shapes that are easier for the machine to deliver. The scope of this thesis is to extend RT planning methods by incorporating geometric and

distributional uncertainty. As such, we focus on the basic setting with a simple objective function and single constraint on target coverage. Planners with a specific definition of plan quality who wish to build custom models can then incorporate all levels of uncertainty using the approaches depicted in this thesis. As such, for now we may take $\mathcal{X}$ to be the empty set.

In the following sections, we will consider the following levels of uncertainty. First, we consider the nominal setting where the position of the tumor is exactly where the initial CT image depicts it. Then, we consider the case where the position of the tumor is uncertain but can be found in one of finitely-many scenarios, each with an estimated probability. The difference between the deterministic model targeting the CTV, the deterministic model targeting the PTV, and the robust model with geometric uncertainty is illustrated in Figure 3.1. Second, we consider the case where the estimated probability distribution is inaccurate, but the true probability lies within a specified margin of error. This type of uncertainty is illustrated in Figure 3.2.

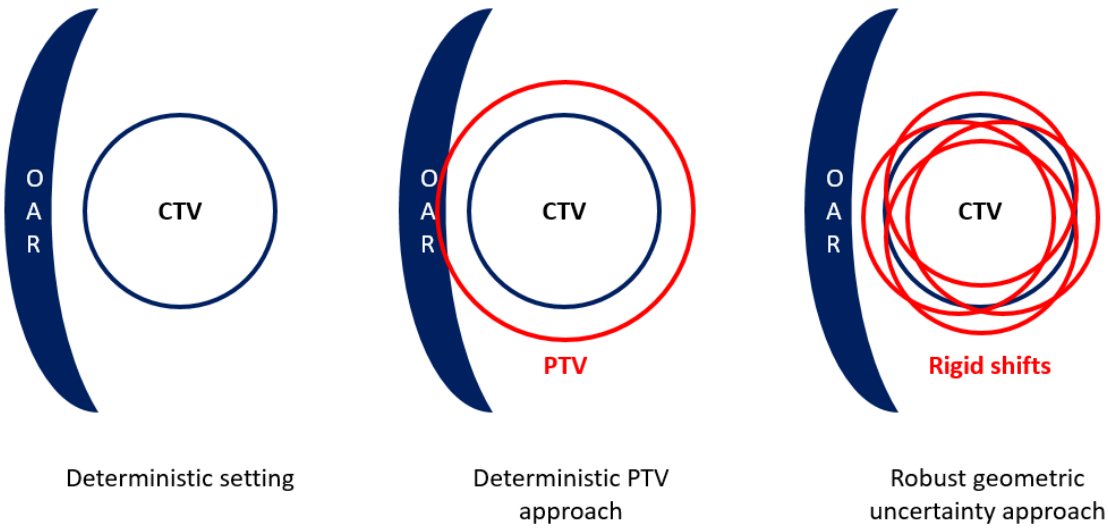


Figure 3.1: A comparison of the clinical setting and various treatment planning approaches.

We will also consider three levels of RT models. First, we define the base FMO model and apply each level of uncertainty to it. Second, we extend the FMO model to create

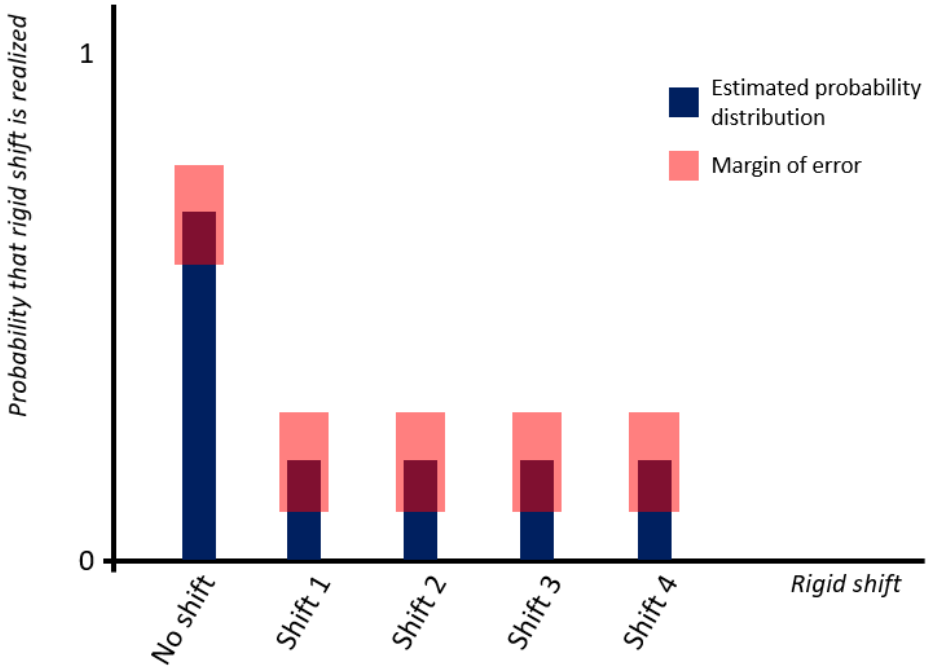


Figure 3.2: An example of the DRO set-up with five possible realizations in the CTV position.

deliverable apertures in a DAO setting. Third, we present the SCP algorithm which enables the solution of the DAO model in large-scale settings and is used for VMAT planning.

Notations will be defined in-line where relevant. Tables detailing all sets, parameters, and variables are provided in Appendix A. In general, we use capital calligraphic letters to denote sets, capital bold letters to denote matrices, lower-case bold letters to denote vectors, where all vectors are column vectors, and lower-case letters to denote one-dimensional variables or parameters, unless otherwise specified.

3.3 Fluence Map optimization (FMO) Models

The FMO model is the simplest RT planning model that takes a predetermined set of beams and individually solves for each beamlet's intensity. The output is a fluence map which can be passed to down-stream LS algorithms to create deliverable apertures. This section first presents the deterministic FMO model in Section 3.3.1. Then, we incorporate the robust (Section 3.3.2) and distributionally robust optimization (Section 3.3.3) approaches discussed above.

3.3.1 Deterministic FMO Model

Assembling our discussion from Section 3.2, model (3.13) depicts the deterministic FMO model.

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.13a}$$

$$\text{s.t.} \quad \rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-\mathbf{d}_i'\mathbf{x})^+\geq \underline{d} \tag{3.13b}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.13c}$$

$$\rho\in\mathbb{R} \tag{3.13d}$$

The objective function (3.13a) minimizes the deviation between the dose received and the dose prescribed. The deviations of each structure are then averaged so that larger structures do not receive a greater emphasis. Constraint (3.13b) is the CVaR-based coverage constraint in the nominal setting. Finally, constraint (3.13d) provides domain constraints. Recall, however, that this model assumes the CTV is located where the imaging indicates its position. In practice, this may not hold during the duration of the treatment plan. If the tumor shifts without an appropriate adjustment by the planner, the treatment quality may degrade. As such, the following subsection handles this using robust optimization.

3.3.2 Robust Chance-Constrained FMO Model

As a less conservative alternative to the PTV approach, we consider the chance-constrained methodology of An et al. (2017) to enforce the CVaR constraint (3.13b) to be met with $100 * (1 - \epsilon)\%$ probability, instead of a single deterministic case. Note that improving plan robustness will inevitably sacrifice plan quality in the nominal scenario (An et al., 2017). As such, we also optimize for the quality in the nominal scenario through the objective function. Let $\tilde{\mathbf{d}}_i'$ be a random vector that represents an uncertain row in the dose-deposition matrix, and let $\mathbf{d}_i'$ denote its value under no shift. The chance-constrained model is shown in formulation (3.14).

$$\min \quad \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} |\mathbf{d}_i' \mathbf{x} - \bar{d}| + \sum_{v \in \Upsilon} \frac{1}{|\mathcal{H}_v|} \sum_{i \in \mathcal{H}_v} \mathbf{d}_i' \mathbf{x} \tag{3.14a}$$

$$\text{s.t.} \quad \mathbb{P}\left(\rho - \frac{1}{(1-\beta)|\mathcal{T}|} \sum_{i \in \mathcal{T}} \left(\rho - \tilde{\mathbf{d}}_i' \mathbf{x} \right)^+ \geq \underline{d} \right) \geq 1 - \epsilon \tag{3.14b}$$

$$\mathbf{x} \geq \mathbf{0} \tag{3.14c}$$

$$\rho \in \mathbb{R} \tag{3.14d}$$

Constraint (3.14b) can be tractably reformulated using the structure of the considered uncertainty set. Let $\mathcal{S}$ represent the uncertainty set consisting of possible scenarios of the geometric position of the CTV, which includes the nominal scenario. Let $\tilde{\mathbf{d}}_i \in \{\mathbf{d}_i^s\}_{s \in \mathcal{S}}$, where $\mathbf{d}_i^s$ is the i^{th} row of the dose deposition matrix associated with scenario $s \in \mathcal{S}$. Let $\mathbf{p}$ represent a known nominal probability distribution, where p^s is the probability of scenario $s \in \mathcal{S}$ being realized. Finally, let $\mathbf{z}$ be a binary vector such that $z^s = 1$ if scenario $s \in \mathcal{S}$ is protected by the chance constraint and $z^s = 0$ otherwise. The tractable reformulation is shown in model (3.15).

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.15a}$$

$$\text{s.t.} \quad \rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-(\mathbf{d}_i^s)'\mathbf{x})^+\geq \underline{d}-M(1-z^s) \quad \forall s\in\mathcal{S} \tag{3.15b}$$

$$\mathbf{p}'\mathbf{z}\geq 1-\epsilon \tag{3.15c}$$

$$\mathbf{x}\geq \mathbf{0} \tag{3.15d}$$

$$\rho\in\mathbb{R} \tag{3.15e}$$

$$\mathbf{z}\in\{0,1\}^{|\mathcal{S}|} \tag{3.15f}$$

Constraint (3.15b) ensures that, for each scenario $s \in \mathcal{S}$, a CVaR surrogate constraint is satisfied if $z^s = 1$ and ignored otherwise. Constraint (3.15c) ensures that the decision variable $\mathbf{x}$ satisfies the CVaR surrogate constraint in a set of scenarios whose total nominal probability under $\mathbf{p}$ is at least $1 - \epsilon$. Together, these constraints form a scenario-based representation of the original chance constraint in (3.14b). However, this approach assumes a known probability distribution which is often not available in practice. For any new patient, although the probability of a shift may be estimated, the estimate may be erroneous. As such, we define the distributionally robust model below.

3.3.3 Distributionally Robust Chance-Constrained FMO Model

Model (3.16) depicts the distributionally robust FMO model with an ambiguous chance constraint. In this setting, we continue to assume the position of the tumor is in one of finitely many scenarios. Each scenario has an associated dose-deposition matrix and nominal probability of occurrence. However, this nominal probability may be erroneous. As such, we model for the worst-case distribution.

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.16a}$$

$$\text{s.t.} \quad \min_{\mathbb{Q}\in\mathcal{P}}\mathbb{Q}\left(\rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-\tilde{\mathbf{d}}_i'\mathbf{x})^+\geq\underline{d}\right)\geq 1-\epsilon \tag{3.16b}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.16c}$$

$$\rho\in\mathbb{R} \tag{3.16d}$$

The difference between models (3.14) and (3.16) is constraint (3.16b), which considers the worst case probability distribution in addition to multiple realizations of the target position. A tractable reformulation of this model is shown below.

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.17a}$$

$$\text{s.t.} \quad \rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}\left(\rho-(\mathbf{d}_i^s)'\mathbf{x}\right)^+\geq\underline{d}-M(1-z^s)\quad\forall s\in\mathcal{S} \tag{3.17b}$$

$$\mathbf{z}'\mathbf{p}+\overline{\mathbf{p}}'\boldsymbol{\eta}+\underline{\mathbf{p}}'\boldsymbol{\theta}\geq 1-\epsilon \tag{3.17c}$$

$$\lambda\mathbf{e}+\boldsymbol{\eta}+\boldsymbol{\theta}=\mathbf{z} \tag{3.17d}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.17e}$$

$$\rho\in\mathbb{R} \tag{3.17f}$$

$$\mathbf{z}\in\{0,1\}^{|\mathcal{S}|} \tag{3.17g}$$

$$\lambda\in\mathbb{R} \tag{3.17h}$$

$$\boldsymbol{\eta}\leq\mathbf{0},\quad\boldsymbol{\theta}\geq\mathbf{0} \tag{3.17i}$$

Model (3.13) can be solved efficiently using commercial solvers and, for reasonably-sized uncertainty sets, so can models (3.15) and (3.17). Recall, however, that FMOs do not yield deliverable apertures. Their outputs must be passed to a subsequent LS algorithm to obtain deliverable machine instructions, which is out of the scope of this thesis. However, this process can be combined using the DAO approach, which is outlined in the next section.

3.4 Direct Aperture optimization (DAO) Models

We use DAO to combine the FMO and LS steps and directly output deliverable apertures. First, we briefly discuss an amendment to FMO models that leads to a DAO formulation in 3.4.1. Then, in 3.4.2, we present the deterministic mixed-integer formulation that can output globally optimal solutions. After, we augment this model with the robust (3.4.3) and distributionally robust counterparts (3.4.4).

3.4.1 Model Preliminaries

Before we formulate the VMAT models, we need to establish the relationship between beamlet intensities and the positions of the MLC leaves. To do so, we employ the mapping function of Dursun et al. (2021) that relates the positions of the left and right MLC leaves with the intensity of each beamlet in their row. Using the variable vectors $\mathbf{l}$ and $\mathbf{r}$ to denote left and right MLC leaf positions, we can calculate the intensity vector of the beamlets $\mathbf{x}$, using the mapping function Φ as a function of the left and right leaf positions and the aperture intensity. Explicitly, for beamlet $j \in \mathcal{J}$ belonging to control point $b \in \mathcal{B}$, located in row t and column c_j, the mapping is:

$$x_j = \Phi(a_b, l_{bt}, r_{bt}) = a_b \times \min\Big\{(\min\{c_j - l_{bt}, 1\})^+, (\min\{r_{bt} + 1 - c_j, 1\})^+, r_{bt} - l_{bt}\Big\}.$$

Figure 3.3 provides a visual illustration of this mapping, in which the left half denotes a sample row of a small aperture where the MLC leaves are in green and the squares represent the beamlets. The aperture intensity is at 10 MU, and the left and right leaves are at positions 1.25 and 3, respectively. The right half of Figure 3.3 shows the associated beamlet intensity after the mapping, in which an intensity of 0, 7.5, and 10 indicates that a beamlet is not covered, partially covered, and fully covered by an MLC leaf, respectively.

This mapping can be incorporated in an optimization model to output deliverable apertures, shown in model (3.18). The objective function and constraint (3.18b) are the same as in the FMO models. Then, constraint (3.18c) relates the beamlet intensities to leaf positions. Constraint (3.18d) specifies that the left leaf cannot clash with the right leaf. Finally, constraint (3.18e) bounds the aperture intensity for each beam.

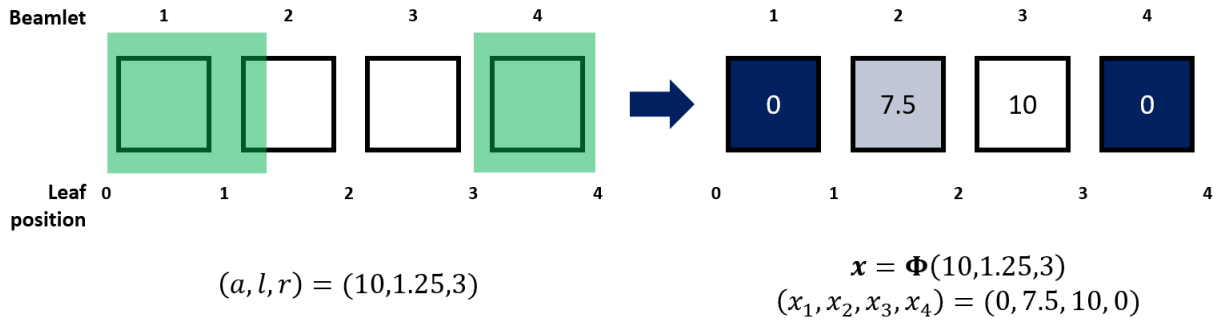


Figure 3.3: Visual schematic of mapping aperture intensity and leaf positions to beamlet intensities.

$$\min \quad \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} |\mathbf{d}_i' \mathbf{x} - \bar{d}| + \sum_{\upsilon \in \Upsilon} \frac{1}{|\mathcal{H}_\upsilon|} \sum_{i \in \mathcal{H}_\upsilon} \mathbf{d}_i' \mathbf{x} \tag{3.18a}$$

$$\text{s.t.} \quad \rho - \frac{1}{(1-\beta)|\mathcal{T}|} \sum_{i \in \mathcal{T}} (\mathbf{d}_i' \mathbf{x} - \rho)^+ \geq \underline{d} \tag{3.18b}$$

$$\mathbf{x} = \Phi(\mathbf{a}, \mathbf{l}, \mathbf{r}) \tag{3.18c}$$

$$\mathbf{0} \leq \mathbf{l} \leq \mathbf{r} \leq \mathbf{N} \tag{3.18d}$$

$$\mathbf{0} \leq \mathbf{a} \leq \mathbf{U} \tag{3.18e}$$

$$\rho \in \mathbb{R} \tag{3.18f}$$

Constraint (3.18c) is non-convex and requires nuanced treatment; Dursun et al. (2021) provide a mixed-integer non-linear programming (MILP) formulation which obtains globally optimal solutions when each beamlet can either be fully open or fully closed. This approach has not been previously integrated with robust optimization to consider geometric or distributional uncertainty in the position of the target. In the subsequent subsections, we will demonstrate this integration.

3.4.2 Deterministic DAO Model

An example of deliverable aperture is shown in figure 3.4. Here, the grey rectangles represent the MLC leaves. Beamlets that are covered by the leaves are off, and are depicted in blue. Beamlets that are not covered by the leaves are on, and are depicted in orange.

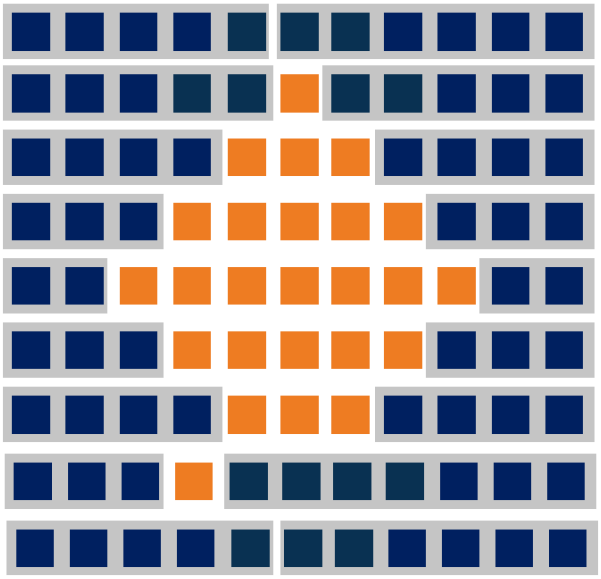


Figure 3.4: An example of a feasible aperture. Orange beamlets are on, blue beamlets are off, and the grey rectangles represent the MLC leaves. Note: rows 7 and 8 present an effect called *interdigitation*. If interdigitation occurs excessively, it may damage the MLC leaves. This effect can be discouraged through feasibility constraints.

In this section, we employ the approach of Dursun et al. (2021) to provide a MILP reformulation of the original non-convex beamlet intensity calculation. In particular, because the binary relaxation is convex, these problems can be solved to global optimality for small-scale settings. Let binary variable $w_j = 1$ if beamlet $j \in \mathcal{J}$ is on, and 0 if it is off. We introduce the set $\mathcal{K}_b$, which represents all rows in the aperture associated with control point $b \in \mathcal{B}$. Following this, we can replace the beamlet intensity mapping constraint (3.18c) with constraints (3.19a)-(3.19g).

$$r_{bk} - c_j w_j \geq 0 \quad \forall b \in \mathcal{B}, k \in \mathcal{K}_b, j \in \mathcal{J} \tag{3.19a}$$

$$(N + 1 - c_j) w_j + l_{bk} \leq N \quad \forall b \in \mathcal{B}, k \in \mathcal{K}_b, j \in \mathcal{J} \tag{3.19b}$$

$$\sum_{j \in \mathcal{J}} w_j = r_{bk} - l_{bk} \quad \forall b \in \mathcal{B}, k \in \mathcal{K}_b \tag{3.19c}$$

$$0 \leq x_j \leq U w_j \quad \forall j \in \mathcal{J} \tag{3.19d}$$

$$a_b - U(1 - w_j) \leq x_j \leq a_b \quad \forall b \in \mathcal{B}, j \in \mathcal{J} \tag{3.19e}$$

$$\mathbf{w} \in \{0, 1\}^{|\mathcal{J}|} \tag{3.19f}$$

$$l_{bk}, r_{bk} \in \mathbb{Z} \quad \forall b \in \mathcal{B}, k \in \mathcal{K}_b, j \in \mathcal{J} \tag{3.19g}$$

Constraints (3.19a)-(3.19c) link the position of the MLC leaves with whether the beamlets are on or off. Specifically, constraint (3.19a) enforces beamlet w_i to be 0 if it is covered by the left leaf. Similarly, constraint (3.19b) enforces beamlet w_i to be 0 if it is covered by the right leaf. Then, constraint (3.19c) enforces beamlet w_i to be 1 if it is not covered by either the left or right leaf while also ensuring that the left and right leaves do not overlap. Next, constraints (3.19d) and (3.19e) relate the on/off beamlets to the radiation delivered. Specifically, they ensure that the dose emitted by a covered beamlet is 0, whereas the dose emitted by an uncovered beamlet is equal to the aperture intensity respectively. This series of links enables us to integrate the CVaR-based tumor coverage constraint as presented in the previous section. Since these additional constraints are linear and our objective function can be linearized using auxiliary variables, our previous VMAT models can all be reformulated as MILP, and the detailed reformulations are provided in the appendix. It follows that model (3.20) presents the full deterministic DAO model.

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.20a}$$

$$\text{s.t.} \quad \rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-\mathbf{d}_i'\mathbf{x})^+\geq\underline{d} \tag{3.20b}$$

$$(3.19a)-(3.19g) \tag{3.20c}$$

$$\mathbf{0}\leq\mathbf{l}\leq\mathbf{r}\leq\mathbf{N} \tag{3.20d}$$

$$\mathbf{0}\leq\mathbf{a}\leq\mathbf{U} \tag{3.20e}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.20f}$$

$$\rho\in\mathbb{R} \tag{3.20g}$$

Contrary to FMO, this model can produce deliverable apertures. However, this model only considers the case where the tumor is exactly where expected, which may not hold in practice. The following subsection addresses by combining robust optimization with the aperture deliverability constraints.

3.4.3 Robust Chance-Constrained DAO Model

Although model (3.20) creates deliverable apertures, it does not account for rigid shifts of the target. Model (3.21) addresses this by adding a chance-constraint to the DAO model. The settings and interpretation mimic the FMO model; the only difference is the inclusion of the deliverability constraints which results in feasible apertures.

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.21a}$$

$$\text{s.t.} \quad \mathbb{P}\left(\rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}\left(\rho-\tilde{\mathbf{d}}_i'\mathbf{x}\right)^+\geq\underline{d}\right)\geq 1-\epsilon \tag{3.21b}$$

$$(3.19a)-(3.19e) \tag{3.21c}$$

$$\mathbf{0}\leq\mathbf{l}\leq\mathbf{r}\leq\mathbf{N} \tag{3.21d}$$

$$\mathbf{0}\leq\mathbf{a}\leq\mathbf{U} \tag{3.21e}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.21f}$$

$$\rho\in\mathbb{R} \tag{3.21g}$$

$$\mathbf{w}\in\{0,1\}^{|\mathcal{J}|} \tag{3.21h}$$

$$l_{bk},r_{bk}\in\mathbb{Z} \qquad \forall b\in\mathcal{B},k\in\mathcal{K}_b \tag{3.21i}$$

The tractable reformulation of this model is shown in (3.22).

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.22a}$$

$$\text{s.t.} \quad \rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-(\mathbf{d}_i^s)'\mathbf{x})^+\geq\underline{d}-M(1-z^s) \qquad \forall s\in\mathcal{S} \tag{3.22b}$$

$$\mathbf{p}'\mathbf{z}\geq 1-\epsilon \tag{3.22c}$$

$$(3.19a)-(3.19e) \tag{3.22d}$$

$$\mathbf{0}\leq\mathbf{l}\leq\mathbf{r}\leq\mathbf{N} \tag{3.22e}$$

$$\mathbf{0}\leq\mathbf{a}\leq\mathbf{U} \tag{3.22f}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.22g}$$

$$\rho\in\mathbb{R} \tag{3.22h}$$

$$\mathbf{w}\in\{0,1\}^{|\mathcal{J}|} \tag{3.22i}$$

$$l_{bk},r_{bk}\in\mathbb{Z} \qquad \forall b\in\mathcal{B},k\in\mathcal{K}_b \tag{3.22j}$$

$$\mathbf{z}\in\{0,1\}^{|\mathcal{S}|} \tag{3.22k}$$

Once again, recall that CCP assumes that the probability distribution of shifts is known. This is not realistic – for any new patient, the probability of a shift can be estimated, but the estimate may be erroneous. As such, we define the distributionally robust model below.

3.4.4 Distributionally Robust Chance-Constrained DAO Model

To ensure robustness to the estimated probability distribution while producing deliverable apertures, model (3.23) uses the ambiguous chance constraint following the previous derivations.

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.23a}$$

$$\text{s.t.} \quad \min_{\mathbb{Q}\in\mathcal{P}}\mathbb{Q}\left(\rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-\tilde{\mathbf{d}}_i'\mathbf{x})^+\geq\underline{d}\right)\geq 1-\epsilon \tag{3.23b}$$

$$(3.19a)-(3.19e) \tag{3.23c}$$

$$\mathbf{0}\leq\mathbf{l}\leq\mathbf{r}\leq\mathbf{N} \tag{3.23d}$$

$$\mathbf{0}\leq\mathbf{a}\leq\mathbf{U} \tag{3.23e}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.23f}$$

$$\rho\in\mathbb{R} \tag{3.23g}$$

$$\mathbf{w}\in\{0,1\}^{|\mathcal{J}|} \tag{3.23h}$$

$$l_{bk},r_{bk}\in\mathbb{Z} \qquad \forall b\in\mathcal{B},k\in\mathcal{K}_b,j\in\mathcal{J} \tag{3.23i}$$

The inner minimization problem yields a tractable reformulation presented in model (3.24).

$$\min \quad \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}|\mathbf{d}_i'\mathbf{x}-\bar{d}|+\sum_{v\in\Upsilon}\frac{1}{|\mathcal{H}_v|}\sum_{i\in\mathcal{H}_v}\mathbf{d}_i'\mathbf{x} \tag{3.24a}$$

$$\text{s.t.} \quad \rho-\frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho-(\mathbf{d}_i^s)'\mathbf{x})^+\geq \underline{d}-M(1-z^s) \qquad \forall s\in\mathcal{S} \tag{3.24b}$$

$$\mathbf{z}'\mathbf{p}+\overline{\mathbf{p}}'\boldsymbol{\eta}+\underline{\mathbf{p}}'\boldsymbol{\theta}\geq 1-\epsilon \tag{3.24c}$$

$$\lambda\mathbf{e}+\boldsymbol{\eta}+\boldsymbol{\theta}=\mathbf{z} \tag{3.24d}$$

(3.19a) – (3.19e) (3.24e)

$$\mathbf{0}\leq\mathbf{l}\leq\mathbf{r}\leq\mathbf{N} \tag{3.24f}$$

$$\mathbf{0}\leq\mathbf{a}\leq\mathbf{U} \tag{3.24g}$$

$$\mathbf{x}\geq\mathbf{0} \tag{3.24h}$$

$$\rho\in\mathbb{R} \tag{3.24i}$$

$$\mathbf{w}\in\{0,1\}^{|\mathcal{J}|} \tag{3.24j}$$

$$l_{bk},r_{bk}\in\mathbb{Z} \qquad \forall b\in\mathcal{B},k\in\mathcal{K}_b,j\in\mathcal{J} \tag{3.24k}$$

$$\mathbf{z}\in\{0,1\}^{|\mathcal{S}|} \tag{3.24l}$$

$$\lambda\in\mathbb{R},\quad \boldsymbol{\eta}\leq\mathbf{0},\quad \boldsymbol{\theta}\geq\mathbf{0} \tag{3.24m}$$

3.4.5 Model Size

Let the reformulated nominal, robust, and distributional robust MILP models be denoted as N-MILP, R-MILP, and DR-MILP, respectively. Table 3.1 notes the size of these problems in terms of the number of continuous, binary, and integer variables. The large size and combinatorial structure of the problem is mostly from the number of beamlets and the number of rows per aperture, which motivates the use of an efficient solution methodology. In what follows, we adapt and extend the approximation algorithm of Dursun et al. (2021) to solve our nominal, robust and distributionally robust VMAT models.

	Variables			
	Continuous	Binary	Integer	Constraints
N-MILP	$\lvert\mathcal{J}\rvert + 3\lvert\mathcal{T}\rvert + 1$	$\lvert\mathcal{J}\rvert$	$\sum_{b\in\mathcal{B}} \lvert\mathcal{K}_b\rvert$	$1 + 3\lvert\mathcal{T}\rvert + \sum_{b\in\mathcal{B}} \lvert\mathcal{K}_b\rvert(2\lvert\mathcal{J}\rvert + 3) + 2\lvert\mathcal{J}\rvert + \lvert\mathcal{B}\rvert(2\lvert\mathcal{J}\rvert + 1)$
R-MILP	$\lvert\mathcal{J}\rvert + 2\lvert\mathcal{T}\rvert + \lvert\mathcal{T}\rvert\lvert\mathcal{S}\rvert + 1$	$\lvert\mathcal{J}\rvert + \lvert\mathcal{S}\rvert$	$\sum_{b\in\mathcal{B}} \lvert\mathcal{K}_b\rvert$	$2 + 2\lvert\mathcal{T}\rvert + \lvert\mathcal{T}\rvert\lvert\mathcal{S}\rvert + \sum_{b\in\mathcal{B}} \lvert\mathcal{K}_b\rvert(2\lvert\mathcal{J}\rvert + 3) + 2\lvert\mathcal{J}\rvert + \lvert\mathcal{B}\rvert(2\lvert\mathcal{J}\rvert + 1)$
DR-MILP	$\lvert\mathcal{J}\rvert + 2\lvert\mathcal{T}\rvert + \lvert\mathcal{T}\rvert\lvert\mathcal{S}\rvert + 2\lvert\mathcal{S}\rvert + 1$	$\lvert\mathcal{J}\rvert + \lvert\mathcal{S}\rvert$	$\sum_{b\in\mathcal{B}} \lvert\mathcal{K}_b\rvert$	$2 + 2\lvert\mathcal{T}\rvert + \lvert\mathcal{T}\rvert\lvert\mathcal{S}\rvert + \lvert\mathcal{S}\rvert + \sum_{b\in\mathcal{B}} \lvert\mathcal{K}_b\rvert(2\lvert\mathcal{J}\rvert + 3) + 2\lvert\mathcal{J}\rvert + 2\lvert\mathcal{B}\rvert(\lvert\mathcal{J}\rvert + 1)$

Table 3.1: Problem size of each MILP instance denoting the number of individual continuous, binary, and integer variables as well as the number of individual constraints.

Chapter 4

Solution Methodology

Recall that the global mixed-integer linear program (MILP) can only be solved in a reasonable amount of time for small-scale cases. As such, Dursun et al. (2021) develop a sequential convex programming (SCP) algorithm to obtain high quality solutions fast at a clinical scale. This section is dedicated to extending the SCP algorithm by incorporating robust and distributionally robust optimization. In Section 4.1, we begin by presenting their SCP algorithm adapted to our problem setting. Then, in Section 4.2, we will adapt the optimization problems within the algorithm to consider geometric and distributional uncertainty. After, we adapt the feasibility check step to consider geometric and distributional uncertainty in Section 4.3. We conclude in Section 4.4 with a summarized implementation of the robust SCP algorithms.

4.1 Deterministic Sequential Convex Programming Algorithm

The speed improvements of the SCP algorithm is due to the convex approximation. Instead of defining a binary variable for each beamlet, it splits an initialized aperture into three categories: a large interior beamlet which remains on between iterations, a large exterior beamlet which remains off between iterations, and a trust region (denoted as the left/right

boundary beamlets) which limits how far the leaves can move between iterations. Then, at each iteration, the trust region is refined until a solution close to the global optimum is obtained. This combination of beamlets discards the need for one binary variable per beamlet, which speeds up performance and enables more granular apertures.

Let $\hat{\mathbf{D}}$ represent the adjusted dose-deposition matrix, which accounts for the interior, exterior, and boundary beamlets at unit intensity, and where $\hat{\mathbf{d}}_i'$ is its i^{th} row. Also, let $\hat{\mathbf{x}}$ be a vector of beamlet intensities, where $\hat{x}_b^{\text{int}}$ is the intensity of the interior beamlet of control point $b \in \mathcal{B}$ and $\hat{x}_{bku}$ is the intensity of the boundary beamlet of control point $b \in \mathcal{B}$ at row $k \in \mathcal{K}_b$ on the left side ($u = 1$) or right side ($u = 2$). Model (4.1) depicts the convex approximation problem used in Dursun et al. (2021) and Dursun et al. (2023) adapted to our problem setting. A visual representation of this formulation is shown in Figure 4.1.

$$\min \quad \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} |\hat{\mathbf{d}}_i' \hat{\mathbf{x}} - \bar{d}| + \sum_{\upsilon \in \Upsilon} \frac{1}{|\mathcal{H}_\upsilon|} \sum_{i \in \mathcal{H}_\upsilon} \hat{\mathbf{d}}_i' \hat{\mathbf{x}} \tag{4.1a}$$

$$\text{s.t.} \quad \rho - \frac{1}{(1-\beta)|\mathcal{T}|} \sum_{i \in \mathcal{T}} (\rho - \hat{\mathbf{d}}_i \hat{\mathbf{x}})^+ \geq \underline{d} \tag{4.1b}$$

$$\hat{x}_b^{\text{int}} = a_b \qquad \forall b \in \mathcal{B} \tag{4.1c}$$

$$0 \leq \hat{x}_{bku} \leq a_b \qquad \forall b \in \mathcal{B}, k \in \mathcal{K}_b, u = 1, 2 \tag{4.1d}$$

$$0 \leq \mathbf{a} \leq U \tag{4.1e}$$

$$\rho \in \mathbb{R} \tag{4.1f}$$

Note that one of the differences between the convex approximation and the global MILP model is the use of $\hat{\mathbf{D}}$ and $\hat{\mathbf{x}}$ instead of $\mathbf{D}$ and $\mathbf{x}$. This represents an adjustment to account for the exterior, boundary, and interior beamlets. As before, we use a linear penalty on the dose deviation as our objective and use constraint (4.1b) to enforce tumor coverage in the nominal setting. Constraint (4.1c) enforces the interior beamlet intensity be equal to the intensity of the aperture. Constraint (4.1d) ensures that the intensity of the boundary beamlet is between 0 and that of the interior beamlet. This intensity is later used to approximate the position of the leaves. Finally, constraint (4.1e) presents lower and upper limits on the intensity of each aperture.

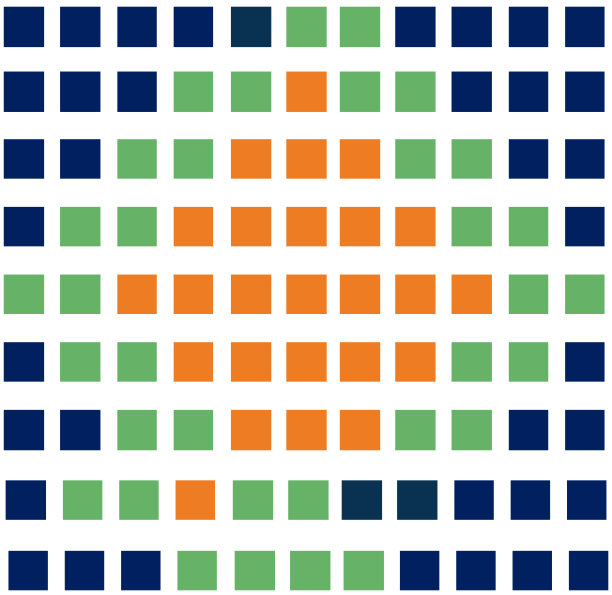


Figure 4.1: An example aperture using the convex approximation formulation. Here, orange beamlets are combined to represents the interior beamlet, which remains on throughout subsequent iterations. Navy beamlets can be combined to represent the exterior beamlet, which remains off throughout subsequent iterations. Green beamlets can be combined to represent the boundary beamlets.

The algorithm for the iterative refinement of the boundary beamlet (referred to as step-size) is described in Algorithm (1). We will refer to this entire algorithm as N-SCP. Line 2 solves the intermediate convex approximation problem to determine the fluence of the boundary beamlets. This optimization problem will need to be changed when considering robustness. Then, line 3 converts the boundary beamlets to machine parameters according to Dursun et al. (2023). For example, for the left leaf of the first row in CP b, the optimal leaf position can be calculated using:

$$l_{b1} = 2\left(1 - \frac{x_{b11}}{a_b}\right).$$

This step is deterministic and does not require adjustment for RO or DRO.

Given that the intermediate problem is an approximation, the (actual) machine parameters may violate some of its (approximately satisfied) constraints. Thus, line 4 checks the feasibility of the constraints under these machine parameters by calculating the value of the constraints under the reconstructed solution. If the feasibility check is left deterministic, other realizations in the position of the target will be ignored. Further, if the feasibility check is made robust, uncertainty in the probability distribution will be ignored. As such, this step will require modification.

If the parameters are infeasible, then a correction problem is solved in line 6. The correction problem fixes the positions of the left and right leaves, then solves for each aperture's intensity. Therefore, the correction problem will not change the shape of the aperture; it only changes the intensity of the radiation beam. The actual correction problem is similar to the intermediate convex approximation problem. The only change necessary is to replace $\hat{\mathbf{D}}\hat{\mathbf{x}}$ with the interior dose-deposition and matrix and interior beamlet intensity only. The calculation and modifications of the dose-deposition matrices is handled using built-in functions in PortPy, as such we take these parameters as given. This problem would, however, require modifications to encompass RO and DRO.

Once a feasible solution is obtained, it is compared to the solution obtained from the previous iteration. If the new solution is better, then the trust region is enlarged. This happens so that the algorithm avoids converging to a local optimum which is not global. Intuitively, enlarging the trust region means expanding the number of green beamlets from

Figure 4.1. This allows the leaf of a row in an aperture to travel further between iterations. If the resulting solution is not better than the previous solution, then the trust region is shrunken. This step aims to improve the accuracy and deliver higher quality solutions. We use the same step-size schedule as Dursun et al. (2021) for all cases.

Algorithm 1 Pseudocode for the Nominal SCP Algorithm (N-SCP)

1: **while** improvement $> \varepsilon$ **do**
2: find the auxiliary solution by optimizing the deterministic intermediate convex approximation problem
3: reconstruct the actual solution $(\mathbf{a}, \mathbf{l}, \mathbf{r})$
4: check whether $(\mathbf{a}, \mathbf{l}, \mathbf{r})$ is feasible to the deterministic intermediate convex approximation problem
5: **if** $(\mathbf{a}, \mathbf{l}, \mathbf{r})$ is infeasible **then**
6: fix $(\mathbf{l}, \mathbf{r})$ and re-optimize $\mathbf{a}$ in the deterministic actual correction problem to generate a feasible solution
7: **if** $(\mathbf{a}, \mathbf{l}, \mathbf{r})$ is better than $(\mathbf{a}^n, \mathbf{l}^n, \mathbf{r}^n)$ **then**
8: update the solution $(\mathbf{a}^{n+1}, \mathbf{l}^{n+1}, \mathbf{r}^{n+1}) = (\mathbf{a}, \mathbf{l}, \mathbf{r})$
9: enlarge the trust-region (step-size = step-size + 1)
10: **else**
11: shrink the trust-region (step-size = round(step-size / 2))
12: $n \leftarrow n + 1$

In summary, lines 2, 4, and 6 of N-SCP (Algorithm (1)) do not extend to RO and DRO, which has yet to be addressed in the literature. Steps 2 and 6 are optimization problems. If the intermediate problem remains deterministic, then any geometric and distributional uncertainty is ignored. If the intermediate problem is robust but the correction problem is deterministic, then adequate convergence will not be achieved. Finally, if the feasibility check is left deterministic, then the algorithm will once again ignore uncertainty. In what follows, we extend N-SCP to RO and DRO.

4.2 Incorporating Robustness in the Intermediate and Correction Problems

The intermediate convex approximation problem and the actual correction problem used in Algorithm 1 do not account for rigid shifts in the tumor. Let $\hat{\mathbf{d}}_i^s$ be the i^{th} row of the adjusted dose deposition matrix associated with scenario $s \in \mathcal{S}$. To incorporate geometric uncertainty, we update the deterministic intermediate convex approximation problem in model (4.1) by replacing constraint (4.1b) with constraints (4.2a)-(4.2d) shown below. The same changes can be made to the actual correction problem, but, as before, one must replace $\hat{\mathbf{D}}\hat{\mathbf{x}}$ with the interior dose-deposition and matrix and interior beamlet intensity only. This gives rise to the robust intermediate convex approximation and robust actual correct problems.

$$\rho - \frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}(\rho - (\hat{\mathbf{d}}_i^s)'\hat{\mathbf{x}})^+ \geq \underline{d} - M(1-z^s) \quad \forall s \in \mathcal{S} \tag{4.2a}$$

$$\mathbf{p}'\mathbf{z} \geq 1-\epsilon \tag{4.2b}$$

$$\rho \in \mathbb{R} \tag{4.2c}$$

$$\mathbf{z} \in \{0,1\}^{|S|} \tag{4.2d}$$

These robust problems assume an estimated known probability distribution. As before, we can protect against estimation errors using distributionally robust optimization. To incorporate distributional uncertainty, we update the deterministic intermediate convex approximation problem in model (4.1) by replacing constraint (4.1b) with constraints (4.3a)-(4.3g) shown below. The same changes can be made to the actual correction problem, but, as before, one must replace $\hat{\mathbf{D}}\hat{\mathbf{x}}$ with the interior dose-deposition and matrix and interior beamlet intensity only. This gives rise to the distributionally robust intermediate convex approximation and distributionally robust actual correction problems.

$$\rho - \frac{1}{(1-\beta)|\mathcal{T}|}\sum_{i\in\mathcal{T}}\left(\rho - (\hat{\mathbf{d}}_i^s)'\hat{\mathbf{x}}\right)^+ \geq \underline{d} - M(1-z^s) \quad \forall s \in \mathcal{S} \tag{4.3a}$$

$$\mathbf{z}'\mathbf{p} + \overline{\mathbf{p}}'\boldsymbol{\eta} + \underline{\mathbf{p}}'\boldsymbol{\theta} \geq 1 - \epsilon \tag{4.3b}$$

$$\lambda\mathbf{e} + \boldsymbol{\eta} + \boldsymbol{\theta} = \mathbf{z} \tag{4.3c}$$

$$\rho \in \mathbb{R} \tag{4.3d}$$

$$\mathbf{z} \in \{0,1\}^{|\mathcal{S}|} \tag{4.3e}$$

$$\lambda \in \mathbb{R} \tag{4.3f}$$

$$\boldsymbol{\eta} \leq \mathbf{0}, \quad \boldsymbol{\theta} \geq \mathbf{0} \tag{4.3g}$$

4.3 Incorporating Robustness in the Feasibility Check

Consider steps 2 to 4 of Algorithm 1. For demonstration, suppose we solve the robust intermediate convex approximation problem in step 2 instead of its deterministic analogue. Then, in step 4, we would need to change the feasibility check to account for the chance constraint. If we do not, the deterministic feasibility check will only ensure that the no-shift scenario receives adequate target coverage. This may result in the chance constraint not meeting the specified tolerance. As such, we propose a modified robust feasibility check in Algorithm 2.

Algorithm 2 Pseudocode for Robust Feasibility Check

1: calculate the specified lower-tail CVaR received by the target in all scenarios $s \in S$
2: define binary vector $\mathbf{y}$ such that $y^s = 1$ if the nominal CVaR constraint is satisfied in scenario $s \in \mathcal{S}$, and 0 if it is violated
3: calculate $\Delta = \mathbf{p}'\mathbf{y}$
4: **if** $\Delta \geq 1 - \epsilon$ **then**
5: the solution is feasible
6: **else**
7: the solution is infeasible

Once again, consider steps 2 to 4 of Algorithm 1. Suppose, for demonstration, that we solve the distributionally robust intermediate convex approximation problem instead of its deterministic analogue. Then, as before, we would need to modify the feasibility check in step 4 to account for the ambiguous chance constraint. If we keep the feasibility check deterministic, then we encounter the same problem where too few rigid shift scenarios may be protected against. However, switching to the robust feasibility check is also not enough; it is possible that a solution which is chance-constraint feasible in the nominal probability distribution is infeasible in the worst-case probability distribution. Note that in the robust setting, the realized probability distribution is given as an input. However, in the distributionally robust setting, the realized probability distribution is taken as the worst-case, which is implicitly solved for in the intermediate problem. As such, we also propose a modified distributionally robust feasibility check in Algorithm 3.

Algorithm 3 Pseudocode for Distributionally Robust Feasibility Check

1: calculate the specified lower-tail CVaR received by the tumor in all $s \in S$
2: define binary vector $\mathbf{y}$ such that $y^s = 1$ if the nominal CVaR constraint is satisfied in scenario $s \in \mathcal{S}$, and 0 if it is violated
3: calculate $\mathbf{p}^{WC}$: the worst-case probability distribution, by solving the inner optimization problem from model (3.6) given $\mathbf{z}$ from the intermediate problem
4: calculate $\Delta = (\mathbf{p}^{WC})'\mathbf{y}$
5: **if** $\Delta \geq 1 - \epsilon$ **then**
6: the solution is feasible
7: **else**
8: the solution is infeasible

4.4 Implementation of the Robust Sequential Convex Programming Algorithm

Upon modifying the optimization problems and feasibility checks in N-SCP, we may define the robust and distributionally robust sequential convex programming algorithms, denoted

R-SCP and DR-SCP, respectively in algorithm (4). The use of SCP enables solutions that result in feasible apertures and appropriate doses at a realistic clinical scale. The use of R-SCP accounts for geometric uncertainty in the position of the target. To the best of our knowledge, this thesis is the first to apply a DRCCP formulation to VMAT treatment planning. The DR-SCP models account for distributional uncertainty in the probabilities of the shifts, in addition to geometric uncertainty in the position of the target. To the best of our knowledge, this thesis is the first to apply DRO with discrete probability distributions to model inter-fraction motion in radiation therapy treatment planning.

Algorithm 4 Pseudocode for R-SCP (or DR-SCP)

while improvement $> \varepsilon$ **do**
    find the auxiliary solution by optimizing the robust/distributionally robust intermediate convex approximation in model (4.2) (or (4.3))
    reconstruct the actual solution $(\mathbf{a}, \mathbf{l}, \mathbf{r})$
    check whether $(\mathbf{a}, \mathbf{l}, \mathbf{r})$ is feasible to the robust (or distributionally robust) intermediate convex approximation using Algorithm (2) (or (3))
    if $(\mathbf{a}, \mathbf{l}, \mathbf{r})$ is infeasible **then**
        fix $(\mathbf{l}, \mathbf{r})$ and re-optimize $\mathbf{a}$ in the robust/distributionally actual correction problem to generate a feasible solution
    if $(\mathbf{a}, \mathbf{l}, \mathbf{r})$ is better than $(\mathbf{a}^n, \mathbf{l}^n, \mathbf{r}^n)$ **then**
        update the solution $(\mathbf{a}^{n+1}, \mathbf{l}^{n+1}, \mathbf{r}^{n+1}) = (\mathbf{a}, \mathbf{l}, \mathbf{r})$
        enlarge the trust-region (step-size = step-size + 1)
    else
        shrink the trust-region (step-size = round(step-size / 2))
    $n \leftarrow n + 1$

Chapter 5

Results

In this chapter, we present the impact of addressing geometric uncertainty as well as geometric and distributional uncertainty. Section 5.1 inspects the impact on FMO models. We note that FMO is not independently used in practice as such the results are not expected to represent a realistic clinical setting. Section 5.2 inspects the impact on DAO models. Although DAO is used in practice, it is computationally expensive. As such, we focus on a small-scale experiment; these results also do not represent a realistic clinical setting. Finally, Section 5.3 inspects the impact on SCP models. This is solved at a realistic scale and the results can be interpreted clinically. However, we only propose these methods to be used by clinicians as a warm start; each patient is unique and would benefit from professional expertise in the treatment planning process.

5.1 Fluence Map Optimization

This section will investigate the impacts of robustness when paired with the FMO model. First, we define our experimental setup in Section 5.1.1. Then, we compare the effects of different levels of robustness using the FMO models. We do this by comparing DVH plots in Section 5.1.2 and through a sensitivity analysis on the size of the ambiguity set in Section 5.1.3.

We remind the reader that FMO models control for the intensity of each beamlet independently, which makes plans seem very conformal. This means that FMO solutions must be paired with a downstream LS algorithm to yield a deliverable plan. Given that LS is out of scope, we present the dose distributions obtained by the FMO for completeness.

5.1.1 Experimental Design

To compare the impacts of robustness, we compare each model in Table 5.1 on a lung patient that was treated using step-and-shoot IMRT from the dataset made available by Memorial Sloan Kettering Cancer Center. We take the six expert-chosen beam angles as control points, use the most granular beamlet size of 2.5 mm by 2.5 mm, and use every voxel available. We define the CTV by placing a 5 mm margin around the GTV to account for cancerous cells that could not be detected by the scan. We develop our models in Python using the PortPy package (Jhanwar et al., 2023) and solve them locally using Gurobi 11.0.1 on a computer with an Intel i5-2320 CPU and 16 GB of RAM.

Model	**Description**
N-FMO	Nominal FMO model with the CTV as the target
R-FMO	Robust FMO model with a chance constraint accounting for rigid shifts of the CTV in the nominal probability distribution
DR-FMO (5%)	Distributionally robust FMO model with an ambiguous chance constraint accounting for rigid shifts of the CTV in the worst-case probability distribution, where the ambiguity set allows for a 5% margin of error around each atom
PTV-Benchmark	Nominal FMO model with the PTV as the target

Table 5.1: FMO models tested in clinical-scale experiment.

In all models, the prescribed target dose is 60 Gy and we enforce the CVaR constraint to deliver an average of at least 59.4 Gy to the 5% least-dosed voxels. Due to pre-calculated dose-deposition matrices in the open-sourced PortPy data, we are only able to accurately

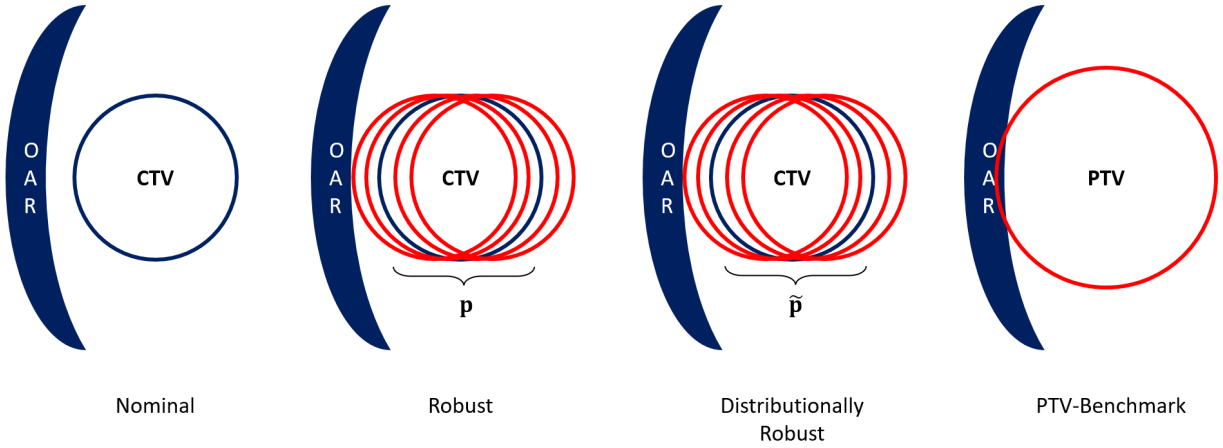


Figure 5.1: A visual depiction of each model.

simulate one-dimensional shifts near the nominal scenario. As such, similarly to Fredriksson et al. (2015), we proceed with this one-dimensional uncertainty set as a proof of concept. In the robust and distributionally robust models, our uncertainty set includes the nominal scenario and simulated shifts to the left and right by 3 mm and 6 mm. We also use out-of-sample simulated left and right shifts by 9 mm to test robustness on unseen cases. For simplicity, we adopt the approach in An et al. (2017) where the nominal scenario occurs with 60% probability and the remaining probability is distributed uniformly across the remainder of the uncertainty set. For the ambiguity set in the distributionally robust model, we define a margin of $\pm 5\%$ on the probability estimates and later run sensitivity with margins of $\pm 2\%$ and $\pm 8\%$. A visualization of the different models' behavior is depicted in 5.1.

5.1.2 Dose-Volume Histograms

In this section, we examine the DVH plots which detail the full dose distribution to each structure. We begin by inspecting the setting in which the nominal position of the tumor is realized. Then, we inspect the effects of rigid shifts and discuss the extent to which RO and DRO protect against uncertainty.

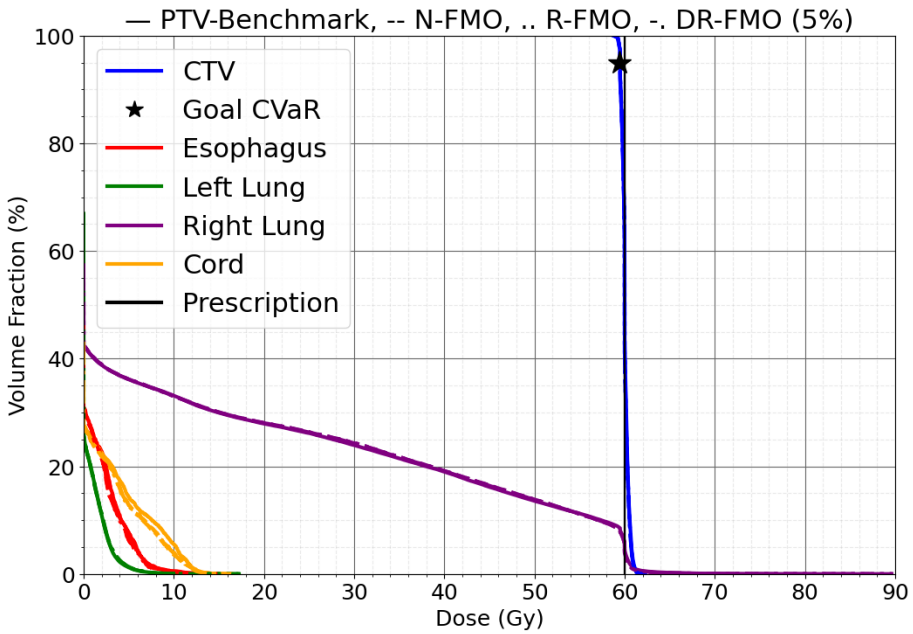


Figure 5.2: Depiction of the dose distribution received by the CTV and all OARs if the nominal position of the ctv is realized.

Nominal Setting

To begin, we inspect the dose-volume histogram realized in the nominal setting, shown in Figure 5.2. This describes the dose distribution received by the CTV and all OARs if the nominal position of the CTV is realized. Recall that the PTV defines a margin around the CTV to account for geometric uncertainty, but this margin contains non-cancerous tissue. As such, we are most interested in how well the PTV-benchmark covers the CTV. Note that the star denotes the DVH criterion, which CVaR approximates. In actuality, we require the mean of the tail above the vertical position of the star to be less than the horizontal position of the star. The dashed curve is the nominal FMO model when the target is the CTV. The dotted curve is the robust FMO model, which accounts for rigid shifts of the CTV. The dash-dot curve is the distributionally robust FMO model, which accounts for rigid shifts of the CTV as well as the worst-case probability distribution.

In all models, the dose distributions appear nearly identical. This is due to the granu-

larity of FMO models; each beamlet can be controlled individually to deliver the desired fluence. As such, it is easier to achieve an exceptional dose distribution in the nominal setting, which diminishes the perceived value of robustness.

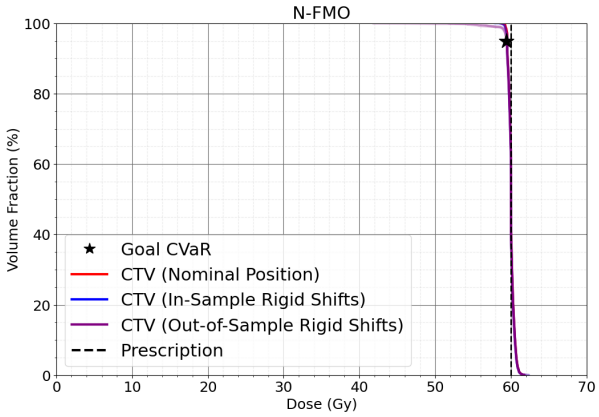
Rigid Shifts

Figure 5.3 examines the dose distribution to the CTV under different realizations of the CTV's position: nominal, in-sample, and out-of-sample, if treated by each of N-FMO, R-FMO, DR-FMO, and PTV-Benchmark. The blue curves each represent the dose distribution to the CTV if an in-sample scenario is realized. The purple curves each represent the dose distribution to the CTV if an out-of-sample scenario is realized. Finally, the star represents the DVH criterion which is approximated by CVaR. Generally, a model with more robust tumor coverage requires that more curves be above and to the right of the star.

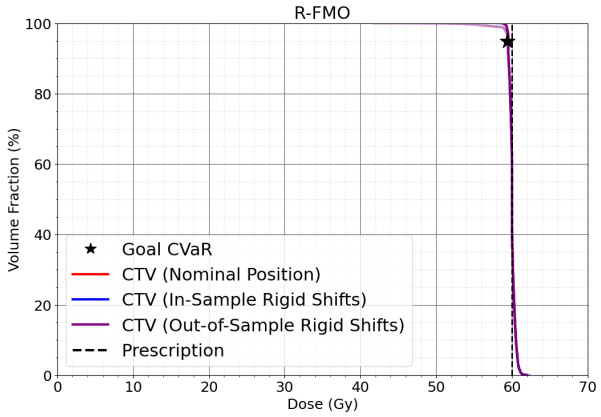
Similarly to the nominal case, the distributions obtained to the rigid shifts appear nearly identical. This corroborates the previous discussion: due to the flexibility of controlling individual beamlets, an excellent dose is obtained in the nominal setting which is also robust. Note that N-FMO and PTV-Benchmark look identical as well. Naturally, this does not mimic what really happens in practice.

5.1.3 Sensitivity Analysis and Out-of-sample Performance

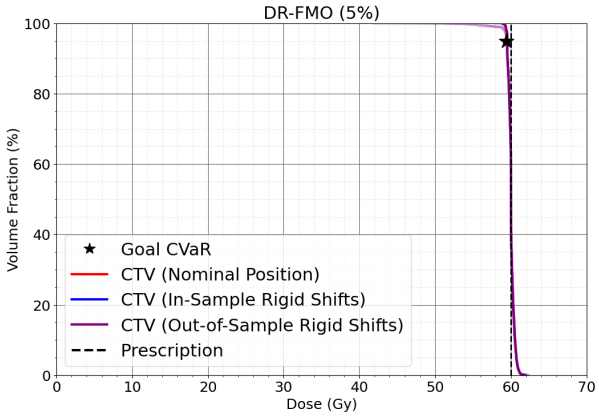
In addition to considering the entire dose distribution of the CTV, we summarize the impact of robustness in Figure 5.4 by examining two particular points on the distribution: the mean dose and the bottom 5%-CVaR generated by each model under each of the rigid shifts. Blue represents the nominal model, green considers geometric uncertainty, different shades of purple consider geometric and distributional uncertainty with varying sizes of the ambiguity set, and red is the nominal model that targets the PTV. Each dot represents one of the shifts. Dots with a border are in-sample, whereas dots without borders are out-of-sample. Finally, the dotted line represents the dose we enforce in the chance constraint.



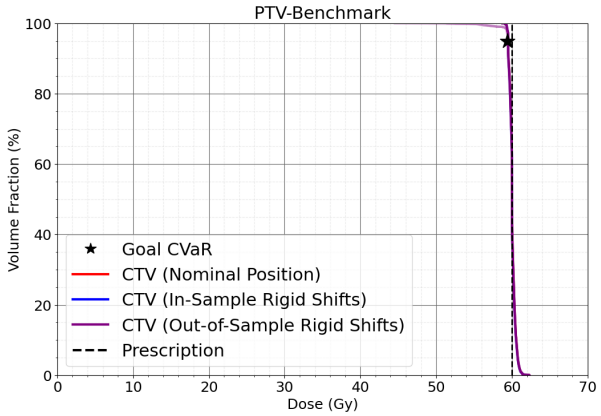
(a) CTV dose distribution under all scenarios for N-SCP.



(b) CTV dose distribution under all scenarios for R-SCP.



(c) CTV dose distribution under all scenarios for DR-SCP (5%).



(d) CTV dose distribution under all scenarios for PTV-Benchmark.

Figure 5.3: Full dose distribution resulting from FMO models in the nominal setting, in-sample shifts, and out-of-sample shifts.

First, we note that adding geometric or distributional robustness does not change the mean dose under different realizations of the uncertainty set. However, the PTV-based model enjoys a tighter mean dose. Second, we note that visualizing the CVaR provides a more meaningful comparison between the models. The differences are slight, but it is

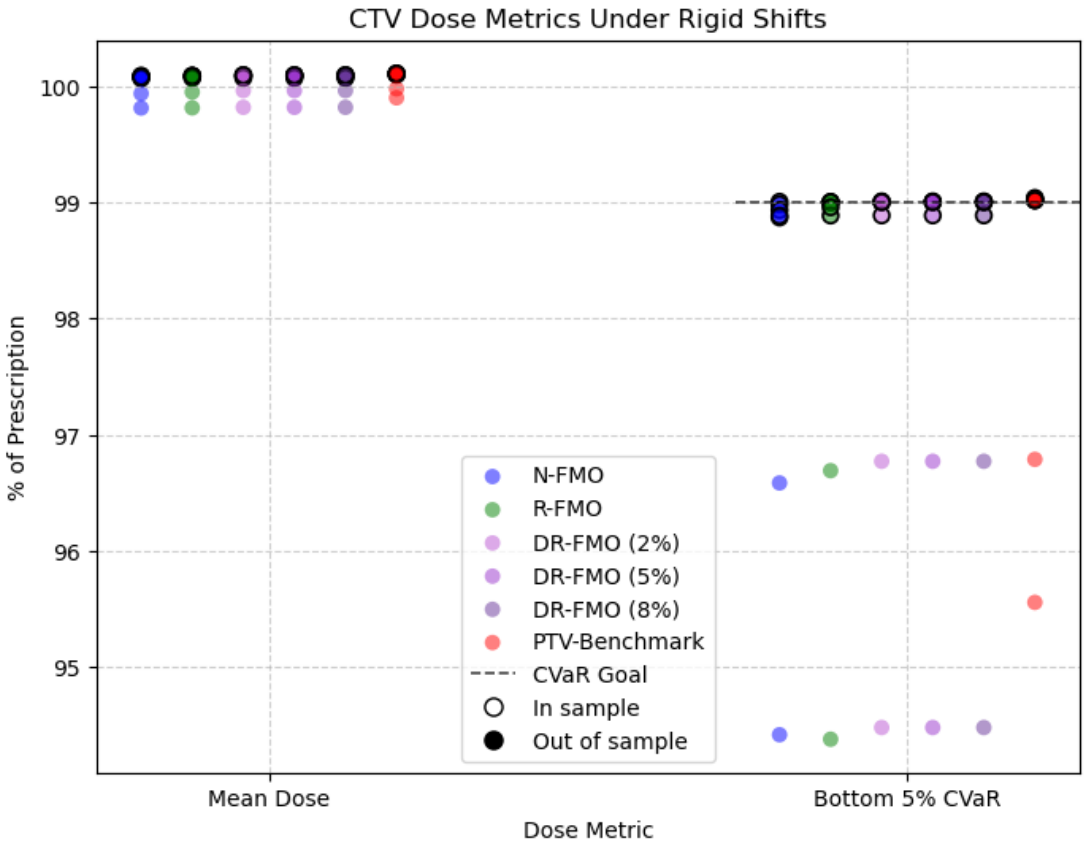


Figure 5.4: A summary of the mean dose and bottom CVaR of the CTV under each model.

clear that N-FMO's lower mean tail begins to slip as the CTV shifts. R-FMO controls this slightly better. DR-FMO shows further improvements for both in-sample and out-of-sample realizations, but the size of the ambiguity set makes no difference. Finally, the PTV-Benchmark achieves the most robust tumor coverage.

Table 5.2 illustrates the numerical CVaR values obtained by each model. As expected, the metrics are similar across models and their differences are not meaningful in practice. To illustrate the behavior of the chance constraints, we use two decimals to illustrate which scenarios are protected. The nominal model meets the CVaR requirement in the nominal case, but falls short on in-sample and out-of-sample rigid shifts. The robust model also meets the nominal CVaR requirement and protects against two of the in-sample rigid shifts. Given the worst case distribution, the distributionally robust model protects against three of the in-sample rigid shifts. This has auxiliary benefits, namely, a higher nominal CVaR, a higher unprotected in-sample CVaR, and higher out-of-sample CVaR. This implies that there is a dependence between rigid shifts: delivering a higher dose to one scenario implicitly satisfies nearby positions, which implies that protected against a small number of rigid shifts is a reasonable proxy to considering every possible movement. DR models with varying ambiguity set sizes achieve identical CVaR metrics. Finally, in terms of satisfying CVaR in many scenarios, the PTV benchmark achieves the most robust performance.

Scenario	N-FMO	R-FMO	DR-FMO (2%)	DR-FMO (5%)	DR-FMO (8%)	PTV-Benchmark
Nominal	**99.00**	**99.00**	**99.00**	**99.00**	**99.00**	**99.02**
3 mm left	98.93	**99.00**	**99.01**	**99.01**	**99.01**	**99.02**
3 mm right	98.98	**99.00**	**99.00**	**99.00**	**99.00**	**99.03**
6 mm left	98.88	98.95	**99.00**	**99.00**	**99.00**	**99.02**
6 mm right	98.87	98.88	98.89	98.89	98.89	**99.03**
9 mm left	96.58	96.69	96.77	96.77	96.77	96.79
9 mm right	94.42	94.38	94.48	94.48	94.48	95.56

Table 5.2: The mean dose received by the 5% least-dosed voxels in the CTV under each shift when treated by each model. Results are separated by nominal, in-sample, and out-of-sample realizations. Bold indicates that the CVaR requirement is satisfied in that scenario.

Next, we examine the dose distribution to the organs at risk. Recall that each model's objective function minimizes the absolute deviation between the dose delivered and the dose prescribed. Since each OAR is prescribed a dose of 0, this term is equivalent to minimizing the average dose each OAR receives. Although more dose requirements can be specified as constraints, we omit them as they clash with the target CVaR constraint, often leading to infeasibility. Dose requirements can also be specified as part of a multi-objective model; the focus of this study is to test the impacts of robustness, so we focus on a simple linear deviation from the prescription similar to the literature. Figure 5.5 shows a box-plot of the dose distribution to each OAR, including the mean and the upper quantiles. Blue depicts the nominal model, green depicts the robust model for geometric uncertainty, shades of purple represent the distributionally robust model for geometric and distributional uncertainty under varying sizes of the ambiguity set, and red represents the PTV-Benchmark model. As such, the nominal positions of the OARs are always realized. As shown in Figure 5.5, the OAR dose distributions are nearly identical due to the flexibility of FMO.

5.2 Direct Aperture optimization

5.2.1 Experimental Design

We take the same experimental design as in 5.1.1 with the following changes. First, instead of FMO models, we compare their DAO counterparts described in Table 5.3. Second, these DAO models cannot be solved at a clinical scale. As such, to reduce the number of integer and binary variables, we increase the beamlet size to 12.5 mm by 12.5 mm. We also double the voxel resolution in the right-left and anterior-posterior directions. This update slightly changes our uncertainty set: we consider in-sample shifts of approximately ± 6 mm and ± 12 mm as well as out-of-sample shifts of approximately ± 18 mm. Third, due to the decrease in granularity from down-scaling and requiring deliverable apertures, we slightly relax the CVaR constraint and require an average of at least 57 Gy to the 10% least-dosed voxels. Table 5.4 compares the number of variables and constraints at each scale. Even at this smaller scale, the models could not be solved locally. As such, we run them on

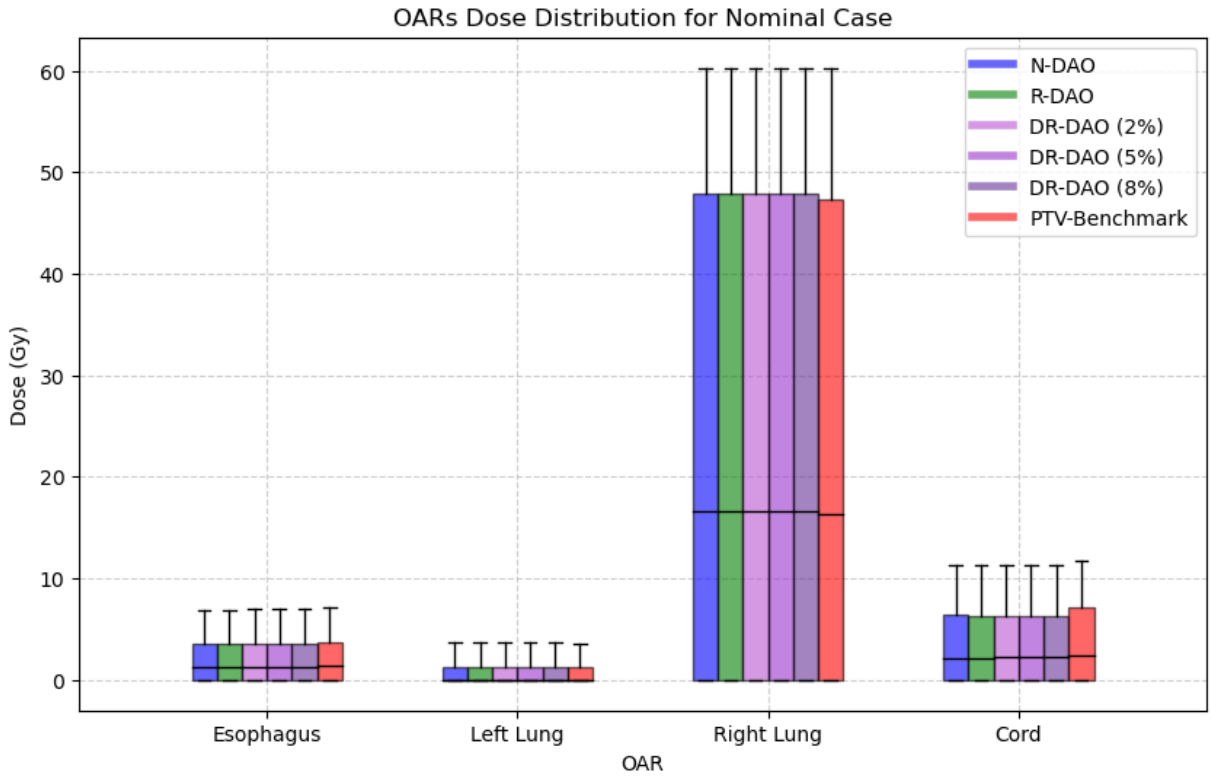


Figure 5.5: Box-plots representing the dose distribution to each OAR when treated by each model.

the Compute Canada Cluster, which uses an AMD EPYC 7532 processor and 512 GB of RAM, and solve them using Gurobi version 12.0.0.

5.2.2 Dose-Volume Histograms

In this section, we examine the DVH plots which detail the full dose distribution to each structure. We begin by inspecting the setting where the nominal position of the tumor is realized in 5.2.2. Then, we inspect the effects of rigid shifts and discuss the extent to which RO and DRO protect against uncertainty in 5.2.2.

Model	Description
N-DAO	Nominal DAO model with the CTV as the target
R-DAO	Robust DAO model with a chance constraint accounting for rigid shifts of the CTV in the nominal probability distribution
DR-DAO (5%)	Distributionally robust DAO model with an ambiguous chance constraint accounting for rigid shifts of the CTV in the worst-case probability distribution, where the ambiguity set allows for a 5% margin of error around each atom
PTV-Benchmark	Nominal DAO model with the PTV as the target

Table 5.3: DAO models tested in clinical-scale experiment.

	Variables						Constraints	
	Continuous		Binary		Integer			
	Full	Small	Full	Small	Full	Small	Full	Small
N-DAO	40,627	2,581	26,802	192	2,160	72	7,638	264
R-DAO	58,963	5,757	26,807	197	2,160	72	7,647	273
DR-DAO	58,974	5,768	26,807	197	2,160	72	7,650	276

Table 5.4: Problem size of each MILP instance, split into full and small scales for continuous, binary, and integer variables, as well as constraints, after compilation in CVXPy.

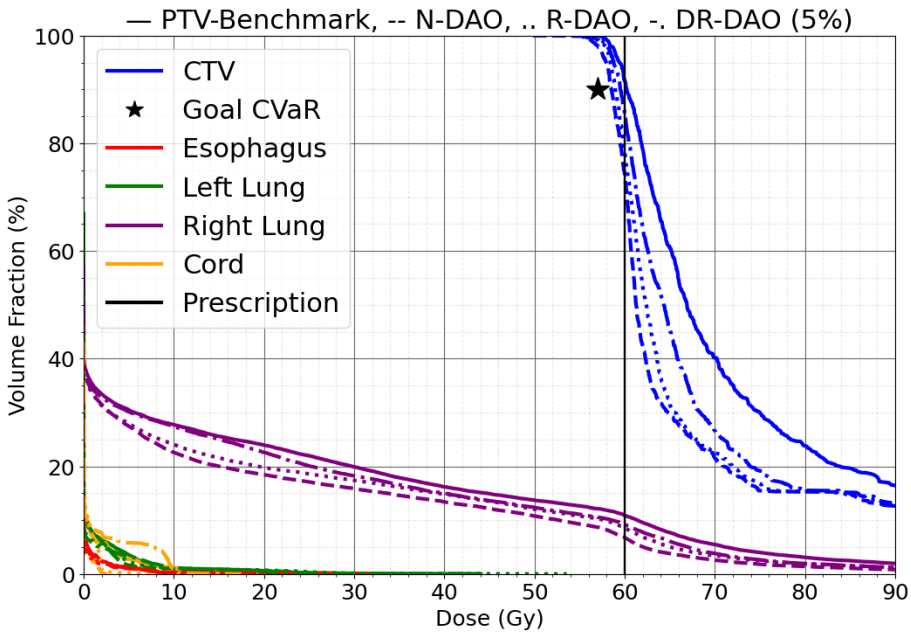


Figure 5.6: Depiction of the dose distribution received by the CTV and all OARs if the nominal position of the CTV is realized.

Nominal Setting

To begin, we inspect the dose-volume histogram realized in the nominal setting, shown in Figure 5.6. The dashed curve is the nominal DAO model when the target is the CTV. The dotted curve is the robust DAO model, which accounts for rigid shifts of the CTV. The dash-dot curve is the distributionally robust DAO model, which accounts for rigid shifts of the CTV as well as the worst-case probability distribution. Finally, the star denotes the DVH criterion, which CVaR approximates.

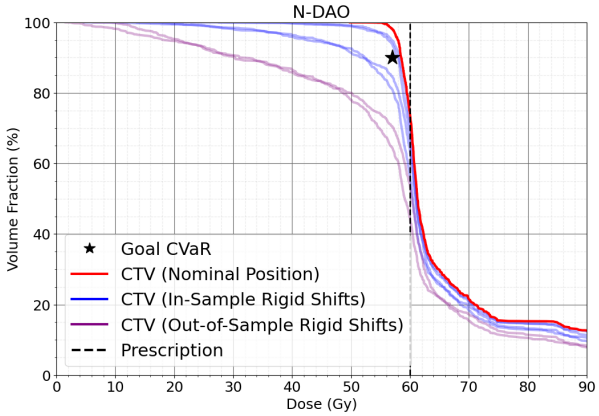
In both CTV coverage and OAR sparing, we see the general pattern: N-DAO delivers the lowest dose, followed by R-DAO, then DR-DAO, and finally PTV-Benchmark delivers the highest dose. This occurs even though our uncertainty set uses larger shifts, which are not contained within the PTV region. As such, N-DAO, which does not consider any rigid shifts at all, is able to better spare the nearby healthy tissues and result in lower (but still adequate) CTV coverage. PTV-Benchmark is radiating the entire area around the CTV,

so it achieves the highest CTV coverage and worst OAR sparing. R-DAO and DR-DAO result are strategically radiating several locations, some of which are outside of the PTV region. As such, they are able to better spare the OARs while adequately covering the CTV.

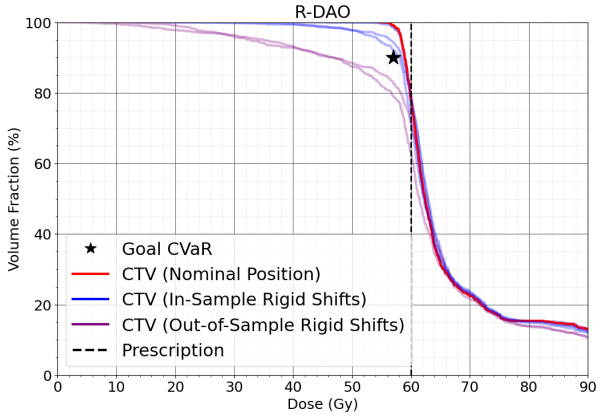
Rigid Shifts

Figure 5.7 examines the dose distribution to the CTV under different realizations of the CTV's position: nominal, in-sample, and out-of-sample, if treated by each of N-SCP, R-SCP, DR-SCP, and PTV-Benchmark. In each graph, the red curve is the dose distribution to the CTV if the nominal scenario is realized. The blue curves each represent the dose distribution to the CTV if an in-sample scenario is realized. The purple curves each represent the dose distribution to the CTV if an out-of-sample scenario is realized. Finally, the star represents the DVH criterion which is approximated by CVaR. Generally, a model with more robust tumor coverage requires that more curves be above and to the right of the star.

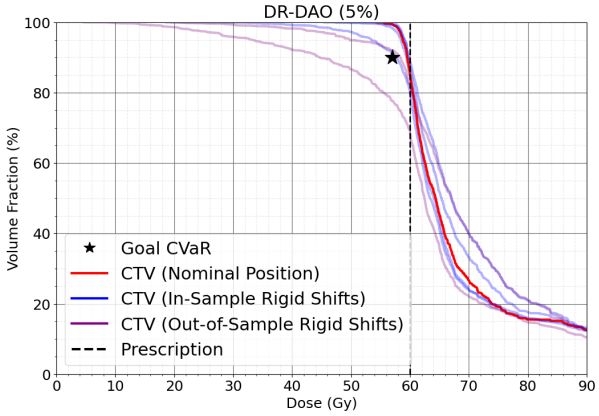
Figure 5.7a depicts N-DAO. As the CTV shifts, the lower tail of its dose distribution is decreased. This is congruent with the treatment approach, because N-DAO does not consider any uncertainty. However, under small in-sample shifts, the CVaR requirement is still approximately satisfied. Figure 5.7b depicts R-DAO. Most notably, distributions are closer to the prescription compared to the nominal treatment plan. Figure 5.7c depicts DR-DAO. We observe a similar trend, where the dose distribution becomes tighter and out-of-sample performance is improved. This shows the value of distributional robustness to protect against uncertainty in the underlying probability distribution in addition the geometric realization. Figure 5.7d depicts PTV-Benchmark. Recall that this examines the dose distribution to the CTV under different shifts if treated by the nominal model with PTV as its target (PTV-Benchmark). This results in the tightest distribution because all in-sample rigid shifts are contained within the PTV margin; however, it also delivers a much higher upper-tail dose, which may be harmful. Note that although the in-sample rigid shifts receive a higher dose in PTV-Benchmark compared to DR-DAO (5%), the distributionally robust model still meets the constraint and has the additional benefit of



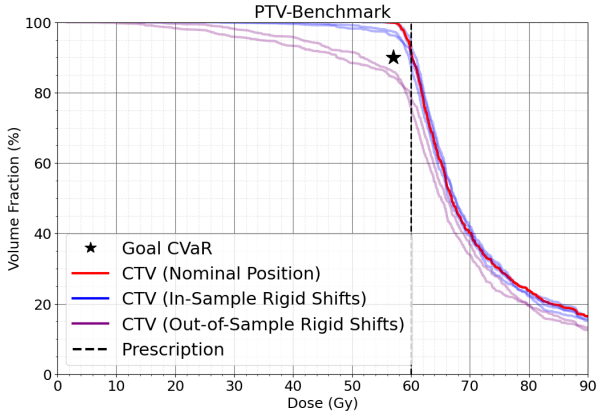
(a) CTV dose distribution under all scenarios for N-DAO.



(b) CTV dose distribution under all scenarios for R-DAO.



(c) CTV dose distribution under all scenarios for DR-DAO (5%).



(d) CTV dose distribution under all scenarios for PTV-Benchmark.

Figure 5.7: Full dose distribution resulting from FMO models in the nominal setting, in-sample shifts, and out-of-sample shifts.

better out-of-sample performance.

5.2.3 Sensitivity Analysis and Out-of-sample Performance

In addition to considering the entire dose distribution of the CTV, we summarize the impact of robustness in Figure 5.8 by examining two particular points on the distribution: the mean dose and the bottom 10%-CVaR generated by each model under each of the rigid shifts. First, we note that adding geometric robustness results in a higher mean dose relative to the nominal model. While this may seem positive at first glance, recall that our objective function minimizes the deviation between the dose received and prescribed. As such, the nominal model enjoys a better mean dose. However, the PTV-based models has the highest mean dose. As such, although R-DAO and DR-DAO have higher mean doses than the nominal model, they perform better than what is used in clinical practice. Second, we note that enforcing the worst-case probability chance constraint results in improvements to the tightness of the bottom 10% CVaR. "Tighter", in this setting, means that a more similar dose is received in more scenarios. The CVaR is tightest in the PTV benchmark plan because of spillover from the higher mean dose. However, the PTV approach neglects two in-sample scenarios, unlike the distributionally robust approach which neglects only one. Further, the DR-DAO models outperform N-DAO and R-DAO in terms of robustness. Finally, we note that varying the radius of the ambiguity set does not result in meaningful changes to these metrics, which is what the models explicitly control for.

Table 5.5 illustrates the numerical CVaR values obtained by each model. As expected, the nominal model meets the CVaR requirement in the nominal case, but falls short on in-sample and out-of-sample rigid shifts. The robust model also meets the nominal CVaR requirement and protects against two of the in-sample rigid shifts. Given the worst case distribution, the distributionally robust model protects against three of the in-sample rigid shifts. This has auxiliary benefits, namely, a higher nominal CVaR, a higher unprotected in-sample CVaR, and higher out-of-sample CVaR. Once again, this implies that there is a dependence between rigid shifts: delivering a higher dose to one scenario implicitly satisfies nearby positions, which implies that protected against a small number of rigid shifts is a reasonable proxy to considering every possible movement. DR models with

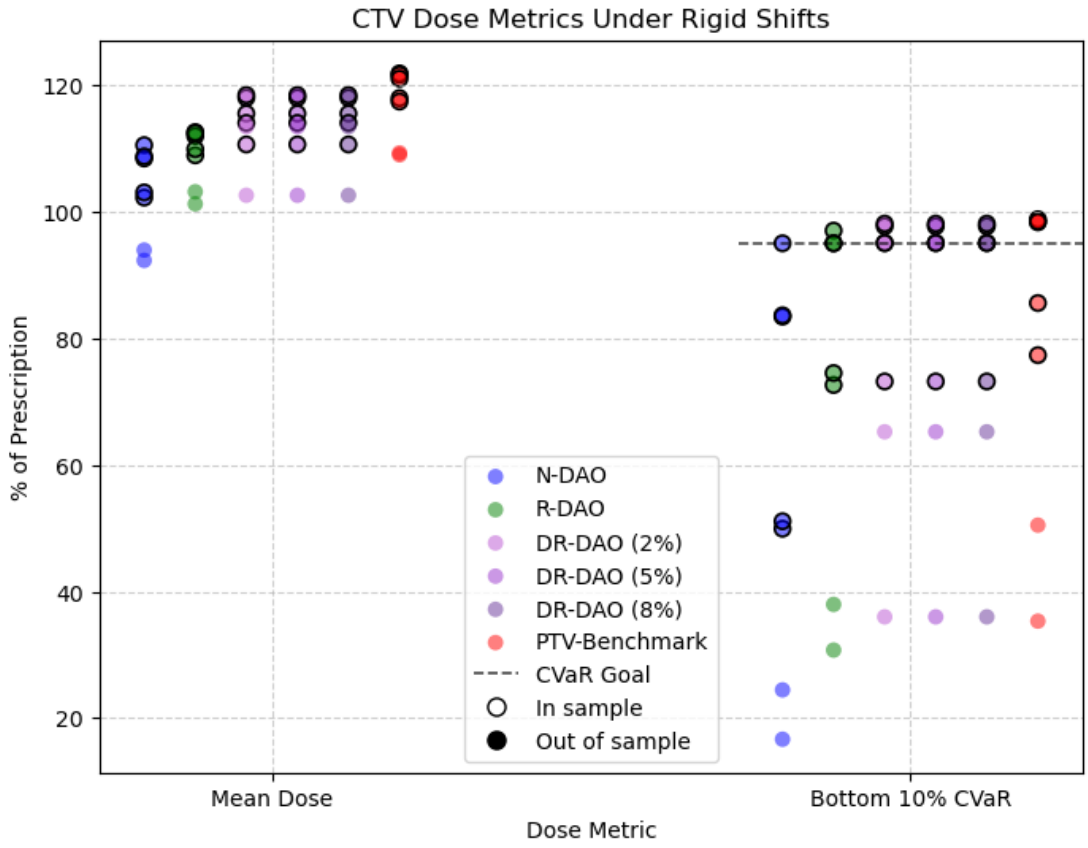


Figure 5.8: A summary of the mean dose and bottom CVaR of the CTV under each model.

Scenario	N-DAO	R-DAO	DR-DAO (2%)	DR-DAO (5%)	DR-DAO (8%)	PTV-Benchmark
Nominal	**95.0**	**97.0**	**97.7**	**97.7**	**97.7**	**98.8**
6 mm left	83.6	**95.0**	**98.1**	**98.1**	**98.1**	**98.5**
6 mm right	83.4	**95.0**	**95.1**	**95.1**	**95.1**	**98.3**
12 mm left	51.1	74.5	**95.0**	**95.0**	**95.0**	85.6
12 mm right	49.9	72.6	73.2	73.2	73.2	77.3
18 mm left	24.5	38.0	65.2	65.2	65.2	50.5
18 mm right	16.7	30.7	36.0	36.0	36.0	35.3

Table 5.5: The mean dose received by the 10% least-dosed voxels in the CTV under each shift when treated by each model. Results are separated by nominal, in-sample, and out-of-sample realizations.

varying ambiguity set sizes achieve identical CVaR metrics. Finally, in terms of satisfying CVaR in many scenarios, the PTV benchmark achieves similar performance to R-DAO but fails to adequately radiate shifts outside of its margin. Hence, it is outperformed by the distributionally robust models.

Next, we examine the dose distribution to the organs at risk. As before, recall that each model's objective function minimizes the average dose each OAR receives and that our problem setting assumes no movement in the OARs. Figure 5.9 shows a box-plot of the dose distribution to each OAR, including the mean and the upper quantiles. For the right lung, which contains the tumor, we make the following observations. First, incorporating geometric uncertainty does not result in material changes to the mean dose. Second, incorporating distributional uncertainty results in an increase. However, the variance of the dose between R-DAO and the DR-DAO models is similar. Third, DR-DAO still maintains a tighter range and lower mean than the PTV-Benchmark. This implies that, despite having a slightly looser CTV distribution, it is able to better spare the nearby OARs. For OARs further from the tumor (namely the esophagus, right lung, and spinal cord), we note that all models generally result in similar distributions. The robust and distributionally robust models provided dose distributions with higher variation, typically followed by the PTV-Benchmark. However, given the magnitude of these changes, we claim that adding

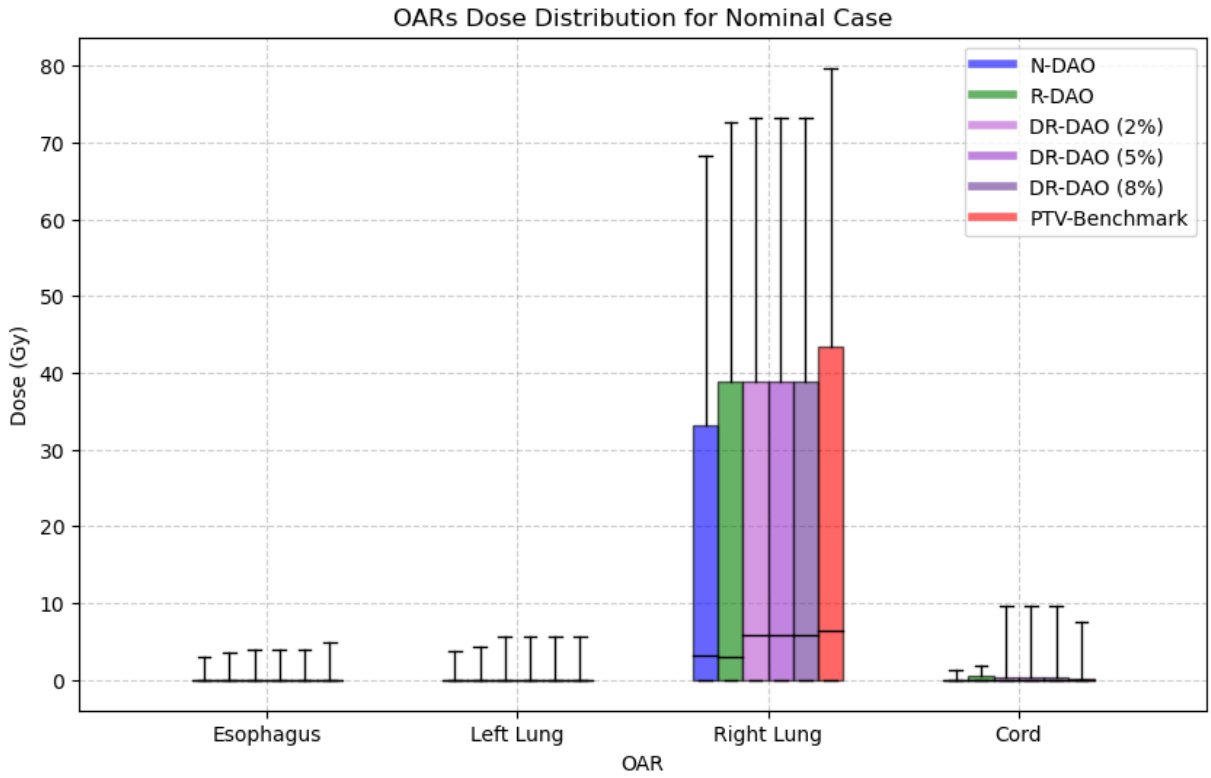


Figure 5.9: Box-plots representing the dose distribution to each OAR when treated by each model.

geometric or distributional robustness does not meaningfully change the dose distribution to OARs far from the tumor. Finally, varying the size of the ambiguity set has little effect on the OAR dose distribution.

5.3 Sequential Convex Programming

This section will investigate the impacts of robustness and the quality of the convex approximation. First, we define our experimental setup in 5.3.1. Then, we compare the effects of different levels of robustness using the SCP models at a clinical scale. We do this by comparing DVH plots in 5.3.2 and through a sensitivity analysis on the size of the

ambiguity set in 5.3.3. In 5.3.4, we examine the quality of the convex approximation by comparing the MILP and SCP models in a small-scale case for which the MILP can be solved to optimality. Finally, we compare the computational burden of each model in 5.3.5.

5.3.1 Experimental Design

To compare the impacts of robustness, we compare each model in Table 5.6 at a clinical scale on a lung patient treated with VMAT from the dataset made available by Memorial Sloan Kettering Cancer Center. To achieve a clinical scale, we discretize the arc into 72 equidistant control points and use the finest beamlet resolution of 2.5 mm by 2.5 mm. As before, we develop our models in Python using the PortPy package (Jhanwar et al., 2023) and solve them using Gurobi version 12.0.0 on the Compute Canada Cluster, which uses an AMD EPYC 7532 processor and 512 GB of RAM. For each patient, we define the CTV by placing a 5 mm margin around the GTV to account for cancerous cells that could not be detected by the scan.

Model	**Description**
N-SCP	Nominal SCP algorithm with the CTV as the target
R-SCP	Robust SCP algorithm with a chance constraint accounting for rigid shifts of the CTV in the nominal probability distribution
DR-SCP (5%)	Distributionally robust SCP algorithm with an ambiguous chance constraint accounting for rigid shifts of the CTV in the worst-case probability distribution, where the ambiguity set allows for a 5% margin of error around each atom
PTV-Benchmark	Nominal SCP algorithm with the PTV as the target

Table 5.6: Models tested in clinical-scale experiment.

In all models, the prescribed target dose is 60 Gy and we enforce the CVaR constraint to deliver an average of at least 57 Gy to the 10% least-dosed voxels. In the robust and distributionally robust models, our uncertainty set includes the nominal scenario and

simulated shifts to the left and right by 3 mm and 6 mm, which is the finest scale available for this patient's voxels. We also use out-of-sample simulated left and right shifts by 9 mm of test robustness on unseen cases. As before, we adopt the approach in An et al. (2017) where the nominal scenario occurs with 60% probability and the remaining probability is distributed uniformly across the remainder of the uncertainty set. For the ambiguity set in the distributionally robust model, we define a margin of ±5% on the probability estimates and later run sensitivity with margins of ±2% and ±8%.

5.3.2 Dose-Volume Histograms

In this section, we examine the DVH plots which detail the full dose distribution to each structure. We begin by inspecting the setting where the nominal position of the tumor is realized in 5.3.2. Then, we inspect the effects of rigid shifts and discuss the extent to which RO and DRO protect against uncertainty in 5.3.2.

Nominal Setting

To begin, we inspect the dose-volume histogram realized in the nominal setting, shown in Figure 5.10. Regarding CTV coverage, all models result in a similar CTV dose distribution. N-SCP and R-SCP have slightly lower tails, but still fulfill the CVaR constraint in the nominal setting. Regarding OAR sparing, we see the hierarchy that the nominal model is better than the robust model, which is better than the distributionally robust model, which is better than the PTV benchmark. This occurs because the uncertainty set we use considers rather small shifts, which are likely contained within the PTV region. As such, N-SCP, which does not consider any rigid shifts at all, is able to better spare the nearby healthy tissues (most notably, the left lung, in which the tumor is situated). R-SCP results in similar sparing of the left lung but higher doses to the other OARs. DR-SCP delivers a higher dose to the OARs due to the increased requirement on robustness to protect against the worst-case probability distribution. Finally, the PTV-Benchmark results in the worst OAR sparing for nearby OARs such as the left lung, because it assumes an entire area around the CTV is cancerous.

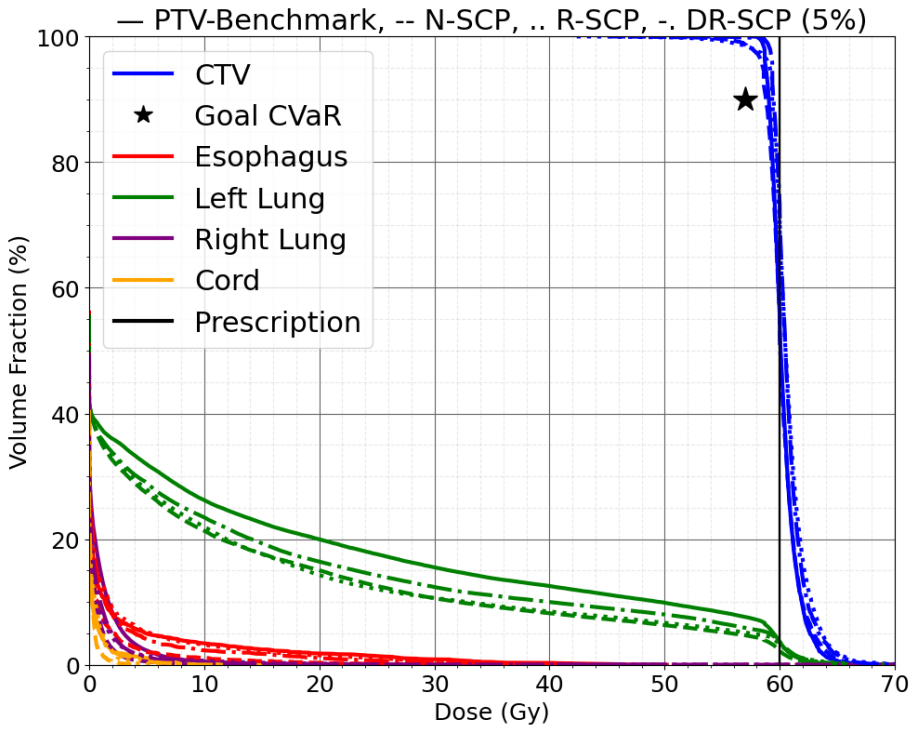


Figure 5.10: Depiction of the dose distribution received by the CTV and all OARs if the nominal position of the CTV is realized.

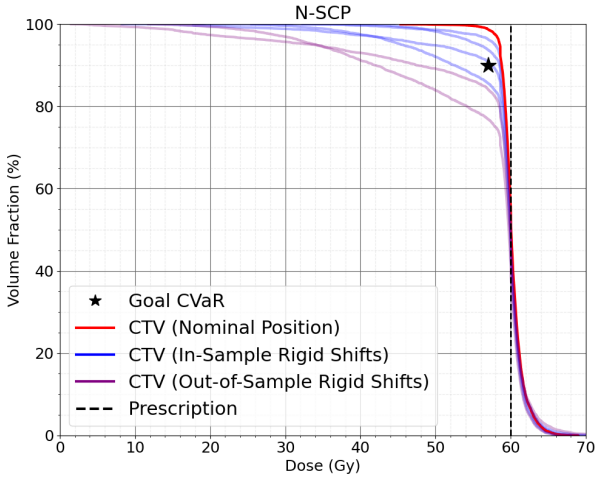
Rigid Shifts

Figure 5.11 examines the dose distribution to the CTV under different realizations of the CTV's position: nominal, in-sample, and out-of-sample, if treated by each of N-SCP, R-SCP, DR-SCP, and PTV-Benchmark. In each graph, the red curve is the dose distribution to the CTV if the nominal scenario is realized. The blue curves each represent the dose distribution to the CTV if an in-sample scenario is realized. The purple curves each represent the dose distribution to the CTV if an out-of-sample scenario is realized. Finally, the star represents the DVH criterion which is approximated by CVaR. Generally, a model with more robust tumor coverage requires that more curves be above and to the right of the star.

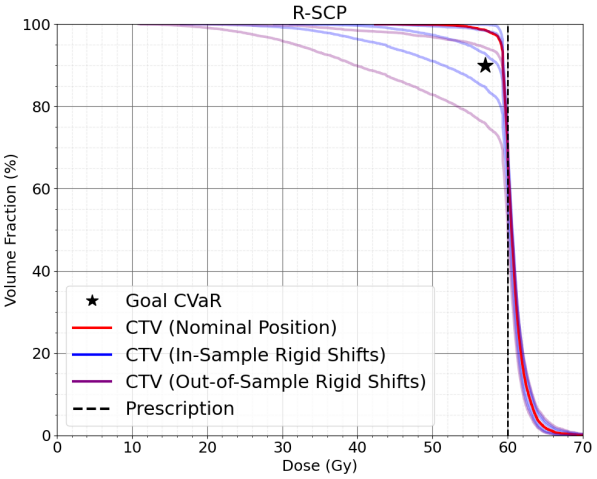
Figure 5.11a depicts N-SCP. As the CTV shifts, the lower tail of its dose distribution is decreased. This is congruent with the treatment approach, because N-SCP does not consider any uncertainty. However, under small in-sample shifts, the CVaR requirement is still approximately satisfied. Figure 5.11b depicts R-SCP. Most notably, distributions are closer to the prescription compared to the nominal treatment plan. This result confirms that the results from the literature hold in VMAT with SCP – incorporating geometric uncertainty as a chance-constraint leads to increased robustness. Figure 5.11c depicts DR-SCP. We observe a similar trend, where the dose distribution becomes tighter and out-of-sample performance is improved. This observation shows the value of distributional robustness to protect against uncertainty in the underlying probability distribution in addition the geometric realization. Figure 5.11d depicts PTV-Benchmark, which delivers the tightest distribution because all rigid shifts are contained within the PTV margin.

5.3.3 Sensitivity Analysis and Out-of-sample Performance

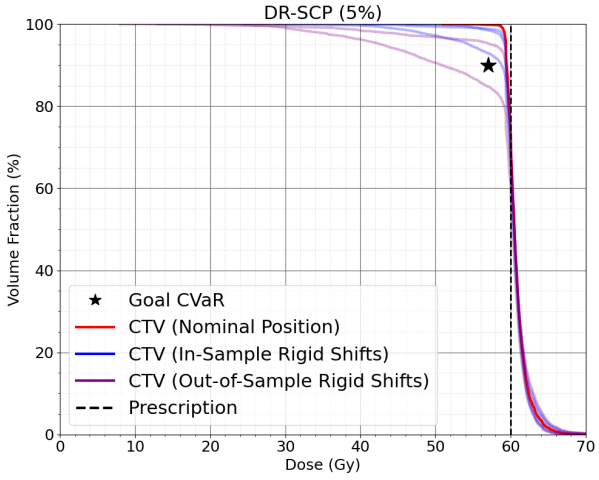
In addition to considering the entire dose distribution of the CTV, we summarize the impact of robustness in Figure 5.12 by examining two particular points on the distribution: the mean dose and the bottom 10%-CVaR generated by each model under each of the rigid shifts. First, we note that adding geometric robustness results in a slightly tighter mean dose relative to the nominal model. Second, we note that enforcing the worst-case



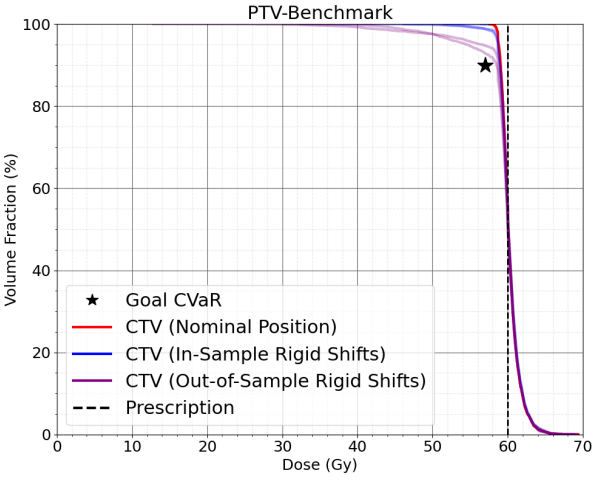
(a) CTV dose distribution under all scenarios for N-SCP.



(b) CTV dose distribution under all scenarios for R-SCP.



(c) CTV dose distribution under all scenarios for DR-SCP (5%).



(d) CTV dose distribution under all scenarios for PTV-Benchmark.

Figure 5.11: Full dose distribution resulting from SCP models in the nominal setting, in-sample shifts, and out-of-sample shifts.

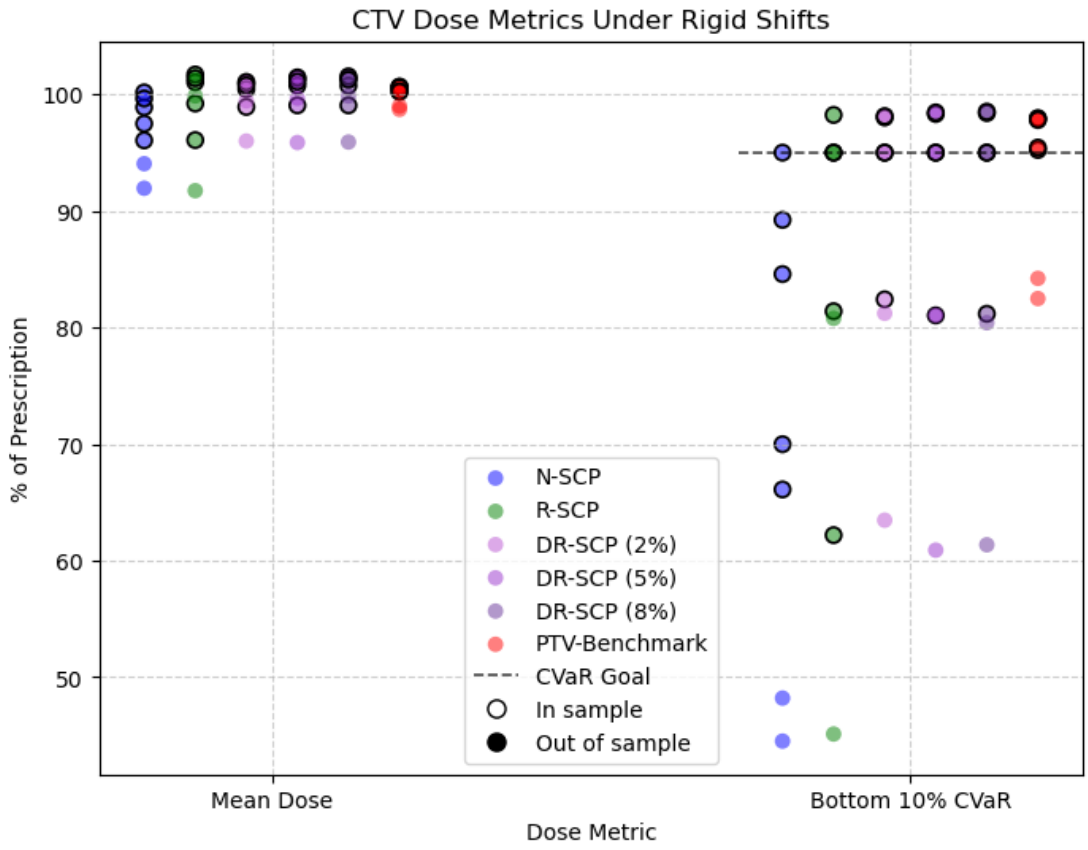


Figure 5.12: A summary of the mean dose and bottom CVaR of the CTV under each model.

probability chance constraint results in even larger improvements to the tightness of the mean dose. Although increasing the radius of the ambiguity set slightly increases tumor coverage, the changes in dose are not meaningful. As before, the mean dose is tightest in the PTV benchmark plan because the margin incorporates the potential shifts.

The bottom CVaR results in more meaningful improvements. In the nominal CTV plan, the average of the 10% least-dosed voxels in the tumor receive as low as 66% of the prescription if in-sample shifts are realized. Coverage drops to lower than 50% if out-of-sample shifts are realized. Incorporating geometric uncertainty improves this for protected scenarios at the slight expense of unprotected scenarios. Incorporating distributional ro-

bustness results in a tighter CVaR because, under the worst case probability distribution, more scenarios must be protected in order to satisfy the chance constraint. This also results in tighter out-of-sample performance. Changing the size of the ambiguity set does not result in meaningful changes to the dose distribution. As with the mean dose, the PTV-benchmark results in the tightest CVaR because its margin is wide enough to consider these scenarios. Recall, however, that the PTV margin includes healthy tissues. Although both the robust CC and distributionally robust CC are set to the same tolerance, the distributionally robust CC protects one extra in-sample scenario. This is because in the worst-case probability distribution, additional protection is needed to satisfy the tolerance.

Scenario	N-SCP	R-SCP	DR-SCP (2%)	DR-SCP (5%)	DR-SCP (8%)	PTV-Benchmark
Nominal	**95.0**	**95.0**	**98.0**	**98.3**	**98.4**	**98.0**
3 mm left	89.2	**98.2**	**98.1**	**98.5**	**98.5**	**97.8**
3 mm right	84.6	81.4	**95.0**	**95.0**	**95.0**	**97.9**
6 mm left	70.0	**95.0**	**95.0**	**95.0**	**95.0**	**95.4**
6 mm right	66.1	62.2	82.4	81.0	81.1	**95.2**
9 mm left	44.5	80.8	81.2	81.0	80.4	84.2
9 mm right	48.2	45.1	63.5	60.9	61.3	82.5

Table 5.7: The mean dose received by the 10% least-dosed voxels in the CTV under each shift when treated by each model. Results are separated by nominal, in-sample, and out-of-sample realizations. Bold indicates that the CVaR requirement is satisfied.

Table 5.7 illustrates the numerical CVaR values obtained by each model. As expected, the nominal model meets the CVaR requirement in the nominal case, but falls short on in-sample and out-of-sample rigid shifts. The robust model also meets the nominal CVaR requirement and protects against two of the in-sample rigid shifts. Given the worst case distribution, the distributionally robust model protects against three of the in-sample rigid shifts. This protection has auxiliary benefits, namely, a higher nominal CVaR, a higher unprotected in-sample CVaR, and higher out-of-sample CVaR. This implies that there is a dependence between rigid shifts: delivering a higher dose to one scenario implicitly satisfies nearby positions, which implies that protected against a small number of rigid shifts is a

reasonable proxy to considering every possible movement. Again, there are few differences between DR models with varying ambiguity set sizes. Finally, the PTV benchmark meets the dose requirement for all shifts as they are contained within the PTV region but fails in out-of-sample shifts.

Next, we examine the dose distribution to the organs at risk in Figure 5.13. For the left lung, which contains the tumor, we make the following observations. First, incorporating more robustness typically results slight increases to the mean dose. Second, incorporating geometric robustness leads to lower quantiles in the dose distribution and the opposite holds for distributional robustness. However, DR-SCP still maintains a tighter range and lower mean than the PTV-Benchmark. This implies that, despite having a slightly looser CTV distribution, it is able to better spare the nearby OARs. Finally, varying the size of the ambiguity set has little effect on the OAR dose distribution. For OARs further from the tumor (namely the esophagus, right lung, and spinal cord), we note that all models result in similar distributions. The robust and distributionally robust models provided dose distributions with higher mean and variation, followed by the PTV-Benchmark. However, given the magnitude of these changes, we claim that adding geometric or distributional robustness does not meaningfully change the dose distribution to OARs far from the tumor.

5.3.4 Quality of Convex Approximation

Recall that the SCP models approximate the respective MILP models, and that the MILP models can only be solved at smaller scales. In this subsection, we empirically test the validity of the convex approximation by comparing the dose distributions under different levels of uncertainty. To reduce the scale while maintaining clinical relevance, we run the SCP models using the experimental design in 5.2.1 and compare them to the DAO counterparts. At this new scale, we define N-SCP, R-SCP, and DR-SCP as well as N-DAO, R-DAO, and DR-DAO as before.

Figure 5.14 shows the dose distributions for each level of robustness and compares the global optimal solution to the one obtained from the convex approximation. The dashed curves are the resulting distributions from the SCP models. For high quality approximations, we hope to see them near the solid curves, which are the ground-truth from

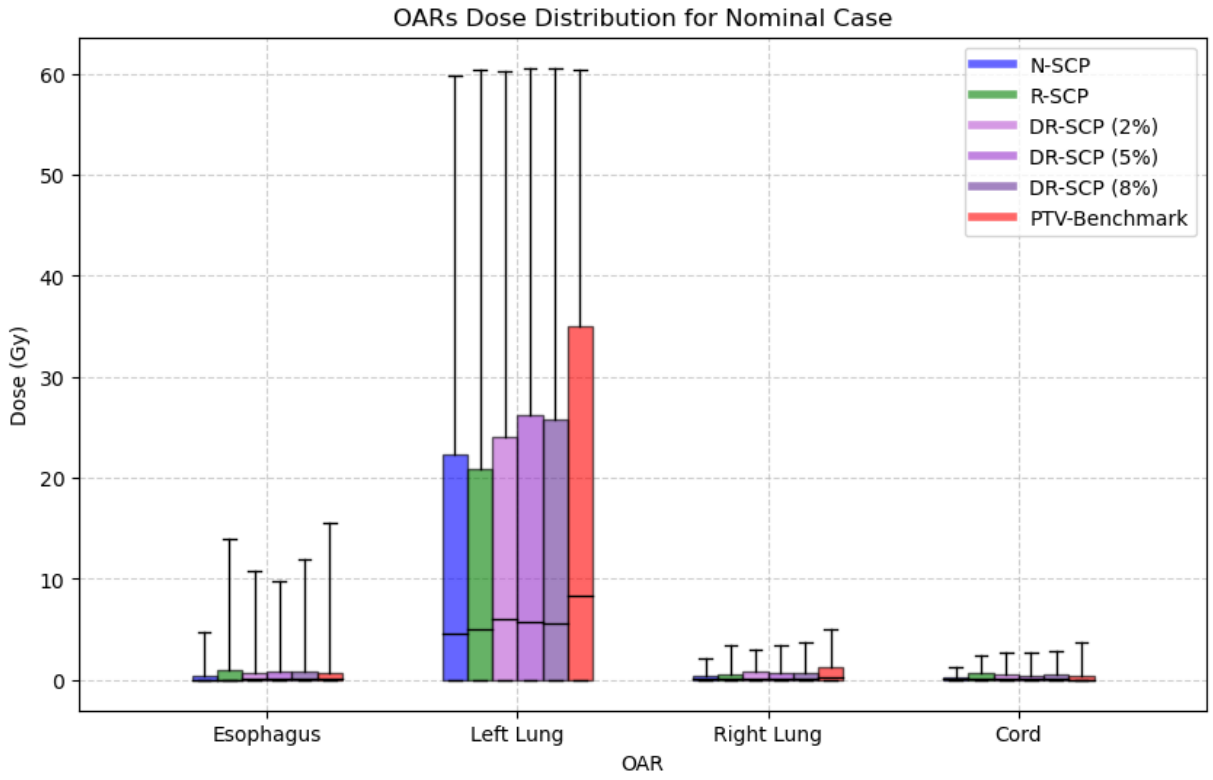


Figure 5.13: Box-plots representing the dose distribution to each OAR when treated by each model.

the DAO variants. In all cases, we note that the the distributions are similar. This implies that the sequential convex programming algorithm remains an accurate approximation in the robust and distributionally robust settings. Note that reducing the problem scale to this extent significantly degrades treatment quality. We remind the reader that the purpose of this section is not to deliver a high quality treatment but to inspect the quality of the convex approximation. Table 5.8 inspects the absolute and relative optimality gaps from each model under each level of uncertainty and notes that integrating RO and DRO with SCP does not degrade the approximation quality.

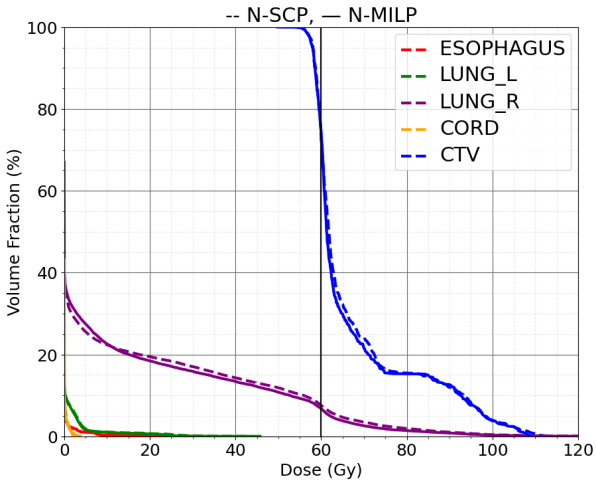
	Objective Values		**Optimality Gaps**	
	MIP	**SCP**	**Absolute**	**Relative (%)**
Nominal	1.27	1.38	0.11	8.34
Robust	1.41	1.49	0.08	5.62
Distributionally Robust	1.58	1.60	0.02	1.23

Table 5.8: Comparison of optimality gaps between MILP and SCP approaches.

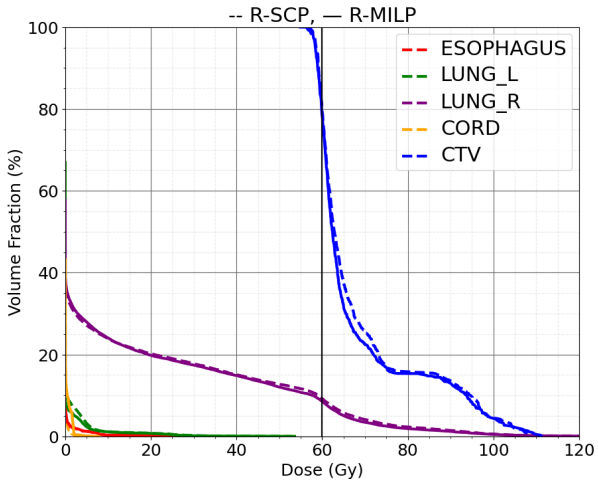
5.3.5 Computational Times

Finally, we illustrate the difference in computational burden between the exact and approximate models in Table 5.9. In the small-scale experiment, the SCP models result in a 99% reduction in solution time. At a clinical scale, the N-MILP problems resulted in an out-of-memory error when run on a computer cluster with 512 GB of RAM. N-SCP, however, was solved in approximately 12 minutes. Incorporating robustness led to an increase in solution time; R-SCP took roughly 8 hours DR-SCP took roughly 15. Despite the increase of compute time, these are still within a clinically acceptable time frame.

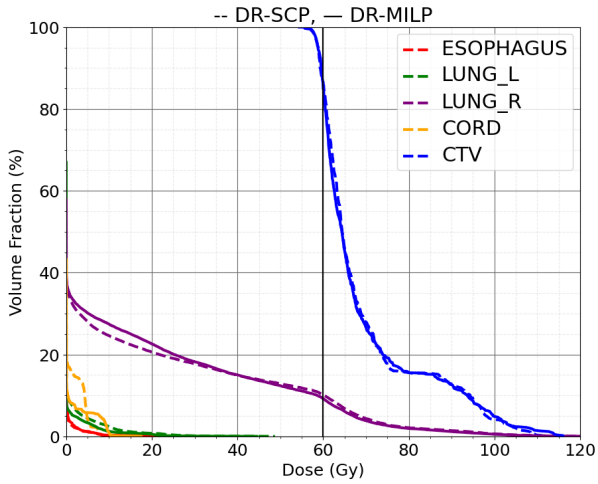
Figure 5.15 demonstrates the convergence behavior of each algorithm. In general, the SCP algorithm terminates when the step size has been 1 for two consecutive iterations and when the objective values of the intermediate problem converges to the actual objective value once the convex approximation is taken into account. The behavior of each variant



(a) Nominal ground-truth and approximation.



(b) RO ground-truth and approximation.



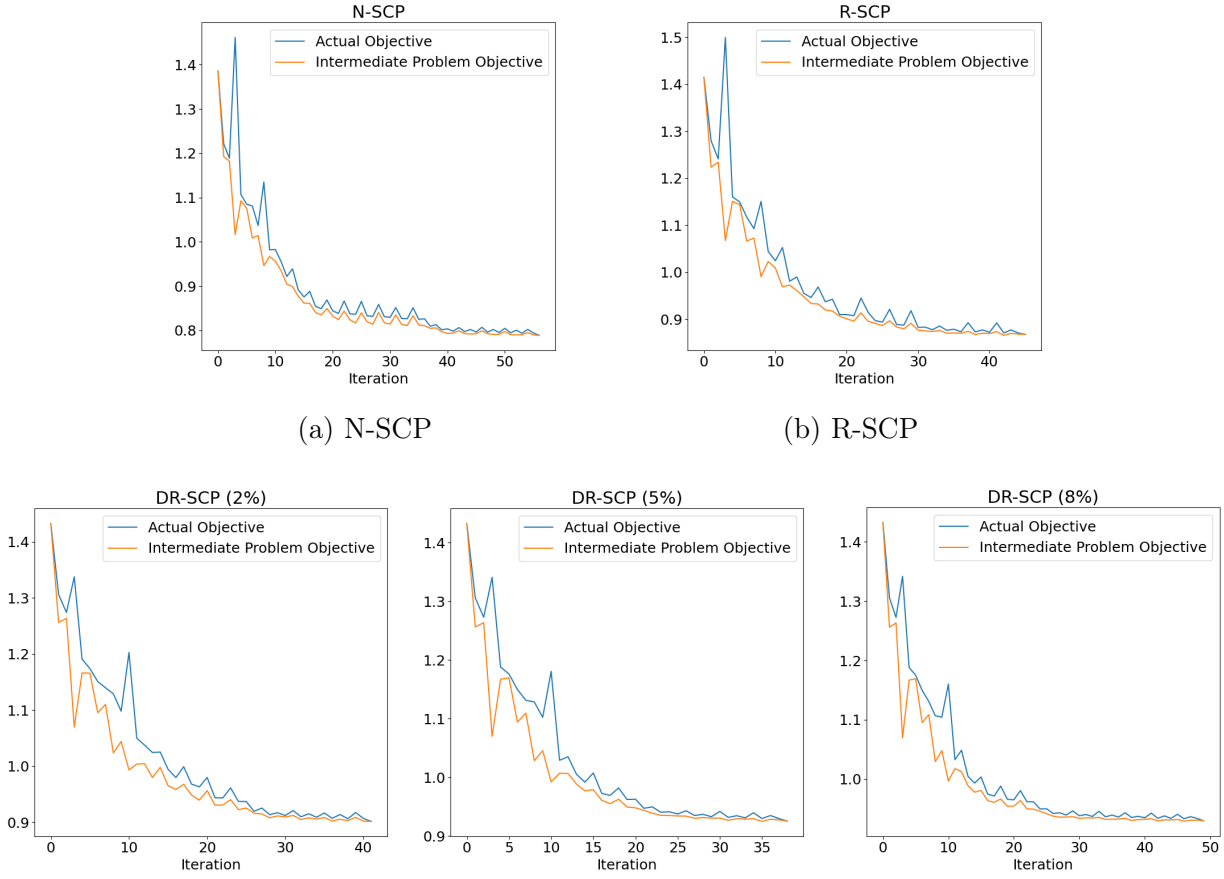
(c) DRO ground-truth and approximation.

Figure 5.14: Comparison of the dose distributions resulting from the ground-truth and the convex approximation models at each level of uncertainty.

is interesting. The robust models converge in less iterations than the deterministic model. However, in the earlier stages, the deterministic model gets the functions much closer; the increased iterations near the end are to get within the tolerated gap. In contrast, at the

	Deterministic		Robust		Distributionally Robust	
	N-SCP	N-DAO	R-SCP	R-DAO	DR-SCP	DR-DAO
Small Scale	1	399	33	1,808	43	26,497
Clinical Scale	753	x	28,977	x	54,132	OOM

Table 5.9: Solution times in seconds for each model at both scales explored. Entries marked with OOM indicate an out-of-memory error when attempting to solve the model with 512 GB of RAM.



(a) N-SCP (b) R-SCP

(c) DR-SCP (2%) (d) DR-SCP (5%) (e) DR-SCP (8%)

Figure 5.15: Convergence behavior for N-SCP, R-SCP, and DR-SCP variants.

earlier iterations, the robust and distributionally robust models struggle. However, once they improve, converging is "easier". That said, the time spent in each iteration is longer with the robust models.

We demonstrated empirically that the accuracy of the convex approximation extends to robust and distributionally robust variants. Finally, we showed the decrease in runtime caused by the convex approximation and observed its potential for clinical use in robust and distributionally robust settings. Note that solving the VMAT problem exactly at a clinical scale remains an open problem.

Chapter 6

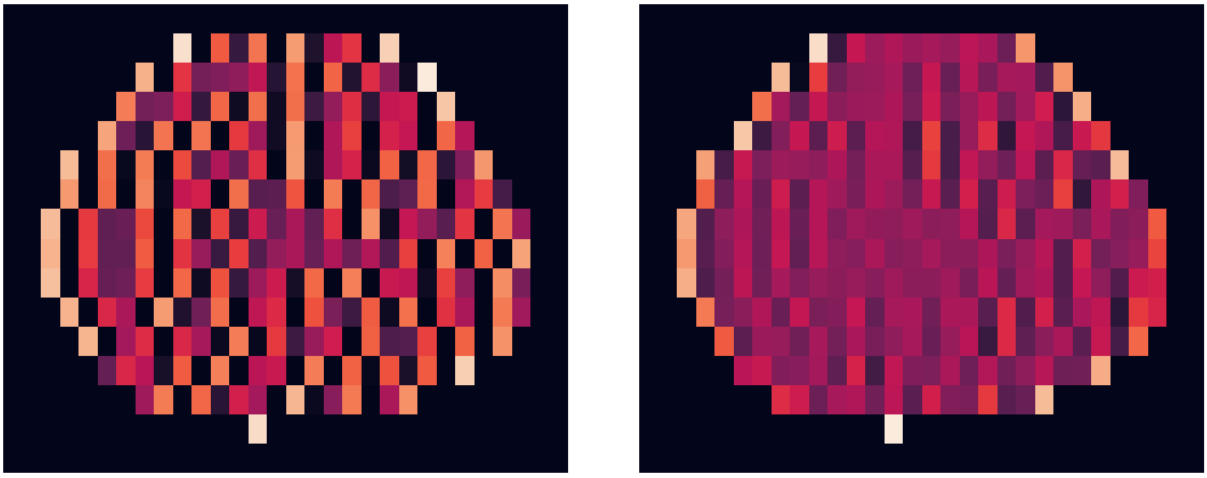
Extensions

This thesis is subject to the following limitations. First, we only consider the FMO model individually, without consideration for the rest of the pipeline. This method is unrealistic in settings that rely on FMO to create RT treatment plans. Second, we do not enforce similarity between neighboring apertures. This approach is not realistic with respect to delivering a VMAT plan in practice. Third, simplify the scenario simulation and objective function due to technological constraints, which limits our relation to how treatment is derived in practice.

Some of these limitations can be easily addressed conceptually, but require more compute resources to implement in practice. Other limitations require nuanced treatment. This chapter will discuss future research ideas to extend this work. Section 6.1 introduces a multi-objective model which results in smoother fluence maps. Smoothness is intended to preserve accuracy and robustness in downstream LS algorithms. Section 6.2 discusses approaches to consider neighboring aperture similarity within the MIP and SCP frameworks. Finally, Section 6.3 discusses several finer details to incorporate more realism in RT optimization models.

6.1 Smooth Fluence Map Optimization

Recall that FMO controls for each beamlet independently and that our results indicate little to no difference in the resulting fluence maps, which were already highly conformal, when robustness was added. However, recall that the quality of the fluence map degrades once it is converted to a deliverable aperture through LS algorithms. One way to mitigate this while retaining the FMO methodology is by incorporating a penalty term on the difference between neighboring beamlets in the objective function (Spirou et al., 2001; Matuszak et al., 2007; Niyas et al., 2016). This is referred to as "smoothness" and is visualized in Figure 6.1.



(a) Example FMO without smoothness considerations

(b) Example FMO with smoothness considerations

Figure 6.1: Side-by-side comparison of an example fluence map resulting from FMO models that do (left) and do not (right) consider smoothness.

Let $f(\mathbf{x})$ be an arbitrary penalty function of beamlet intensities. Let x_j^R and x_j^L represent the intensity of the right and lower neighbors of beamlet j, respectively, each with associated importance weights w_R and w_L. Also, let w_{sm} be the weight associate with smoothness and w_f be the weight associated with the dose penalty function. Finally, let $\sigma(\mathbf{x})$ be a smoothing function which we can formally define as:

$$\sigma(\mathbf{x}) = w_R \sum_{b\in\mathcal{B}} \sum_{j\in\mathcal{J}_b} (x_j - x_j^R)^2 + w_L \sum_{b\in\mathcal{B}} \sum_{j\in\mathcal{J}_b} (x_j - x_j^L)^2.$$

Here, the first term penalizes the squared difference in intensity between a beamlet and its right neighbor, whereas the second term modifies the squared difference between a beamlet and its lower neighbor. Smoothness can then be incorporated into FMO through multiobjective optimization as shown in (6.1).

$$\min \quad w_f f(\mathbf{x}) + w_{sm}\sigma(\mathbf{x}) \tag{6.1a}$$
$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X} \tag{6.1b}$$

In (6.1), the planner is free to specify any objective function of the dose distribution, such as minimizing the quadratic deviation from the prescription. Further, the planner can specify any required constraints. The only change is the addition of the smoothness function to the objective and using weights to balance the two objectives.

Future work can explore the effects of robust and distributionally robust optimization on fluence maps with smoothness. The addition of smoothness results in less tailored maps, which implies that the value of robustness may be more visible in the FMO setting. Further, since LS algorithms degrade the FMO quality, they may also decrease the visible difference that robustness makes. Therefore, it may also be worthwhile to explore how LS algorithms can be modified to be more sensitive to changes in robustness. Given the large number of weights associated with this multiobjective problem, exploration of the Pareto front may reveal interesting insights.

6.2 Neighboring Aperture Similarity

Recall that VMAT is delivered in a continuous arc and that MLC leaves govern the shape of each aperture. Since the arc is discretized, it is importance that neighboring apertures be similar. Otherwise, it is possible that the machine is instructed to deliver an impossible

plan. If that occurs, the gantry will stall and treatment quality will degrade. As such, it is important to consider similar neighboring apertures in the optimization process.

Dursun et al. (2019a) enforce similar neighboring apertures using constraints. Let δ be the number of beamlets a leaf may travel between neighboring control points. Then, the following constraints can be added to a MIP formulation:

$$l_{kb} - l_{k(b+1)} \leq \delta \quad \forall k \in \mathcal{K}_b, b \in \{1, ..., |\mathcal{B}| - 1\}, \tag{6.2a}$$
$$l_{k(b+1)} - l_{kb} \leq \delta \quad \forall k \in \mathcal{K}_b, b \in \{1, ..., |\mathcal{B}| - 1\}, \tag{6.2b}$$
$$r_{kb} - r_{k(b+1)} \leq \delta \quad \forall k \in \mathcal{K}_b, b \in \{1, ..., |\mathcal{B}| - 1\}, \tag{6.2c}$$
$$r_{k(b+1)} - r_{kb} \leq \delta \quad \forall k \in \mathcal{K}_b, b \in \{1, ..., |\mathcal{B}| - 1\}. \tag{6.2d}$$

Here, constraint (6.2a) specifies that for the left leaf in row k, the difference in its position at CP b and CP $b+1$ must be no more than δ. Constraint (6.2b) does the same but for CPs b and $b-1$. Then, constraints (6.2c) and (6.2d) achieve the same for the right leaves.

Aperture similarity constraints limit the feasible region. As such, it would be interesting to see their impact when geometric and distributional uncertainty are considered. In this thesis, we were only able to solve a small-scale example of the full MILP. At this scale, the beamlets were so large that any attempt to limit their motion was redundant. Future work can explore the use of decomposition algorithms to incorporate robustness into the global MIP formulation at a larger scale.

Aperture similarity could also be incorporated within an SCP framework, however the implementation is more nuanced. Recall the convex approximation formulation from (4.1). Note that it does not explicitly control for any leaf positions; the only decision variables are the interior/exterior/boundary beamlet intensities as well as the aperture intensity. One may attempt to bound the boundary intensity between neighboring apertures, but this does not guarantee that the shape will remain similar over subsequent iterations of the algorithm. To encourage similar apertures, Dursun et al. (2023) use multiobjective optimization using the following relationships.

First, recall the relationship between MLC leaf positions and boundary beamlet intensity depicted in Φ. For the left leaf of the first row in CP b, the optimal leaf position can

be found using:

$$l_{b1} = 2\left(1 - \frac{x_{b11}}{a_b}\right).$$

This can be rearranged as:

$$l_{b1} \times a_b = 2(a_b - x_{b11}).$$

Extending this to all rows and also right leaves yields the general relationship:

$$l_{bt} \times a_b = (\min_{j \in \mathcal{J}_{bt}}\{c_j\} - 1) \times a_b + |\mathcal{J}_{bt}| \times (a_b - x_{bt1})$$

and

$$r_{bt} \times a_b = (\min_{j \in \mathcal{J}_{bt}}\{c_j\} - 1) \times a_b + |\mathcal{J}_{bt}| \times x_{bt2}.$$

Note that the right hand side of each of these is convex. As such, we can define "$l_{bt} \times a_b$" as a single variable, and similarly with "$r_{bt} \times a_b$". These can now be incorporated into a penalty function F^{AS} as follows:

$$F^{AS} = \sum_{b=1}^{|\mathcal{B}|-1} \sum_{k=1}^{|\mathcal{T}_b|} \Big(\big((l_{(b+1)k} \times a_{b+1})^2 - (l_{bk} \times a_b)^2\big) + \big((r_{(b+1)k} \times a_{b+1})^2 - (r_{bk} \times r_b)^2\big) \Big).$$

Note that, since these are all defined as single variables, this formulation is convex. The left term minimizes the difference between the left leaf position weighted by the intensity of CP $b+1$ and the left leaf position weighted by the intensity of CP b. The second term does the same for the right leaves. F^{AS} encourages both similar leaf positions as well as similar aperture intensities. However, it is not perfect. For example, one control point could have a small leaf position with a high intensity while the neighboring CP could have a large leaf position with a low intensity, resulting in differing apertures. Nevertheless, this is used to treat patients with VMAT in Memorial Sloan Kettering Cancer Center.

Future research can integrate F^{AS} in the objective function of the SCP optimization problems to investigate the impact of aperture similarity on robustness. Further research should also investigate opportunities to explicitly consider aperture similarity constraints in the model. Note that the relationship between leaf positions and beamlet intensities, Φ, is non-convex. As such, an interesting mathematical problem would be to include it using some iterative scheme or to define convex bounds and assess the performance.

6.3 Towards Clinical Realism

Currently, we simulate inter-fraction motion of the CTV by rigidly shifting the dose-deposition matrix. While this is useful as an approximation, it does not capture the underlying physics of a body's deformable motion and how that affects the radiation it receives. We integrated a PortPy API with an existing RT planning software to generate the rigid shifts and turn them into a compatible format with our optimization models. However, the software was not able to successfully export data even in the smallest cases attempted. Given that the main advantage of SCP is the ability to optimize for clinical scale settings, we pivoted away from this direction. Future research should consider improvements to the speed of the API and the efficiency of the RT planning software. Although seemingly menial, these improvements would unlock many doors to future research projects. For instance, one could take data from any partner cancer clinic, accurately simulate a large number of scenarios, and export it to a compatible format. This would enable a large-scale clinical study on many different sites. It would be very useful to examine how our algorithms affect treatment quality in cases such as prostate cancer, which are prone to geometric uncertainty. It would also be interesting to compare this to more nuanced scenarios such as brain cancer, where a slight overdose could change a patient's life, or even in veterinary care with animals significantly smaller than humans.

Another step towards realism would involve using a quadratic objective on the penalty between the dose prescribed and the dose delivered. We experimented with using a quadratic objective on the SCP model and found, although the problem can be solved hypothetically, it took over a day to complete three iterations. Given that some variants exceed 50 iterations, it is not yet practical to incorporate non-linear objectives. This limits clinical applicability. As such, future research could explore efficient solution schemes for the intermediate problems with robustness, which in this case would admit a large-scale mixed-integer quadratic program (MIQP) formulation.

Due to the rounded ends of the MLC leaves, it is possible for radiation to leak through even when the aperture is closed. This effect is called "leakage" or "tongue and groove", and it implies that, although the planned aperture and delivered aperture may be identical, the resulting dose may still be inaccurate. Salari et al. (2011) assume that leakage is

uncertain and develop a robust optimization approach around it. Steciw (2023) develop an accurate way to measure leakage using clinical treatment planning systems. It would be worthwhile to extend both of these works. For example, by combining robustness to leakage with robustness to the CTV's position and its probability distribution. Alternatively, by incorporating a leakage metric into an objective function. Interestingly, there is some debate between clinicians regarding leakage; some believe it is important while others believe its effect is negligible. This highlights an exciting gap and, as the clinical software improves, it would be interesting to see it become widely integrated within optimization and to have its impact examined.

Chapter 7

Conclusion

In this thesis, we considered the use of mathematical programming to create effective radiation therapy treatment plans robust to uncertainty in the position of the tumor and the underlying probability distribution of its shifts.

Our (useful) novel contributions to this body of knowledge included modeling and algorithmic developments. First, we introduced the robust chance-constrained model presented in An et al. (2017) to VMAT. Second, we applied the distributionally robust CCP derivation from Yanıkoğlu (2019) to radiation therapy in order to account for uncertainty in the distribution of the chance constraint. Finally, we extended the sequential convex programming scheme developed in Dursun et al. (2021) to create treatment plans at a clinical scale which are close to the global optimum. This involved updating the convex approximation problems as well as the feasibility checks to consider robust and distributionally robust optimization.

Overall, we examined three levels of uncertainty: nominal, robust, and distributionally robust, on three treatment modalities: FMO, DAO, and VMAT. When possible, we tested each model on a clinical-size lung patient with simulated one-dimensional rigid shifts and a probability distribution similar to the literature. For the global MIP formulation, only a small scale example could be solved to optimality. Aside from FMO, which was not impacted by robustness, our experiments verified that incorporating geometric robustness can improve target coverage in the case of rigid shifts at a slight expense to the OARs.

Next, we showed that incorporating distributional robustness can further improve target coverage in the case of rigid shifts, resulting in better coverage to out-of-sample shifts. This is important because it showed that even slight errors in estimating the nominal probability distribution necessitate additional robustness. Then, we benchmarked the robust plans against the nominal PTV-based plan which is widely used in practice. This experiment demonstrated that knowledge of the potential shifts of a target can be leveraged to achieve similar target coverage with improved OAR sparing. Finally, we compared the robust SCP models to the robust ground-truth models in a small-scale setting and showed that the convex approximation quality does not degrade.

We also outlined several ideas we began exploring as potential avenues to future research. For example, ways to consider aperture similarity and ways to simulate more realistic motion. These future projects have potential to address the limitations of this study and enable the scaling of our models to more realistic settings.

References

Akartunalı, K., Mak-Hau, V., and Tran, T. (2015). A unified mixed-integer programming model for simultaneous fluence weight and aperture optimization in VMAT, Tomotherapy, and Cyberknife. *Computers & Operations Research*, 56:134–150.

An, Y., Liang, J., Schild, S. E., Bues, M., and Liu, W. (2017). Robust treatment planning with conditional value at risk chance constraints in intensity-modulated proton therapy. *Medical physics*, 44(1):28–36.

Baatar, D., Hamacher, H. W., Ehrgott, M., and Woeginger, G. J. (2005). Decomposition of integer matrices and multileaf collimator sequencing. *Discrete Applied Mathematics*, 152(1-3):6–34.

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, 53(3):464–501.

Bertsimas, D., Nohadani, O., and Teo, K. M. (2010). Nonconvex robust optimization for problems with constraints. *INFORMS journal on computing*, 22(1):44–58.

Bohoslavsky, R., Witte, M. G., Janssen, T. M., and Van Herk, M. (2013). Probabilistic objective functions for margin-less IMRT planning. *Physics in Medicine & Biology*, 58(11):3563.

Bortfeld, T. (2006). IMRT: a review and preview. *Physics in Medicine & Biology*, 51(13):R363.

Brenner, D. R., Gillis, J., Demers, A. A., Ellison, L. F., Billette, J.-M., Zhang, S. X., Liu, J. L., Woods, R. R., Finley, C., Fitzgerald, N., et al. (2024). Projected estimates of cancer in canada in 2024. *Cmaj*, 196(18):E615–E623.

Bzdusek, K., Friberger, H., Eriksson, K., Hårdemark, B., Robinson, D., and Kaus, M. (2009). Development and evaluation of an efficient approach to volumetric arc therapy planning. *Medical physics*, 36(6Part1):2328–2339.

Cameron, C. (2005). Sweeping-window arc therapy: an implementation of rotational IMRT with automatic beam-weight calculation. *Physics in Medicine & Biology*, 50(18):4317.

Canadian Cancer Statistics Advisory Committee in collaboration with the Canadian Cancer Society, S. C. and the Public Health Agency of Canada (2024). Canadian cancer statistics: A 2024 special report on the economic impact of cancer in canada.

Chan, T. C., Mahmoudzadeh, H., and Purdie, T. G. (2014). A robust-CVaR optimization approach with application to breast cancer therapy. *European Journal of Operational Research*, 238(3):876–885.

Charnes, A. and Cooper, W. W. (1959). Chance-constrained programming. *Management science*, 6(1):73–79.

Craft, D., Papp, D., and Unkelbach, J. (2014). Plan averaging for multicriteria navigation of sliding window IMRT and VMAT. *Medical physics*, 41(2):021709.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management science*, 1(3-4):197–206.

Dursun, P., Hong, L., Jhanwar, G., Huang, Q., Zhou, Y., Yang, J., Pham, H., Cervino, L., Moran, J. M., Deasy, J. O., et al. (2023). Automated VMAT treatment planning using sequential convex programming: algorithm development and clinical implementation. *Physics in Medicine & Biology*, 68(15):155006.

Dursun, P., Taşkın, Z. C., and Altınel, İ. K. (2019a). The determination of optimal treatment plans for volumetric modulated arc therapy (VMAT). *European Journal of Operational Research*, 272(1):372–388.

Dursun, P., Taşkın, Z. C., Altınel, I. K., Bilge, H., Kesen, N. D., Okutan, M., and Oral, E. N. (2019b). A column generation heuristic for VMAT planning with adaptive CVaR constraints. *Physics in Medicine & Biology*, 64(20):205024.

Dursun, P., Zarepisheh, M., Jhanwar, G., and Deasy, J. O. (2021). Solving the volumetric modulated arc therapy (VMAT) problem using a sequential convex programming method. *Physics in Medicine & Biology*, 66(8):085004.

Fredriksson, A. (2012). A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization. *Medical physics*, 39(8):5169–5181.

Fredriksson, A., Forsgren, A., and Hårdemark, B. (2011). Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical physics*, 38(3):1672–1684.

Fredriksson, A., Forsgren, A., and Hårdemark, B. (2015). Maximizing the probability of satisfying the clinical goals in radiation therapy treatment planning under setup uncertainty. *Medical Physics*, 42(7):3992–3999.

Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767.

Ho-Nguyen, N., Kilinç-Karzan, F., Küçükyavuz, S., and Lee, D. (2023). Strong formulations for distributionally robust chance-constrained programs with left-hand side uncertainty under Wasserstein ambiguity. *INFORMS Journal on Optimization*, 5(2):211–232.

Jhanwar, G., Tefagh, M., Taasti, V. T., Alam, S. R., Tuomaala, S., Nadeem, S., and Zarepisheh, M. (2023). Portpy: An open-source python package for planning and optimization in radiation therapy including benchmark data and algorithms. *AAPM 65th Annual Meeting & Exhibition*.

Kamath, S., Sahni, S., Li, J., Palta, J., and Ranka, S. (2003). Leaf sequencing algorithms for segmented multileaf collimation. *Physics in Medicine & Biology*, 48(3):307.

Küçükyavuz, S. and Jiang, R. (2022). Chance-constrained optimization under limited distributional information: A review of reformulations based on sampling and distributional robustness. *EURO Journal on Computational Optimization*, 10:100030.

Liu, W., Li, Y., Li, X., Cao, W., and Zhang, X. (2012a). Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques. *Medical physics*, 39(6Part1):3089–3101.

Liu, W., Zhang, X., Li, Y., and Mohan, R. (2012b). Robust optimization of intensity modulated proton therapy. *Medical physics*, 39(2):1079–1091.

Luan, S., Wang, C., Cao, D., Chen, D. Z., Shepard, D. M., and Yu, C. X. (2008). Leaf-sequencing for intensity-modulated arc therapy using graph algorithms. *Medical physics*, 35(1):61–69.

MacFarlane, M., Hoover, D. A., Wong, E., Battista, J. J., and Chen, J. Z. (2020). A fast inverse direct aperture optimization algorithm for volumetric-modulated arc therapy. *Medical physics*, 47(4):1558–1565.

Masi, L., Doro, R., Favuzza, V., Cipressi, S., and Livi, L. (2013). Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy. *Medical physics*, 40(7):071718.

Matuszak, M. M., Larsen, E. W., and Fraass, B. A. (2007). Reduction of IMRT beam complexity through the use of beam modulation penalties in the objective function. *Medical physics*, 34(2):507–520.

Mayo Clinic (2025). Intensity-modulated radiation therapy (IMRT).

National Research Council Canada (2023). Developing new standards for cancer treatments.

Niyas, P., Abdullah, K. K., Noufal, M. P., and Sankaran Nair, T. (2016). Effect of fluence smoothing on the quality of intensity-modulated radiation treatment plans. *Radiological physics and technology*, 9:202–213.

Otto, K. (2008). Volumetric modulated arc therapy: IMRT in a single gantry arc. *Medical physics*, 35(1):310–317.

Papp, D. and Unkelbach, J. (2014). Direct leaf trajectory optimization for volumetric modulated arc therapy planning with sliding window delivery. *Medical physics*, 41(1):011701.

Peng, F., Jia, X., Gu, X., Epelman, M. A., Romeijn, H. E., and Jiang, S. B. (2012). A new column-generation-based algorithm for VMAT treatment plan optimization. *Physics in Medicine & Biology*, 57(14):4569.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471.

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

Salari, E., Men, C., and Romeijn, H. E. (2011). Accounting for the tongue-and-groove effect using a robust direct aperture optimization approach. *Medical physics*, 38(3):1266–1279.

Shepard, D. M., Earl, M. A., Li, X. A., Naqvi, S., and Yu, C. (2002). Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. *Medical physics*, 29(6):1007–1018.

Spirou, S. V., Fournier-Bidoz, N., Yang, J., Chui, C.-S., and Ling, C. C. (2001). Smoothing intensity-modulated beam profiles to improve the efficiency of delivery. *Medical Physics*, 28(10):2105–2112.

Steciw, S. (2023). A convolution-superposition fluence model for the Siemens HD120 multi leaf collimator with application to a 3D VMAT dose engine. *Biomedical Physics & Engineering Express*, 9(6):065004.

Sun, B., Song, J., Zhu, G., and Shi, L. (2013). A two-stage approach for VMAT treatment plan optimization. In *2013 IEEE international conference on automation science and engineering (CASE)*, pages 455–460. IEEE.

Teoh, M., Clark, C. H., Wood, K., Whitaker, S., and Nisbet, A. (2011). Volumetric modulated arc therapy: a review of current literature and clinical use in practice. *The British journal of radiology*, 84(1007):967–996.

Ulrich, S., Nill, S., and Oelfke, U. (2007). Development of an optimization concept for arc-modulated cone beam therapy. *Physics in Medicine & Biology*, 52(14):4099.

Unkelbach, J., Alber, M., Bangert, M., Bokrantz, R., Chan, T. C., Deasy, J. O., Fredriksson, A., Gorissen, B. L., Van Herk, M., Liu, W., et al. (2018). Robust radiotherapy planning. *Physics in Medicine & Biology*, 63(22):22TR02.

Unkelbach, J., Bortfeld, T., Craft, D., Alber, M., Bangert, M., Bokrantz, R., Chen, D., Li, R., Xing, L., Men, C., et al. (2015). Optimization approaches to volumetric modulated arc therapy planning. *Medical physics*, 42(3):1367–1377.

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

Uryasev, S. and Rockafellar, R. T. (2001). Conditional value-at-risk: optimization approach. *Stochastic optimization: algorithms and applications*, pages 411–435.

Wada, T., Kawahara, D., Murakami, Y., Nakashima, T., and Nagata, Y. (2022). Robust optimization of VMAT for prostate cancer accounting for geometric uncertainty. *Journal of Applied Clinical Medical Physics*, 23(9):e13738.

Wang, C., Luan, S., Tang, G., Chen, D. Z., Earl, M. A., and Yu, C. X. (2008). Arc-modulated radiation therapy (AMRT): a single-arc form of intensity-modulated arc therapy. *Physics in Medicine & Biology*, 53(22):6291.

Wang, Y., Liu, H., Yang, Y., and Lu, B. (2022). A practical algorithm for VMAT optimization using column generation techniques. *Medical Physics*, 49(7):4335–4352.

Witte, M. G., Van Der Geer, J., Schneider, C., Lebesque, J. V., Alber, M., and Van Herk, M. (2007). IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP. *Medical Physics*, 34(9):3544–3555.

World Health Organization and International Agency for Research on Cancer (2024). Global cancer burden growing, amidst mounting need for services. Accessed: 2025-05-11.

Xie, W. (2021). On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, 186(1):115–155.

Yahalom, J., Illidge, T., Specht, L., Hoppe, R. T., Li, Y.-X., Tsang, R., Wirth, A., Group, I. L. R. O., et al. (2015). Modern radiation therapy for extranodal lymphomas: field and dose guidelines from the international lymphoma radiation oncology group. *International Journal of Radiation Oncology* Biology* Physics*, 92(1):11–31.

Yanıkoğlu, İ. (2019). Robust reformulations of ambiguous chance constraints with discrete probability distributions. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)*, 9(2):236–252.

Yu, C. X. (1995). Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Physics in Medicine & Biology*, 40(9):1435.

Zhang, P., Happersett, L., Hunt, M., Jackson, A., Zelefsky, M., and Mageras, G. (2010). Volumetric modulated arc therapy: planning and evaluation for prostate cancer cases. *International Journal of Radiation Oncology* Biology* Physics*, 76(5):1456–1462.

APPENDICES

Appendix A

Notations

A.1 Sets

Set	Definition
Υ	Set of OARs
$\mathcal{H}_\upsilon$	Set of voxels in OAR $\upsilon \in \Upsilon$
$\mathcal{T}$	Set of tumor voxels
$\mathcal{B}$	Set of control points (also referred to as "beams")
$\mathcal{J}$	Set of beamlets, where $\mathcal{J}_{bk}$ represents the beamlets in row k of control point b
$\mathcal{K}_b$	Set of row indices for control point $b \in \mathcal{B}$
$\mathcal{S}$	Set of scenarios for positions of the target
$\mathcal{P}$	Ambiguity set containing probability distributions, where $\mathbb{Q} \in \mathcal{P}$ defines possible probabilities for each $s \in S$ being realized

Table A.1: Notations used for sets.

A.2 Parameters

Parameter	Definition
k_j	Row index of beamlet $j \in \mathcal{J}$
c_j	Column index of beamlet $j \in \mathcal{J}$
$\mathbf{D}$	Dose-deposition matrix, where $\mathbf{d}_i'$ is its i^{th} row, and d_{ij} is the dose from beamlet $j \in \mathcal{J}$ to voxel $i \in \mathcal{V}$ at unit intensity
$\hat{\mathbf{D}}$	Adjusted dose-deposition matrix for interior, exterior, and boundary beamlets, where $'_i$ is its i^{th} row, and $\hat{d}_{ij}$ is the dose from beamlet $j \in \mathcal{J}$ to voxel $i \in \mathcal{V}$ at unit intensity
$\bar{d}$	Prescribed dose to the target
$\underline{d}$	Desired lower bound for mean tail of target dose distribution
β	Risk tolerance for conditional value at risk
γ	Targeted fraction of the prescribed dose
U	Upper bound intensity of each aperture
N	Width of aperture rows
$\mathbf{p}$	Nominal probability distribution $\mathbf{p} = (p^s)_{s \in \mathcal{S}}$, where p^s represents the probability of scenario $s \in \mathcal{S}$ being realized
M	An arbitrarily large constant
ϵ	Violation tolerance for the chance constraint
$\mathbf{e}$	Vector of all ones
$\underline{\mathbf{p}}$	Vector of lower bounds for the error on the estimate of the nominal probability distribution
$\overline{\mathbf{p}}$	Vector of upper bounds for the error on the estimate of the nominal probability distribution

Table A.2: Notations used for parameters.

A.3 Variables

Variable	Definition
$\mathbf{x}$	Vector of beamlet intensities
ρ	Auxiliary variable for CVaR approximation
$\mathbf{a}$	Vector of aperture intensities
$\mathbf{l}$	Vector of positions for left MLC leaves, where l_{bk} represents the position of the left leaf in row k of control point b corresponding to beamlet $j \in \mathcal{J}$
$\mathbf{r}$	Vector of positions for right MLC leaves, where r_{bk} represents the position of the right leaf in row k of control point b corresponding to beamlet $j \in \mathcal{J}$
$\mathbf{z}$	Binary vector where $z^s = 1$ if scenario $s \in \mathcal{S}$ is protected, and 0 otherwise
$\hat{\mathbf{p}}$	Vector of errors for the nominal probability distribution
λ	Dual variable for the valid probability distribution sum constraint
$\boldsymbol{\eta}$	Dual variable for the upper bound constraint on $\hat{\mathbf{p}}$
$\boldsymbol{\theta}$	Dual variable for the lower bound constraint on $\hat{\mathbf{p}}$
ψ_i^{over}	Auxiliary variable for the overdose on voxel $i \in \mathcal{T}$
ψ_i^{under}	Auxiliary variable for the underdose on voxel $i \in \mathcal{T}$
ω_i	Auxiliary variable to linearize CVaR constraint for the underdose to voxel $i \in \mathcal{T}$

Table A.3: Notations used for decision variables.