Title
Deterministic Direct Aperture Optimization Using Multiphase Piecewise Constant Segmentation

Author
Nguyen, Dan Minh

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Deterministic Direct Aperture Optimization Using Multiphase Piecewise Constant Segmentation

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biomedical Physics

by

Dan Minh Nguyen

2017

ABSTRACT OF THE DISSERTATION

Deterministic Direct Aperture Optimization Using Multiphase Piecewise Constant Segmentation

by

Dan Minh Nguyen

Doctor of Philosophy in Biomedical Physics

University of California, Los Angeles, 2017

Professor Ke Sheng, Chair

Purpose

Direct Aperture Optimization (DAO) attempts to incorporate machine constraints in the inverse optimization to eliminate the post-processing steps in fluence map optimization (FMO) that degrade plan quality. Current commercial DAO methods utilize a stochastic or greedy approach to search a small aperture solution space. In this study, we propose a novel deterministic direct aperture optimization that integrates the segmentation of fluence map in the optimization problem using the multiphase piecewise constant Mumford-Shah formulation.

Methods

The deterministic DAO problem was formulated to include an L2-norm dose fidelity term to penalize differences between the projected dose and the prescribed dose, an anisotropic total variation term to promote piecewise continuity in the fluence maps, and the multiphase piecewise constant Mumford-Shah function to partition the fluence into pairwise discrete segments. A proximal-class, first-order primal-dual solver was implemented to solve the large scale optimization problem, and an alternating module strategy was implemented to update fluence and delivery segments. Three patients of varying complexity-one glioblastoma multiforme (GBM) patient, one lung (LNG) patient, and one bilateral head and neck (H&N) patient with 3 PTVs-were selected to test the new DAO method. For comparison, a popular and successful approach to DAO known as simulated annealing-a stochastic approach-was replicated. Each patient was planned using the Mumford-Shah based DAO (DAO_{MS}) and the simulated annealing based DAO (DAO_{SA}). PTV coverage, PTV homogeneity (D95/D5), and OAR sparing were assessed for each plan. In addition, high dose spillage, defined as the 50% isodose volume divided by the tumor volume, as well as conformity, defined as the van't Riet conformation number, were evaluated.

Results

DAO_{MS} achieved essentially the same OAR doses compared with the DAO_{SA} plans for the GBM case. The average difference of OAR Dmax and Dmean between the two plans were within 0.05% of the plan prescription dose. The lung case showed slightly improved critical structure sparing using the DAO_{MS} approach, where the average OAR Dmax and Dmean were reduced by 3.67% and 1.08%, respectively, of the prescription dose. The DAO_{MS} plan substantially reduced

OAR Dmax and Dmean by over 10% of the prescription dose for the H&N patient. The DAO_{MS} plan substantially improved the H&N PTV coverage, increasing D99 by 6.98% of the prescription dose. For the GBM and LNG patients, the DAO_{MS} and DAO_{SA} plans had comparable high dose spillage but slightly worse conformity with the DAO_{MS} approach. For the H&N plan, DAO_{MS} was considerably superior in high dose spillage and conformity to the DAO_{SA}. The deterministic approach can solve the DAO problem 9.5 to 40 times faster than the simulated annealing approach.

Conclusions

A novel deterministic direct aperture optimization formulation was developed and evaluated. It combines fluence map optimization and the multiphase piecewise constant Mumford-Shah segmentation into a unified framework, and the resulting optimization problem can be solved efficiently. Compared to the widely and commercially used simulated annealing DAO approach, it showed comparable dosimetry behavior for simple plans, and substantially improved OAR sparing, PTV coverage, PTV homogeneity, high dose spillage, and conformity for the more complex H&N plan.

The dissertation of Dan Minh Nguyen is approved.

Dan Ruan

Daniel A. Low

Lieven Vandenberghe

Ke Sheng, Committee Chair

University of California, Los Angeles

2017

For Tiffany, Lily, Mom, and Dad

"Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution."

~Einstein, Albert. *Cosmic religion: With other opinions and aphorisms*. Covici-Friede, 1931.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

CLN

Clinical 45, 46, 47, 48, 52, 53, 54, 55

Co-60

Cobalt-60 33, 34, 35, 41, 43, 55, 56, 57, 156, 158

CT

Computed Tomography 1

CVSP

Convolution/Superposition 12

DAO

Direct Aperture Optimization ii, iii, iv, 4, 5, 6, 32, 33, 57, 58, 91, 93, 114, 115, 117, 118, 119, 129, 131, 135, 144, 147

DDR

Dose Domain Regularization 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 31, 32, 35, 41, 42, 79, 80

DMPO

Direct Machine Parameter Optimization 32

DMS

Direct MLC Segmentation 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 26, 27, 28, 66, 67, 69, 70, 71, 76, 77, 78, 79, 80, 84, 85, 86, 87

DVH

Dose Volume Histogram 19, 25, 26, 76, 80, 135, 137

FMO

Fluence Map Optimization ii, 4, 56, 117, 125

FROST

Fluence Regularization and Optimized Selection of Triplets 35, 36, 43, 45, 46, 47, 48, 49, 52, 53, 54, 55, 57

GBM

Glioblastoma Multiforme iii, iv, 12, 13, 14, 16, 17, 20, 21, 22, 23, 24, 25, 26, 28, 65, 76, 78, 83, 85, 86, 99, 101, 102, 103, 104, 107, 108, 109, 110, 130, 135, 136, 137, 139, 140, 141, 142, 143

H&N

Head and Neck iii, iv, 12, 23, 28, 46, 52, 65, 76, 78, 83, 85, 86, 102, 107, 109, 130, 132, 135, 136, 137, 139, 140, 141, 143

IMRT

Intensity Modulated Radiation Therapy 2, 4, 5, 6, 7, 12, 27, 33, 34, 57, 58, 59, 60, 67, 87, 88, 89, 90, 91, 92, 110, 112, 113, 114, 116, 117, 118, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161

KKT

Karush-Kuhn-Tucker 39, 40, 56

LNG

Lung iii, iv, 12, 23, 28, 43, 45, 47, 48, 52, 53, 54, 65, 76, 83, 85, 86, 99, 107, 108, 109, 110, 130, 135, 136, 137, 139, 140, 141, 143

MLC

Multileaf Collimator 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 23, 27, 31, 32, 34, 35, 36, 41, 42, 43, 44, 49, 55, 56, 57, 58, 59, 66, 69, 70, 71, 79, 87, 89, 90, 92, 93, 94, 103, 104, 113, 114, 117, 118, 123, 124, 125, 141, 144, 145, 146, 154, 155, 156, 157

MRI

Magnetic Resonance Imaging 33, 34, 43, 55, 156, 157, 158

OAR

Organs at Risk iii, iv, 7, 14, 17, 31, 36, 37, 38, 46, 67, 100, 109, 115, 132, 137

PTV

Planning Target Volume iii, iv, 7, 12, 13, 14, 16, 17, 20, 21, 22, 23, 24, 25, 28, 29, 33, 36, 37, 38, 44, 45, 46, 48, 52, 53, 54, 62, 65, 66, 67, 76, 77, 79, 80, 81, 88, 94, 99, 100, 107, 108, 109, 115, 123, 131, 132, 135, 136, 137

RAO

Rectangular Aperture Optimization 66, 67, 70, 71, 76, 77, 78, 79, 80, 84, 85, 86, 87, 88, 90

SBRT

Stereotactic Body Radiation Therapy 65, 154, 157

SOC

Sparse Orthogonal Collimator 59, 60, 69, 86, 87, 89, 90, 91

TSP

Traveling Salesman Problem 68, 69

TV

Total Variation 30, 95, 112, 113

UPD

Unprocessed Dose 14, 22, 23, 27, 28

VMAT

Volumetric Modulated Arc Therapy 65, 91, 92, 93, 94, 100, 111, 112, 113, 114, 115, 154, 160, 161

LIST OF APPENDICES

ACKNOWLEDGMENTS

When I first entered UCLA as a graduate student, I was excited, anxious, and optimistic about starting a new chapter in my life learning about medical physics. I would soon realize my ideas of the field to be an extremely naïve and limited view, but that only motivated me more as I understood that there was so much to learn. Reaching the final destination of my graduate journey would have been impossible without the guidance and support of my advisor, Ke Sheng. He has been the best mentor any student could ask for, providing wisdom and direction for all of my research projects and dissertation work.

I would also like to thank Dan Ruan, Daniel A. Low, and Lieven Vandenberghe for all of their help and support during my graduate career and as my dissertation committee. They have provided valuable insight and knowledge the work I have done over the years.

Thank you Victoria, Angelia, Kaley, Qihui, Ryan, Wenbo, and Elizabeth for all of your help and support during my graduate years. It's been fun! Special thanks to Daniel O'Connor. My graduate career would've have been very different if I had not met Daniel. Thank you for teaching me all the optimization knowledge that I know.

I would like to thank my sister, Lily, my mother, Kim, and my father, Huy, for being there all my life, supporting me and enabling for me to attain my dreams. Finally, I would like to thank Tiffany, whose love, encouragement, and support have made it possible for me to reach the finish line!

VITA

EDUCATION

Ph.D. in Biomedical Physics	University of California Los Angeles	Spring 2017
M.S. in Biomedical Physics	University of California Los Angeles	Spring 2015
B.S. in Physics	The University of Texas at Austin	Spring 2012

PUBLICATIONS

First Author Publications

1) Nguyen, D., Dong, P., Long, T., Ruan, D., Low, D. A., Romeijn, E., & Sheng, K. (2014). Integral dose investigation of non-coplanar treatment beam geometries in radiotherapy. *Medical physics*, *41*(1).

2) Nguyen, D., Rwigema, J. C. M., Victoria, Y. Y., Kaprealian, T., Kupelian, P., Selch, M., ... & Sheng, K. (2014). Feasibility of extreme dose escalation for glioblastoma multiforme using 4π radiotherapy. *Radiation Oncology*, *9*(1), 239.

3) Nguyen, D., O'Connor, D., Yu, V. Y., Ruan, D., Cao, M., Low, D. A., & Sheng, K. (2015). Dose domain regularization of MLC leaf patterns for highly complex IMRT plans. *Medical physics*, *42*(4), 1858-1870.

4) Nguyen, D., Ruan, D., O'Connor, D., Woods, K., Low, D. A., Boucher, S., & Sheng, K. (2016). A novel software and conceptual design of the hardware platform for intensity modulated radiation therapy. *Medical physics*, *43*(2), 917-929.

5) Nguyen, D., Lyu, Q., Ruan, D., O'Connor, D., Low, D. A., & Sheng, K. (2016). A comprehensive formulation for volumetric modulated arc therapy planning. *Medical Physics*, *43*(7), 4263-4272.

6) Nguyen, D., Thomas, D., Cao, M., O'Connor, D., Lamb, J., & Sheng, K. (2016). Computerized triplet beam orientation optimization for MRI-guided Co-60 radiotherapy. *Medical Physics*, *43*(10), 5667-5675.

Co-author Publications

1) Dong, P., Yu, V., Nguyen, D., Demarco, J., Woods, K., Boucher, S., ... & Sheng, K. (2014). Feasibility of using intermediate x-ray energies for highly conformal extracranial radiotherapy. *Medical physics*, *41*(4).

2) Dong, P., Nguyen, D., Ruan, D., King, C., Long, T., Romeijn, E., ... & Sheng, K. (2014). Feasibility of prostate robotic radiation therapy on conventional C-arm linacs. *Practical radiation oncology*, *4*(4), 254-260.

3) Rwigema, J. C. M., Nguyen, D., Heron, D. E., Chen, A. M., Lee, P., Wang, P. C., ... & Steinberg, M. L. (2015). 4π noncoplanar stereotactic body radiation therapy for head-and-neck cancer: Potential to improve tumor control and late toxicity. *International Journal of Radiation Oncology* Biology* Physics*, *91*(2), 401-409.

4) Victoria, Y. Y., Nguyen, D., Pajonk, F., Kupelian, P., Kaprealian, T., Selch, M., ... & Sheng, K. (2015). Incorporating cancer stem cells in radiation therapy treatment response modeling and the implication in glioblastoma multiforme treatment resistance. *International Journal of Radiation Oncology* Biology* Physics*, *91*(4), 866-875.

5) Yu, V. Y., Tran, A., Nguyen, D., Cao, M., Ruan, D., Low, D. A., & Sheng, K. (2015). The development and verification of a highly accurate collision prediction model for automated noncoplanar plan delivery. *Medical physics*, *42*(11), 6457-6467.

6) Woods, K., Nguyen, D., Tran, A., Victoria, Y. Y., Cao, M., Niu, T., ... & Sheng, K. (2016). Viability of Noncoplanar VMAT for liver SBRT compared with coplanar VMAT and beam orientation optimized 4π IMRT. *Advances in radiation oncology*, *1*(1), 67-75.

7) Yu, V. Y., Tran, A., Nguyen, D., Cao, M., Ruan, D., Low, D. A., & Sheng, K. (2015). The development and verification of a highly accurate collision prediction model for automated noncoplanar plan delivery. *Medical physics*, *42*(11), 6457-6467.

8) Tran, A., Zhang, J., Woods, K., Yu, V., Nguyen, D., Gustafson, G., ... & Sheng, K. (2017). Treatment planning comparison of IMPT, VMAT and 4π radiotherapy for prostate cases. *Radiation Oncology*, *12*(1), 10.

1 INTRODUCTION

1.1 Radiation Therapy and Fluence Map Optimization

Since the discovery of x-rays in 1895[1], radiation has had an intimate role in the world of medicine for treating tumors. Radiation therapy techniques have evolved at an astounding rate, becoming one of the most common modalities of cancer treatments today. In modern external beam radiation therapy, the workflow can be divided into four stages: CT/MR simulation, tumor and critical structure segmentation, treatment planning, and treatment delivery. The success of each category is crucial to the efficacy of radiation therapy, and poor performance in any of these will result in inadequate patient outcome. The series of studies that will be presented are focused on improving the treatment planning aspect for external beam radiation therapy. Particularly, the emphasis will be in optimization formulations and algorithms for more effective and efficient delivery of the radiation to the patient.

Intensity modulated radiation therapy (IMRT) was theorized in the 1980s[2] and subsequently developed in several landmark papers[3-5]. Since then, IMRT has been widely accepted into radiation therapy clinics as the staple approach to radiotherapy. Modern radiotherapy equipment for IMRT typically consist of a gantry housing a linear accelerator (linac) and equipped with a multileaf collimator (MLC). An example radiation therapy station is shown in Figure 1.1.

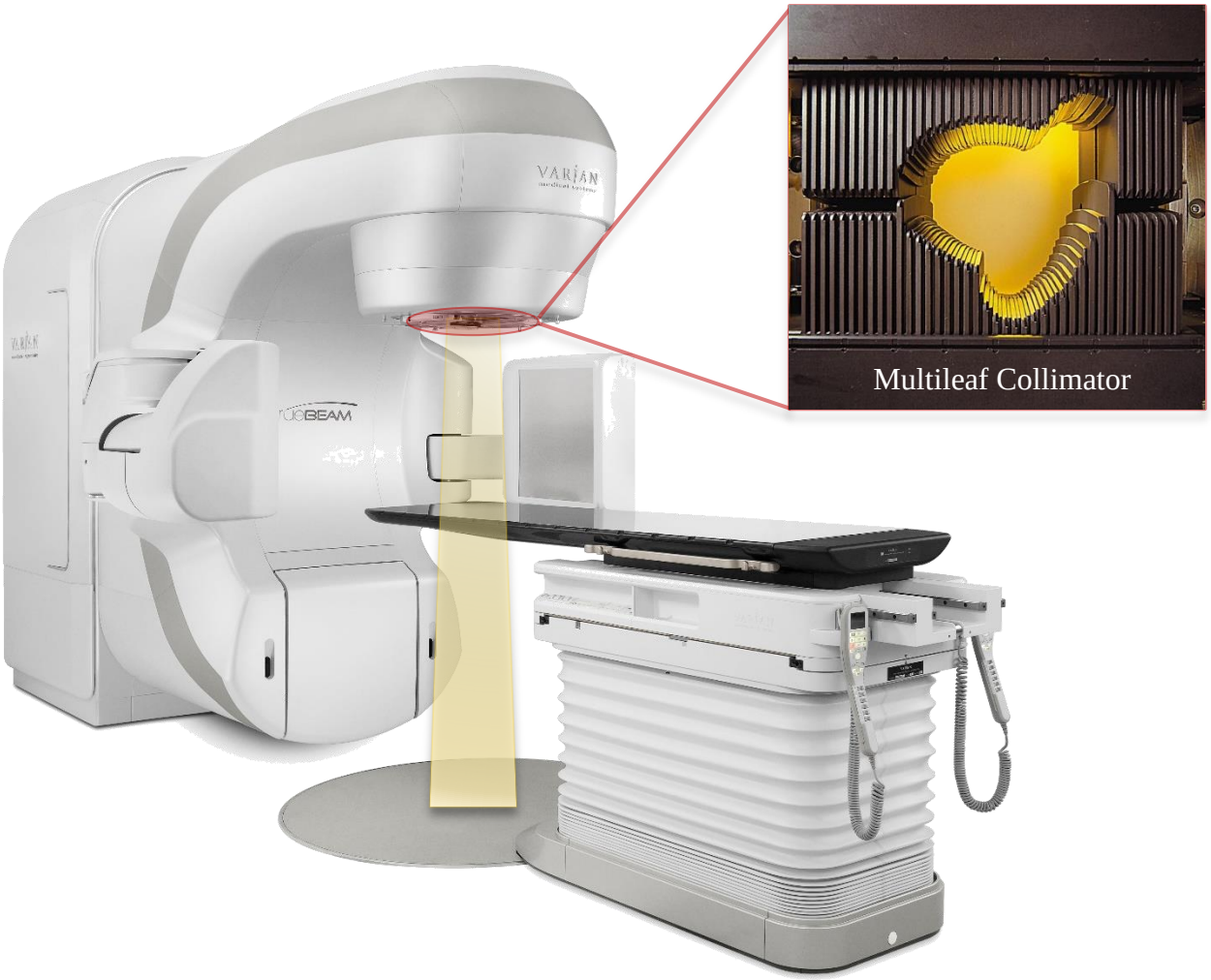


Figure 1.1: C-arm gantry with multileaf collimators and patient couch. Radiation beam represented in yellow. Image courtesy of Varian Medical Systems, Inc. All rights reserved.

The linac accelerates electrons to a specific energy, typically 6 MeV. These electrons then collide into a tungsten target, producing photons with an energy spread ranging from 0 to the electron energy. The MLCs are then responsible for modulating the radiation field to have a desired distribution, which is achieved by the MLCs forming apertures of various shapes, and allowing the radiation to pass through these apertures at a specified fluence rate and time. Complicated fluence maps are broken down into several delivery segments, where, for each deliverable segment, the MLCs form one or more apertures and the beam is turned on. Figure 1.2 shows a basic schematic of the intensity modulation done by the MLCs on the radiation field.

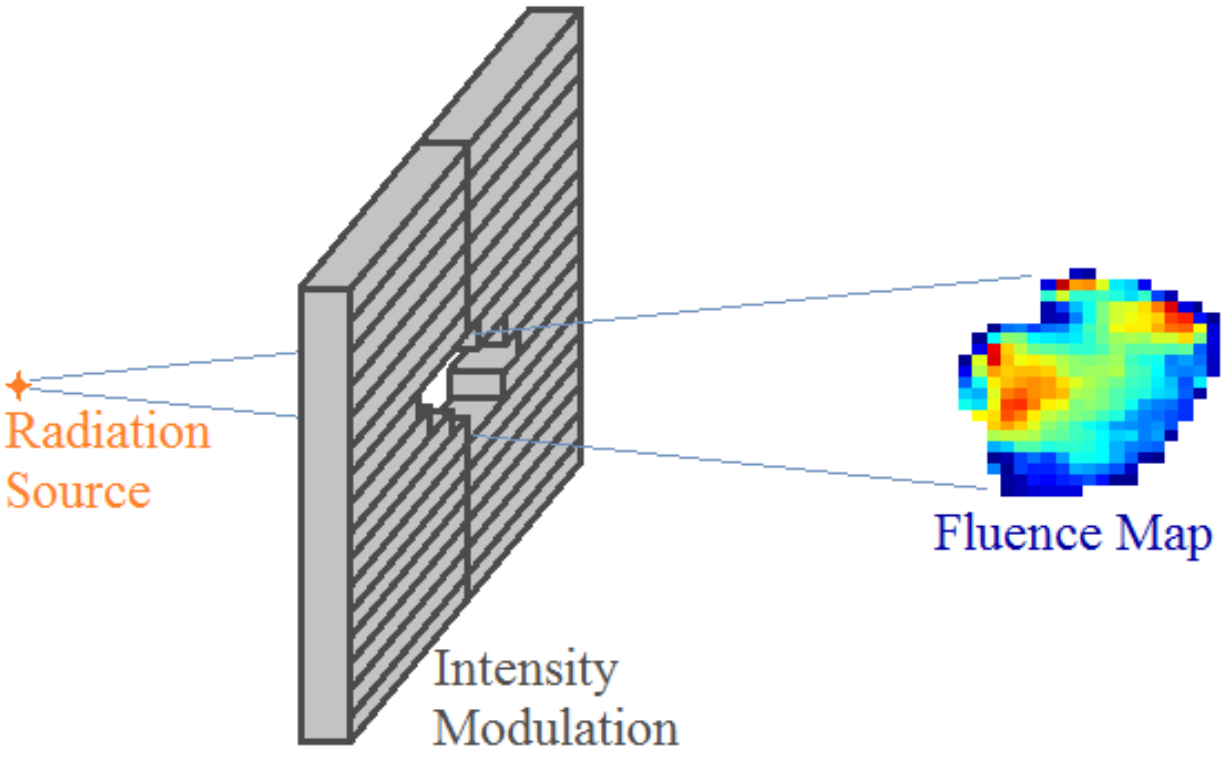


Figure 1.2: Schematic of the intensity modulation process of the radiation field using the multileaf collimator. For linacs, the radiation source is a tungsten target, which is collided with electrons to produce photons.

While IMRT has been proved exceptional in controlling dose distributions, one major weakness lies in the fluence map optimization (FMO), which does not consider machine constraints in the inverse optimization. MLCs are physically limited in how they can modulate the radiation field. The FMO problem does not take this physical limitation into account, producing fluence maps that the MLCs cannot deliver. The result is that these optimized fluence maps then need to be approximated into a set of MLC deliverable segments. While several clever sequencing methods have been developed, such as sliding window or the reducing level method[6], this additional post processing step separate from initial FMO inverse optimization leads to varying degrees of degradation in the dose quality.

1.2 Direct Aperture Optimization

Direct Aperture Optimization (DAO) was invented to solve the problem by incorporating machine constraints in the optimization to eliminate the post-processing step that follows FMO. One significant challenge with DAO is that the number of possible apertures as the combination of beamlets is mathematically intractable. To manage the challenge, stochastic and greedy approaches have been implemented. The stochastic DAO method utilizes a simulated annealing process[7] to iteratively update the aperture shapes and intensity values[8-14]. In the greedy approach, DAO plans are created by optimizing a predetermined library of apertures, and a column generation method to expand the aperture library until an acceptable treatment plan is acquired[15-17]. These DAO methods result in plans that are directly deliverable by the machine without the need for a separate MLC segmentation step but there are significant limitations. Due to the

computational cost, both methods can only search a small fraction of the possible apertures, which may be insufficient for complex IMRT plans.

We aim to develop a new deterministic DAO method that is not limited to a narrow range of apertures in the optimization. Instead, the MLC segmentation problem is formulated as a piecewise constant segmentation problem that can be efficiently solved using methods such as Mumford-Shah[18-21]. Piecewise-constant segmentation solves an optimization problem that partitions a scaler image into piecewise constant regions that the original image. Following the original Mumford-Shah formulation, recent developments have led to multiphase piecewise-constant segmentations[19,20,22-25]. Pertinent to the IMRT optimization problem, because the piecewise-constant segmented regions are arbitrarily shaped, it provides a mathematically complete description of any possible MLC aperture for a specific fluence map. We integrate the multiphase piecewise-constant Mumford-Shah function with the fluence map optimization problem into a multiconvex formulation[26,27]—a non-convex problem that can be solved by alternating among convex modules. We then test the feasibility of this optimization on several patient cases and compare its performance to the DAO method using simulated annealing.

In this dissertation, four essential preliminary studies that were done during my graduate studies will be presented, followed by the deterministic DAO with multiphase piecewise-constant Mumford-Shah study. The four preliminary studies each played a significant role in the development and expansion of the ideas that eventually lead to the creation of the deterministic DAO approach.

2 PRELIMINARY STUDIES

Four key studies lay the groundwork to the ideation and development of the deterministic direct aperture optimization approach. For each preliminary study, the following sections will be presented: 1) Introduction, 2) Methods, 3) Results, 4) Discussion, and 5) Contribution to Deterministic DAO approach. The first four sections are specific to the preliminary study, and the last section will tie in its significance and influence to the deterministic DAO approach.

2.1 Dose domain regularization of MLC leaf patterns for highly complex IMRT plans[28]

2.1.1 Introduction

Recent development in automated beam orientation optimization has shown the feasibility of further improving IMRT plan dose conformality[17,29,30]. The integrated beam orientation and fluence optimization technique was particularly suited to selecting beams in a large solution

space such as the non-coplanar solution space, where the ability of manual beam orientation selection was limited. These new planning methods have been shown to reduce organs at risk (OAR) dose while maintaining high dose coverage to the planning target volume (PTV)[31-34]. For radioresistant tumors, the markedly increased dose conformality could be used for safe dose escalation and potentially improved tumor control probability. On the other hand, more than 20 highly modulated fields have been generally needed to fully take advantage of the non-coplanar platform[31-34]. This had created a practical challenge in delivery because the number of complex fluences that needed to be converted to deliverable MLC segments substantially increased.

In theory, the optimized fluence could be precisely translated into MLC motion if the number of MLC segments was unrestricted, but this approach would have resulted in an impractical number of MLC segments. Large numbers of MLC segments increased not only delivery time, but also dosimetric errors due to the increase of small MLC segments[35] and accumulation of small field dosimetry error, and increased leakage doses[36], which were proportional to the total monitor units used in the plan. On the other hand, fluence map and subsequent MLC segment simplification inevitably degraded dose distribution from the plan optimized dose and resulted in undesired hot and cold spots in the PTV that often required mitigation such as plan renormalization, which worsened the normal tissue doses.

Fluence map simplification to reduce the complexity of MLC segments had been previously researched for step-and-shoot IMRT delivery based on three main approaches. The first was to reduce the fluence map complexity, typically represented by the number of discretized intensity levels, while minimizing the difference between the original fluence map and the simplified fluence map[6,37-40]. The second approach was to include the fluence smoothness in the optimization objective function[41,42]. In the third approach, the hardware limitations, such as the

number of MLC segments, were included in plan optimization as constraints[43-48]. All approaches have shown improvement in plan deliverability. However, both of these approaches have limitations that are intensified in the automated beam orientation optimization scenario. The first approach did not promise minimization of the difference between the delivered dose and optimized dose for a given MLC segment number[49]. Since there was no mechanism to compensate the errors in one fluence map of a beam by another beam, the approximation of using fluence as the surrogate of the optimization problem tended to cumulate dose delivery errors with increasing number of fields. Although plan quality degradation, caused by generic fluence smoothness, can be mitigated by approaches including minimal surface smoothing constraints[50] and a method to constrain the smoothing based on the cost function gradients[51], the objective function is still not transparent to the dose objectives. In the second approach, the modified objective function does not guarantee a certain level of dose fidelity or fluence smoothness. The relative weights of the two terms are unknown for each individual patient, requiring a trial and error process to determine. The process can be time consuming, unpredictable and irreproducible. The third approach results in either significantly increased optimization units (apertures vs. beamlets) or rendering the objective function non-convex. Due to the enormous size of automated beam orientation and fluence optimization problem, a mathematically tractable solution of automated beam orientation and optimization problem using the third approach is currently unavailable.

A third approach to solve the problem is by directly optimizing the MLC segments based on the dose domain objective function. This is performed after solving for the optimal dose distribution from the original plan optimization. A sub-problem to optimize the number of fluence intensity

steps had been previously proposed by Suss et al[49]. The full fluence-and-dose-domain problem was **N**on-deterministic **P**olynomial-time (NP)-hard, *i.e.*, the running time to solve this problem is not upper bounded by a polynomial expression in the size of the input for the algorithm. The authors proposed methods to solve a partition problem that provided an upper error bound for a simplified equal level stratification problem. The approach, on the other hand, did not minimize the number of segments and dosimetric deviation. In this study we reformulate the full problem to be a convex minimization problem and then implement an efficient primal-dual algorithm to iteratively solve it.

2.1.2 Methods

The overall MLC segment optimization problem was performed in two successive steps. In the first step, the optimized fluence maps were regularized for piecewise smoothness while minimizing the dosimetric changes. In the second step, the regularized fluence was stratified and converted to MLC segments.

2.1.2.1 Optimization Formulation

In the first step, the complexity of the fluence map was reduced by minimizing a convex objective function subject to a non-negativity constraint:

$$\underset{\boldsymbol{x}}{\boldsymbol{argmin}} \quad \frac{\mathbf{1}}{\mathbf{2}}\|\boldsymbol{Ax}-\boldsymbol{d}\|_{\mathbf{2}}^{\mathbf{2}}+\boldsymbol{\lambda}\|\boldsymbol{D}^{\parallel}\boldsymbol{x}\|_{\mathbf{1}}+\boldsymbol{\lambda}\|\boldsymbol{D}^{\perp}\boldsymbol{x}\|_{\mathbf{1}}$$

$$\boldsymbol{subject\ to} \quad \boldsymbol{x}\geq\mathbf{0},$$

Equation 2.1

where x is the vectorized fluence maps, concatenated from all optimized beam fluences. d is the vectorized optimal dose distribution within the body contour, which was resulted from the plan optimization. A is the transformation matrix from the fluence to the dose domain. The matrices $D^{\parallel}$ and $D^{\perp}$ take the derivatives of the fluence maps in the orthogonal directions defined by the MLC leaf orientation. λ is the weighting coefficient. The first L2-data fidelity term of the equation penalizes deviations from the originally optimized dose, and the last two terms with L1 norms are known as total variation regularization to encourage piecewise smoothness in the fluence map. The definition of norm, for some vector a of length n, is $\|a\|_p = \sqrt[p]{\sum_{i=1}^{n}|a_i|^p}$.

As elegant as it is, Equation 2.1 is difficult to solve directly due to the large computational cost required for transformation from the fluence to the dose domain. The transform matrix A contains tens of millions of non-zero entries, and the number of beamlets in x was is on the order of thousands for even 20 radiation beams. Classic convex optimization algorithms such as interior point methods[52-54], as well as general purpose software based on such methods, are typically inefficient at solving problems of this scale. In the past several years, much research in convex optimization algorithms focused on developing first-order methods which are capable of efficiently solving large scale, nondifferentiable, constrained optimization problems. For every convex optimization problem, there is a corresponding "dual problem". Instead of only solving the original "primal" problem, optimization algorithms known as "primal-dual" algorithms simultaneously solve both the primal and the dual problem more efficiently. Of these methods, the Chambolle-Pock algorithm[20] is particularly well suited for our problem. Unlike many other first-order methods such as alternating direction method of multipliers [55], the Chambolle-Pock algorithm does not require solving a linear system involving the matrix A at each iteration, which

is prohibitively expensive due to the size of A. Rather, at each iteration, Chambolle-Pock requires only that we perform multiplications by A and by A^T. Further details of the Chambolle-Pock algorithm can be found in the Appendix.

2.1.2.2 Stratification and MLC Calculation

In order to calculate MLC positions from fluence maps, the fluence map was stratified into a finite number of discrete levels. A larger number of levels, correlated to a smaller step size in the discretization, tends to lead to more complicated fluence maps. After stratification, the MLC sequencing step was performed. A reducing level method, described by Xia and Verhey[6], was utilized for the sequencing. S_k is the k^{th} segment, which is $n \times n$ in size, that is delivered with an intensity level of d_k. R_{k+1} is the residual intensity matrix after S_k and d_k were found, with R_1 being the original intensity matrix. The reduction level method then follows as:

$k = 1$

$L_{max} = Max\ R_1$

$while\ Max\ R_k > 0$

$d_k = 2^{round(log_2(L_{max}))-k}$

$if\ d_k < 1$

$d_k = 1$

end

$$S_k(i,j) = \begin{cases} 0 \; if \; R_k(i,j) < d_k \\ 1 \; if \; R_k(i,j) \geq d_k \end{cases} \qquad for \; all \; 1 < i,j < n$$

$$R_{k+1} = R_k - d_k S_k$$

$$k = k + 1$$

$end.$

Equation 2.2

Since MLCs have to deliver a continuous "on" region along the leaf motion axis for a single segment, a search was performed on S_k to find any bixels that were "off" between "on" bixels inside the segments. A bixel is a basic square unit on an IMRT fluence map that cannot be further divided. It is typically determined by the width of a MLC at the treatment depth. Based on the location of the 'off' bixels, a minimum number of cuts was calculated, and S_k was then subdivided into smaller segments that were MLC deliverable. In order to reduce the small field dose uncertainty, single bixel apertures were removed from the segments. Aperture continuity is defined to be the direct four neighboring bixels. Any identical segments in each beam were combined to reduce MLC complexity. These segments were re-ordered to reduce MLC motion.

2.1.2.3 Evaluation

Six plans for 2 GBM patients, 2 head & neck patients (H&N), and 2 lung patients (LNG), with prescription doses and PTV volumes shown in table 1, were evaluated in this study. The head & neck case #2 has 3 prescription dose levels. These patients were first planned on the 4π radiotherapy platform with 20 beam angles to obtain the optimal beam angle and fluence maps.

Their corresponding planning details can be found in [32,56,57]. Convolution/superposition (CVSP) was used to calculate the beamlet and full doses. Equation 2.1 and Chambolle-Pock were utilized to simplify the optimized fluence maps. This regularized fluence was then stratified and an MLC sequencing was performed.

	Prescription Dose (Gy)	**PTV Volume (cc)**
GBM #1	60	262.5
GBM #2	59.4	377.8
Head & Neck #1	40	18.9
Head & Neck #2	54, 59.4, and 69.94	197.5, 432.6, and 255.0
Lung #1	50	47.8
Lung #2	50	138.8

Table 2.1: Prescription doses and PTV volumes for each of the six cases.

A bisection algorithm was used to modify the step size until exactly 300 total segments were calculated with the reduction level method and after small aperture removal for the 20 beams. Since the total number of segments was limited, a global stratification method, where the step size was held constant for all of the beams, was utilized to allow beams with higher total relative fluence to be more complex than beams with low relative intensities.

Performance of the dose domain regularized (DDR) fluence method was compared to direct MLC segmentation (DMS) of the fluence maps using the reducing level method without dose domain regularization of the fluence maps.

Without losing generality, to further investigate the effects of stratification and MLC sequencing, two additional studies were performed on GBM case #1: The removal of small apertures from the segments, and the reduction of the total number of segments. In the first study, apertures equal to or smaller than 1, 2 or 3 bixels were removed during MLC calculation, and the total number of segments was kept at 300 segments. In the second study, in addition to removing single-bixel apertures, the total number of MLC segments was reduced to 250, 200 and 150 segments, respectively.

The treatment plans were normalized to 100% of the prescription dose to 95% of the volume. PTV homogeneity, D98, D99, and maximum dose, as well as OAR maximum and mean doses were evaluated. Homogeneity was defined as $\frac{D95}{D5}$, and maximum dose, following the recommendation by ICRU-83[58], was defined as the dose to 2% of the structure volume, D2.

Dose differences were normalized and reported as a percentage of their respective prescription dose. The head & neck case #2 was normalized to the 59.4 Gy prescription dose.

$$Percent\ Difference_{PD} = \frac{|Dose_{DDR} - Dose_{DMS}|}{Prescription\ Dose} * 100$$

Equation 2.3

The dose after fluence map optimization but before fluence regularization, stratification, or MLC segmentation is labeled as the unprocessed dose (UPD). A gamma evaluation [59,60] was performed for three sets of dose comparisons: DDR vs. DMS, DDR vs. UPD, and DMS vs. UPD. The first plan stated in each comparison was used as the reference plan. The difference in dose/distance to agreement for the gamma criteria were separately set and tested at different values of 1%/1mm, 2%/2mm, and 3%/3mm, with failure defined at $\gamma > 1$. Gamma calculation was normalized to the maximum dose of the reference plan. Volumes receiving <20% of the maximal dose were suppressed from the calculation.

2.1.3 Results

The average time to solve the dose domain regularization problem with the Chambolle-Pock algorithm was approximately 5 minutes on an i7-3960X CPU with 32GB memory.

Figure 2.1 shows the colormaps of two different beam's fluences. It illustrates how the optimization regularized fluences, and how they were differently affected by the stratification and MLC calculation step. For the first fluence pattern, both DDR and DMS smoothed the fluence while maintaining an overall fluence profile with similar peaks and troughs. However, using DDR, the second fluence map changed significantly due to the coupling effect of all fluence maps in a single optimization equation, in which the deviation of individual fluence map not only was not penalized, but also was compensated by the other beams.

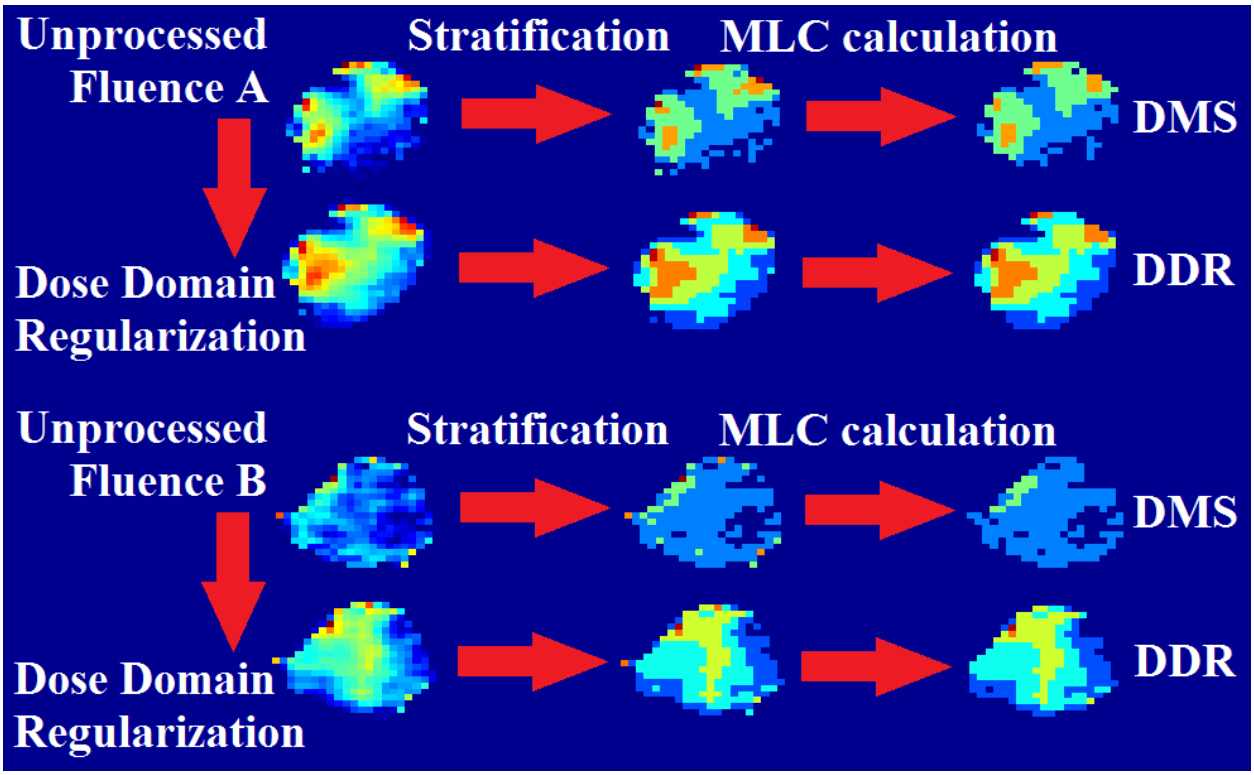


Figure 2.1: Schematic of 2 different fluence maps that were processed using DDR (bottom row of each set) and DMS (top row of each set). DDR minimally changed the fluence map in the first case but substantially changed the fluence map in the second case while DMS maintained the basic fluence map shapes in both cases.

Figure 2.2 shows the DVHs for the GBM, head & neck, and the lung cases, respectively. Increases and decreases in dosimetric indices using DDR compared to DMS method are marked with + and – symbols preceding the number.

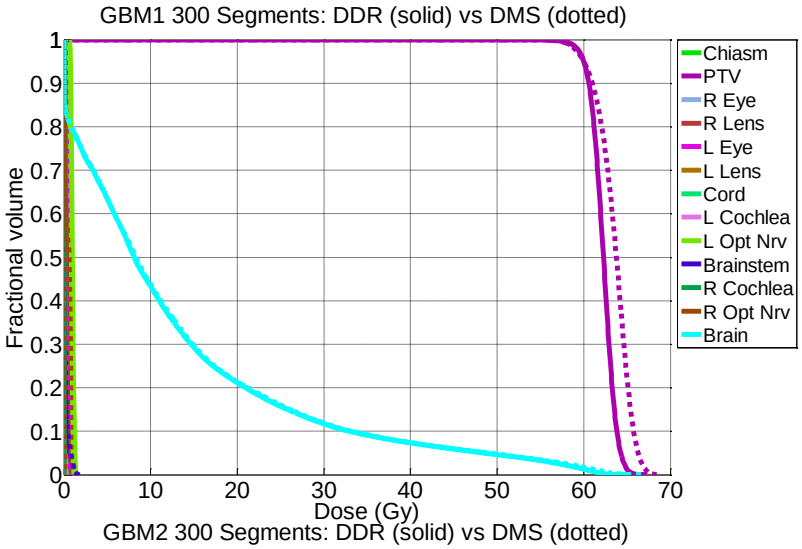
The DDR method globally increased PTV homogeneity and PTV coverage, when comparing to DMS. The average percent increase of D98/D99 between all cases was +1.458%/+2.270%. PTV Dmax was also decreased, with an average percentage decrease of -10.295%. Only the 59.4 Gy

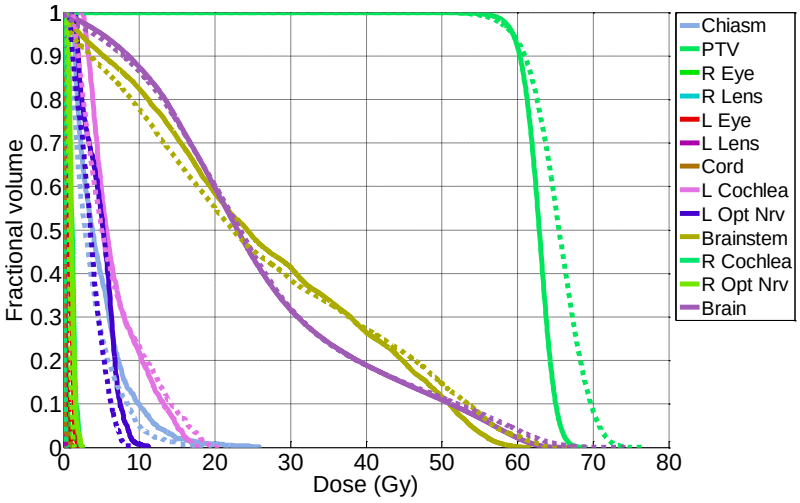
PTV from the head & neck case #2 was included in these averages since plan normalization was based on this structure. Numerical results are presented in table 2.

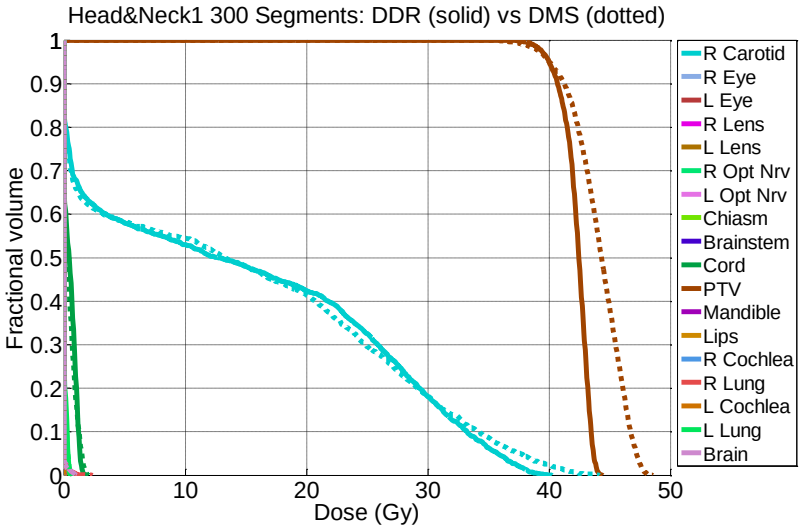
The head & neck case #2 illustrated the greatest differences in doses between the DDR and DMS methods. For the DDR case, there is slight underdosing to the 54 Gy and 69.96 Gy PTVs, as the 95% of the PTVs are only receiving 52.5 Gy and 68.2 Gy, which are consistent to the clinical plans created in Eclipse due to the clinical decision to reduce hot spots. For the DMS case, there is overdosing to most critical structures and to the 69.96 Gy PTV, where 95% of the PTV is receiving 80.9 Gy. Therefore, without the dose domain fluence regularization, the plan re-normalization to maintain the 59.4 Gy PTV coverage resulted in substantial increase in doses to OARs including the pharynx, larynx, mandible, and oral cavity.

The largest and smallest dose value differences for an OAR can be found in table 3 using 300 segments after removing single bixels.

The DDR method improved OARs that have some direct overlap with the PTV, such as the brainstem from the GBM case #2 case and the pharynx, larynx, mandible and oral cavity from the head & neck case #2. The DDR method, in comparison to the DMS method, was able to spare the max dose to these OARs by 4.5 Gy, 19.3 Gy, 5.7 Gy, 16.5 Gy, and 11.9 Gy, respectively. However, dosimetric improvement was not always observed for OARs abutting the PTV. For example, while the DDR method decreased the max dose to the carotid by 2.8 Gy, it increased the max dose to the proximal bronchus by 2.4 Gy. This non-uniform performance is due to the fact that the fidelity term in Eq. (1) is not organ weighted.


GBM1 300 Segments: DDR (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
Chiasm
PTV
R Eye
R Lens
L Eye
L Lens
Cord
L Cochlea
L Opt Nrv
Brainstem
R Cochlea
R Opt Nrv
Brain


GBM2 300 Segments: DDR (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
Chiasm
PTV
R Eye
R Lens
L Eye
L Lens
Cord
L Cochlea
L Opt Nrv
Brainstem
R Cochlea
R Opt Nrv
Brain


Head&Neck1 300 Segments: DDR (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
R Carotid
R Eye
L Eye
R Lens
L Lens
R Opt Nrv
L Opt Nrv
Chiasm
Brainstem
Cord
PTV
Mandible
Lips
R Cochlea
R Lung
L Cochlea
L Lung
Brain

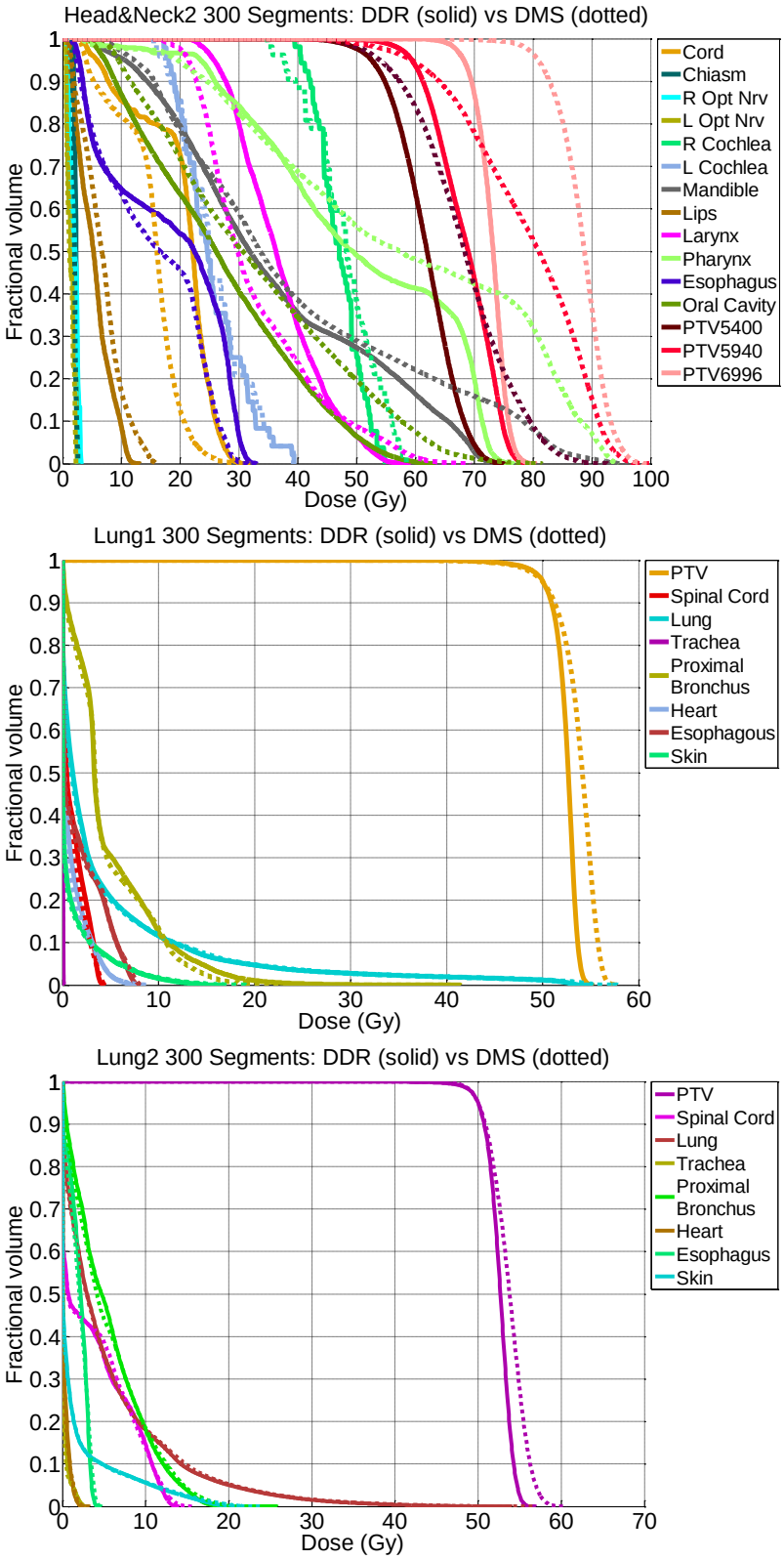


Figure 2.2: DVH comparison of the 6 plans.

PTV statistics		Homogeneity	D95	D98	D99	Dmax
		DMS to DDR	DDR – DMS (Gy)			
GBM #1		0.905 to 0.934	0	+0.302	+0.316	-2.073
GBM #2		0.839 to 0.905	0	+0.777	+1.094	-5.533
Head & Neck #1		0.847 to 0.916	0	+0.259	+0.841	-3.712
Head & Neck #2	PTV5400	0.657 to 0.753	-1.384	-0.743	-0.531	-13.502
	PTV5940	0.640 to 0.785	0	+3.296	5.387	-17.877
	PTV6996	0.856 to 0.891	-12.657	-12.072	-11.372	-18.509
Lung #1		0.891 to 0.927	0	+0.431	+0.281	-2.243
Lung #2		0.878 to 0.914	0	-0.062	-0.239	-2.570

Table 2.2: Comparison of PTV homogeneity, D98, D99, and Dmax for the GBM, head & neck, and lung case. Cases are at 300 segments with single bixel apertures removed.

Dose difference (Gy) (DDR – DMS)	Dmax		Dmean	
	Largest value	Smallest value	Largest value	Smallest value
GBM #1	+0.808 R Eye	-0.895 Brain	+0.701 R Eye	-0.144 Brain
GBM #2	+3.100 Chiasm	-4.538 Brainstem	+1.266 L Opt Nrv	-0.150 Brain
Head & Neck #1	+0.115 R Lung	-2.825 R Carotid	+0.056 Spinal Cord	-0.067 R Carotid
Head & Neck #2	+3.463	-19.262	+5.023	-7.471

	Spinal Cord	Pharynx	Spinal Cord	Pharynx
Lung #1	+2.375 Prox Bronch	-0.287 Esophagous	+0.271 Prox Bronch	+0.007 Trachea
Lung #2	-0.015 Trachea	-0.693 Prox Bronch	+0.102 Prox Bronch	-0.088 Spinal Cord

Table 2.3: Largest and smallest values found for dose differences (DDR – DMS) for the Dmax and Dmean. Cases were calculated using 300 segments with single bixel apertures removed.

GBM #1 Aperture Removal: PTV Statistics	**Homogeneity**	**D98**	**D99**	**Dmax**
	DMS to DDR	DDR – DMS (Gy)		
0 bixels	0.931 to 0.939	+0.091	+0.187	-0.703
1 bixel	0.905 to 0.934	+0.302	+0.316	-2.073
2 bixels	0.898 to 0.918	+0.531	+0.921	-1.172
3 bixels	0.851 to 0.909	+0.761	+1.118	-4.606

Table 2.4: Comparison of PTV homogeneity, D98, D99, and Dmax for the GBM case #1 when small apertures for varying sizes are removed. Cases are all calculated using 300 segments.

GBM #1 Segment	**Homogeneity**	**D98**	**D99**	**Dmax**

Reduction: PTV Statistics				
	DMS to DDR	DDR – DMS (Gy)		
300 segments	0.905 to 0.934	+0.302	+0.316	-2.073
250 segments	0.906 to 0.927	+0.136	+0.220	-1.604
200 segments	0.872 to 0.891	+0.383	+0.701	-1.846
150 segments	0.832 to 0.867	+0.118	+0.542	-3.031

Table 2.5: Comparison of PTV homogeneity, D98, D99, and Dmax for the GBM case #1 when the number of segments are reduced. All cases have single bixel apertures removed from the segments.

GBM #1 Aperture Statistics	**Number of Apertures**		**Max Aperture Size (bixels)**		**Mean Aperture Size (bixels)**	
	DDR	DMS	DDR	DMS	DDR	DMS
300 segments	382	386	189	70	13.98	10.92
250 segments	310	324	184	185	15.65	12.32
200 segments	244	246	162	96	17.76	14.19
150 segments	186	192	111	99	20.24	15.99

Table 2.6: Aperture statistics for the GBM case #1 for both the DDR and DMS cases.

Gamma Pass Rate		DDR vs. DMS			DDR vs. UPD			DMS vs. UPD		
Dose diff/ DTA		1%/ 1mm	2%/ 2mm	3%/ 3mm	1%/ 1mm	2%/ 2mm	3%/ 3mm	1%/ 1mm	2%/ 2mm	3%/ 3mm
GBM		34.97%	75.79%	88.47%	42.37%	82.42%	90.92%	32.27%	73.74%	83.28%

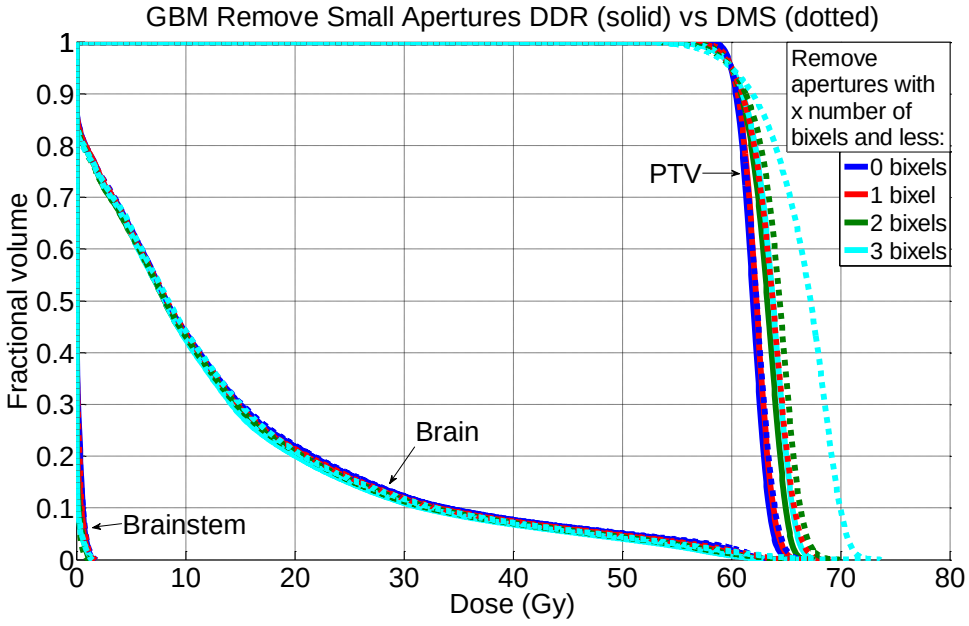
#1	No PTV	39.58%	85.15%	95.76%	50.46%	94.63%	98.92%	39.35%	88.67%	97.23%
GBM #2		34.63%	62.29%	78.28%	44.79%	73.12%	85.42%	35.45%	65.07%	80.33%
	No PTV	18.00%	52.69%	72.76%	30.75%	66.28%	81.72%	18.99%	56.16%	75.31%
H&N #1		19.08%	60.81%	77.35%	20.26%	71.04%	86.94%	9.28%	57.10%	78.10%
	No PTV	19.66%	63.24%	79.98%	21.60%	75.45%	91.83%	9.83%	60.43%	82.43%
H&N #2		11.93%	34.93%	50.08%	15.68%	48.87%	66.51%	8.99%	38.99%	56.44%
	No PTV	8.92%	34.07%	50.48%	12.66%	48.27%	66.68%	5.89%	39.03%	57.99%
LNG #1		41.31%	73.43%	89.43%	59.04%	85.93%	94.11%	41.06%	77.42%	92.46%
	No PTV	42.61%	75.21%	90.92%	61.93%	89.67%	97.32%	42.97%	90.82%	96.02%
LNG #2		28.12%	72.15%	89.27%	37.00%	84.12%	94.34%	19.12%	74.68%	91.46%
	No PTV	28.33%	72.98%	90.08%	38.18%	86.57%	96.34%	19.34%	76.72%	93.49%
Mean		28.34%	63.23%	78.81%	36.52%	74.25%	86.37%	24.36%	64.50%	80.35%
Mean No PTV		26.18%	63.89%	80.00%	35.93%	76.81%	88.80%	22.73%	68.64%	83.75%

Table 2.7: Gamma passing rates of various dose comparisons and at various dose differences and distances to agreement (DTA). The region of interest is the 20% isodose volume for each case. Unprocessed dose (UPD) is the dose before any fluence regularization, stratification, or MLC segmentation. Rows marked as "No PTV" are passing rates excluding the PTV region in the calculation.

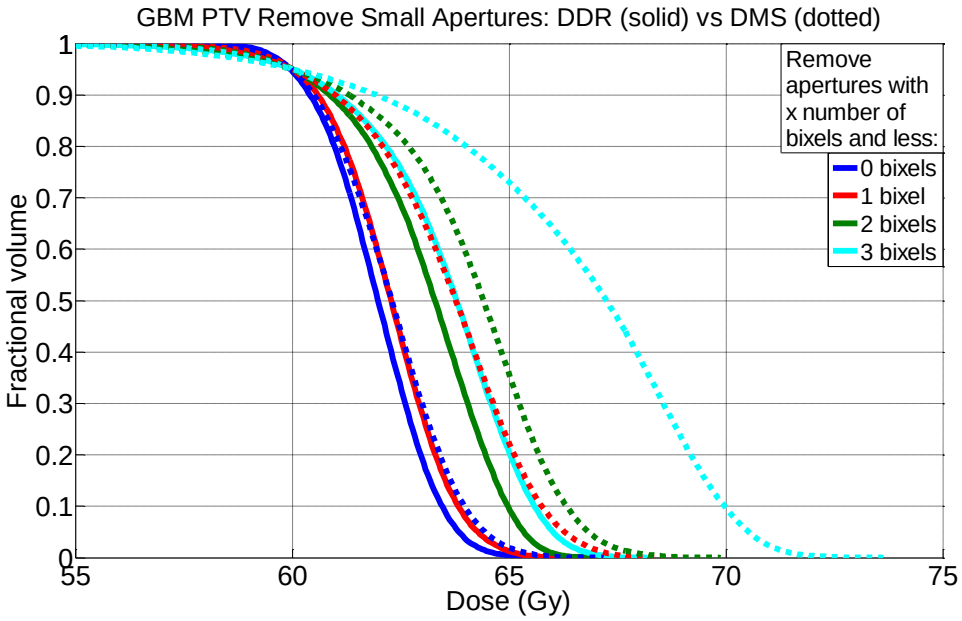
For investigation of stratification and segmentation effects on the dose distribution, Table 2.4 and Figure 2.3 show that the DDR method resulted in an overall increase in PTV homogeneity and PTV dose coverage, as opposed to the DMS method, when bixels of varying size were

removed. Table 2.5 and Figure 2.4 indicate the same trends as the plans are recalculated with 250, 200 and 150 segments.

Superior PTV homogeneity was achieved with dose domain fluence optimization. The homogeneities from the DDR fluences were consistently greater than that from the DMS fluences, regardless of the number of bixels removed or segments reduced. The DMS method had homogeneity drop as low as 0.851 in the GBM case #1 case where 3 bixel or smaller apertures were excluded, as seen in figure 3B, and as low as 0.832 for the GBM 150 segment study. Between the 6 cases, at 300 segments with single bixel apertures removed, PTV D99, on average, increased by 2.27%, indicating better PTV coverage with the DDR method. This excludes the 54 Gy and 69.96 Gy PTVs from the head & neck case #2, since the plan was normalized to the 59.4 Gy PTV.



A



B

Figure 2.3: A) DVH of PTV, brain, and brainstem for the GBM case #1 when aperture sizes less than or equal to 0 bixels, 1 bixel, 2 bixels, and 3 bixels are removed. B) Magnified DVH to show the PTV. All cases are re-stratified and re-calculated to 300 segments.

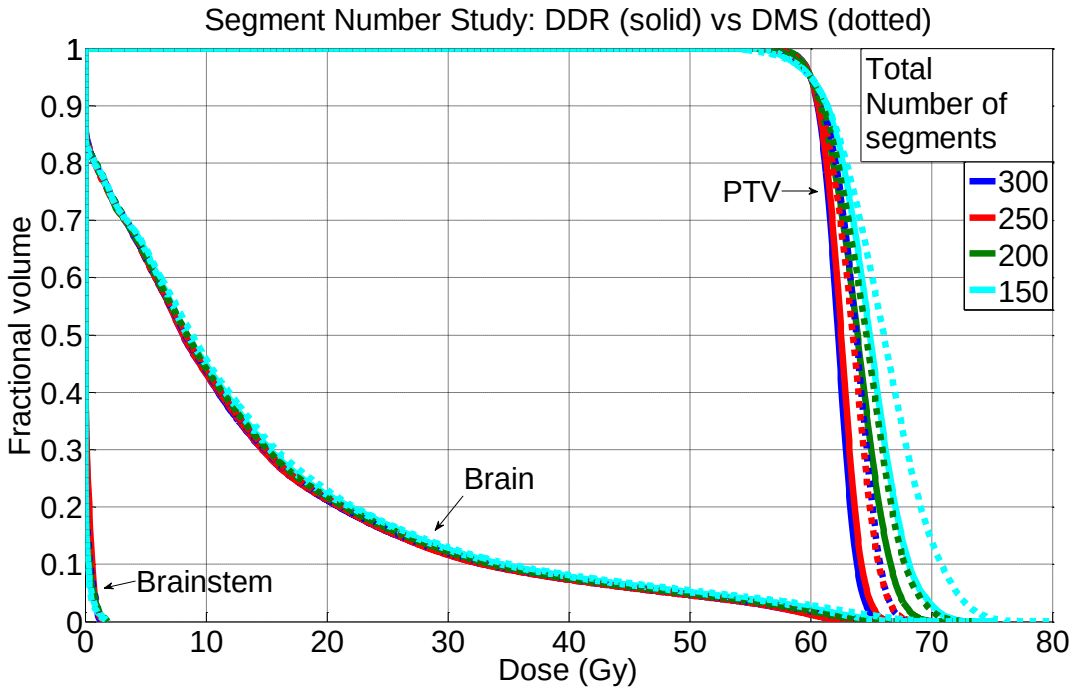


Figure 2.4: DVH comparing plans with 300, 250, 200, and 150 segments for the GBM case #1.

Figure 2.5 shows that the DDR method created overall larger apertures than the DMS method. Total number of apertures for DDR segments was slightly less than that of DMS segments, and with larger maximum and mean aperture sizes, shown in Table 2.6. For the DDR method, with a decreasing number of total segments, the mean aperture size increased, and the maximum aperture size decreased. Reducing the number of segments for the DMS method also resulted in an increase in the mean aperture size, but the maximum aperture size varied without a clear pattern.

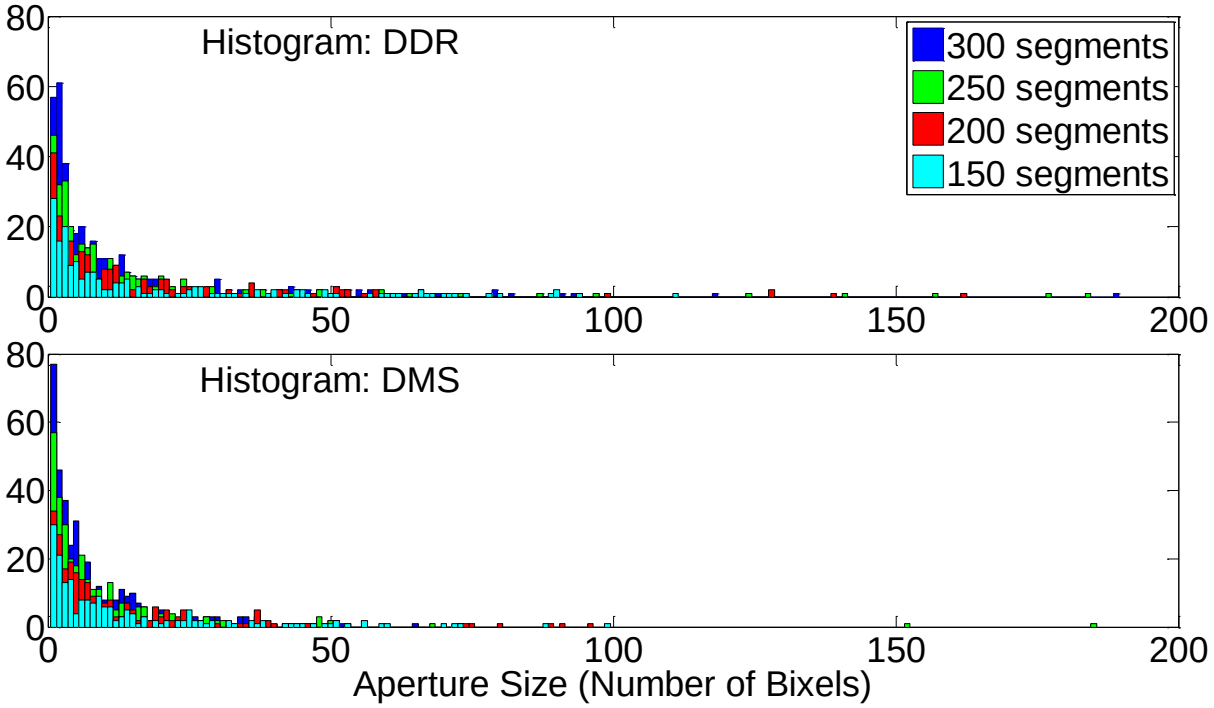


Figure 2.5: Histograms comparing aperture sizes for dose domain regularized (DDR) fluences and direct MLC sequenced (DMS) fluences for varying number of total segments.

The 3D gamma calculation indicated that there was a global increase in the gamma passing rate comparing DDR vs. UPD to DMS vs. UPD, with a mean passing rate of 86.37% for DDR vs. UPD and 80.35% for DMS vs. UPD for 3% and 3mm criteria. This reveals that the dose domain optimization better restored the original dose. Table 2.7 shows the individual gamma passing rates for each case. The gamma metric was utilized as a tool to quantify dose differences from differently modified fluence maps, instead of the typical use for IMRT QA that is based on the same fluence map, so the lower gamma passing rates should be interpreted accordingly. Figure 2.6 indicates where gamma fails ($\gamma > 1$) as both the red and blue colors with a 1% dose difference and 1 mm distance to agreement. Qualitatively, the DDR dose resulted in reduced

high dose in the PTV when compared to the DMS dose, which agreed with the observed high dose tails from the DMS case in DVHs of the PTVs in figures 2-4. For DDR vs. DMS from Table 2.7, only 28.34% of the volume, on average, pass the 1% and 1 mm criteria. For DDR vs. DMS from figure 6, a ratio of the volume of the blue region divided by that of the red region inside the PTV yielded ratios of 3.55 (GBM #1), 4.63 (GBM #2), 5.39 (H&N #1), 8.90 (H&N #2) , 8.51(LNG #1), and 2.94 (LNG #2), indicating that a larger volume of the PTV is receiving higher dose from the DMS case. Both of these observations are mainly due to the greater dose required to maintain the same PTV prescription coverage for the DMS case. The same ratio applied to the body, excluding the PTV, gives ratios of 1.28 (GBM #1), 1.12 (GBM #2), 1.31 (H&N #1), 1.06 (H&N #2), 0.93 (LNG #1), and 1.33 (LNG #2), indicating that for five of the six cases, more healthy tissue was spared using the dose domain filtering. Visual inspection of DDR and DMS dose versus the UPD in Figure 2.6 indicated that the DDR dose had fewer gamma failures than the DMS dose. Figure 2.7 shows the dose color washes for all six cases. It is evident from the color washes that the DMS case produces more hotspots in the PTV, resulting in higher dose heterogeneity. This agrees with the larger amount of gamma failures inside the PTV for the DMS case shown in Figure 2.6.

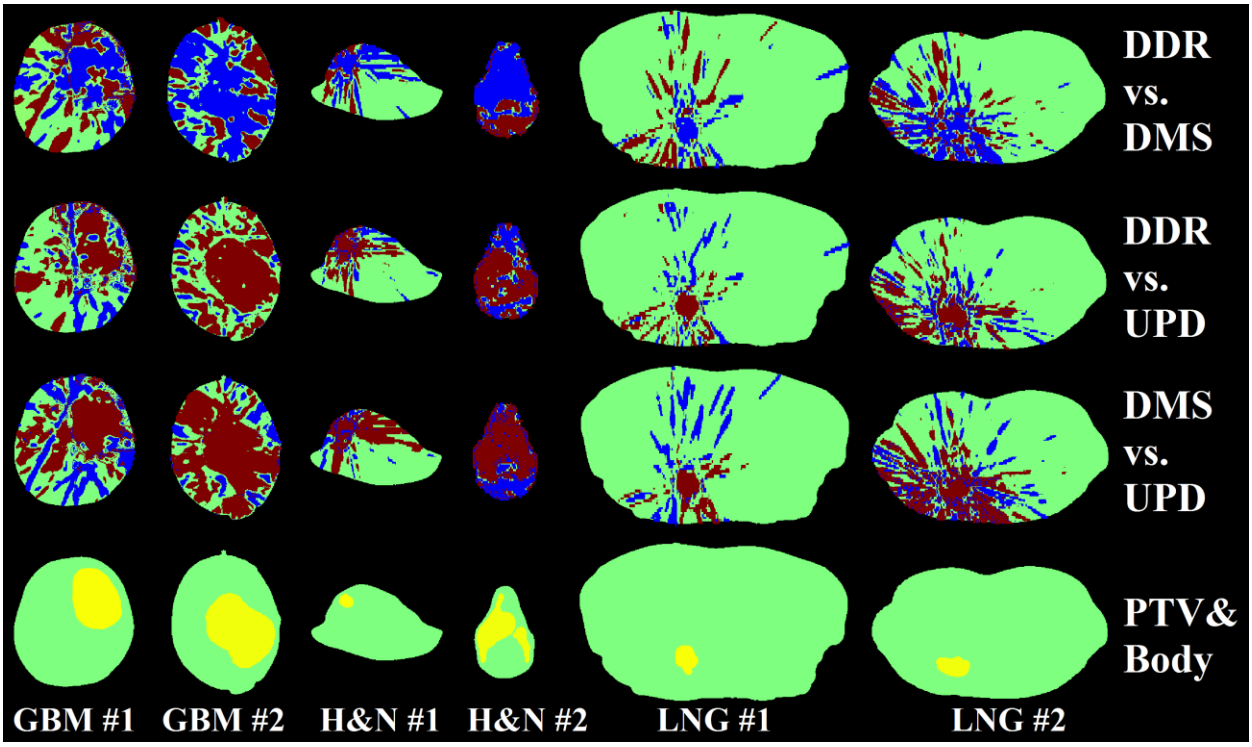


Figure 2.6: Axial slices showing where the gamma fails ($\gamma > 1$) comparing plan 1 vs. plan 2 by either blue, which indicate lower dose from plan 1, or red, which indicates greater dose from plan 1. The body and PTV contours are shown in the last row for reference.

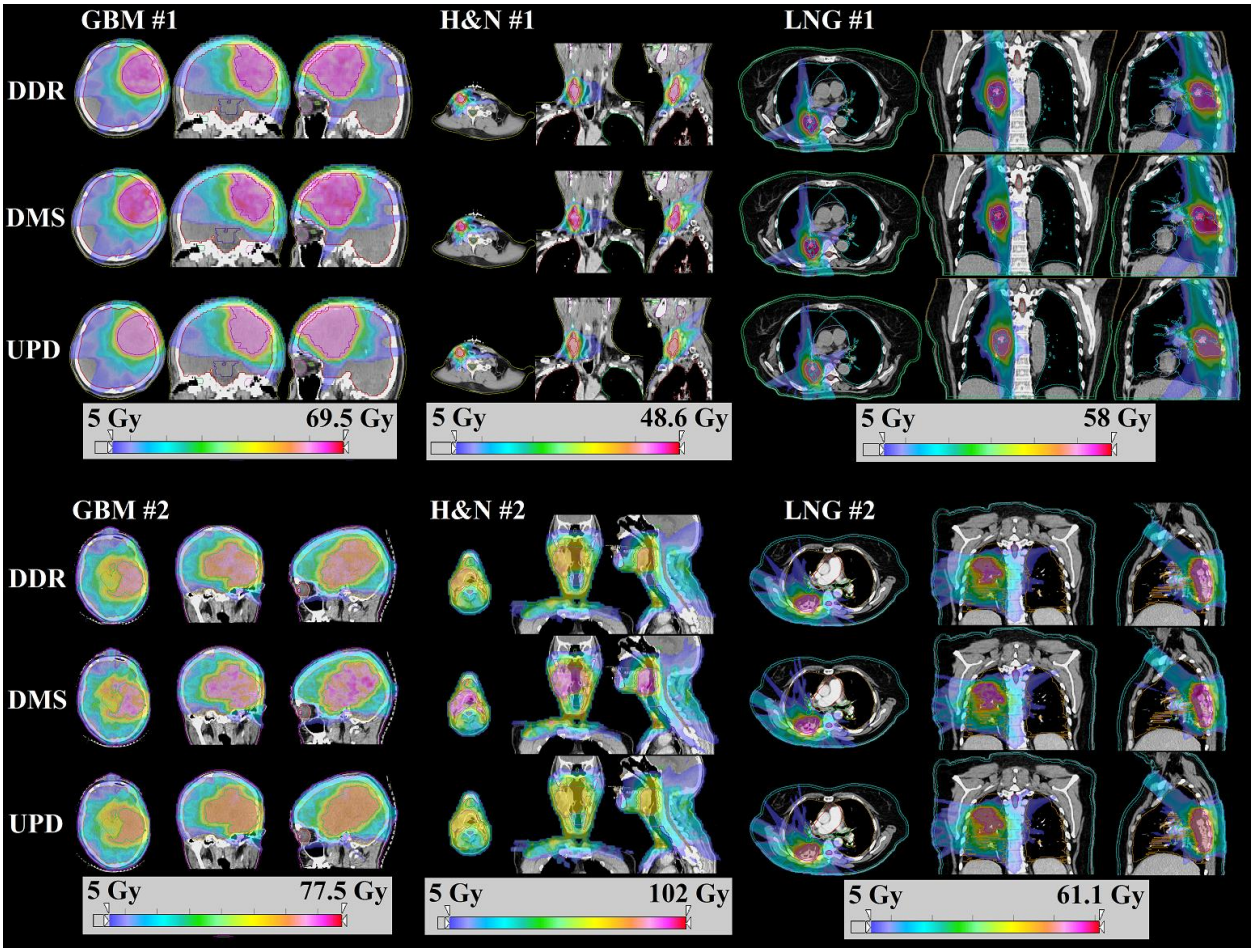


Figure 2.7: Dose color washes of axial, coronal and sagittal slices of all six cases.

2.1.4 Discussion

A novel algorithm for fluence map regularization in the *dose domain* was introduced. To optimally retain plan quality and maximize delivery efficiency, accuracy and plan quality, we first formulated it as a L2 fidelity optimization problem with TV regularization on the fluence map. The objective function in this convex optimization problem contained a fidelity term that directly penalized deviations from the optimal dose distribution, as well as a regularization total variation term that encouraged piecewise smoothness of the fluence map. We used a recently developed Chambolle-Pock algorithm based on first-order primal-dual configuration[20], which

allowed all beams to be considered simultaneously in the dose domain optimization. Since all beams were mapped to the same dose domain, each beam was no longer a separate problem, but was coupled together with the rest. Completely different from previous fluence map smoothing and MLC segmentation studies aiming at minimizing the difference in fluences, our method could be better understood as a secondary global optimization step after the initial fluence optimization. With this approach, individual original fluence maps could be substantially modified to simplify MLC segmentation and still retain the plan fidelity. The increased fluence map regularization flexibility and robustness was more clearly demonstrated when the average number of MLC segments was further reduced to fewer than 10 per field. A limitation of the current method as a secondary optimization is that the organ weighting is not included in the fidelity term, resulting in inconsistent OAR dose improvements as shown in table 3, despite the clear advantage in preserving the physical dose similarity. We expect more consistently improved OAR sparing by including such OAR specific weighting in the future research.

In addition to the ability to better retain the optimized dosimetry, compared to the direct MLC segmentation method, the larger average aperture sizes from DDR fluences might be advantageous in providing more accurately delivered dose. Plans with many small segments are typically less accurate to deliver[35] for several reasons. It has been previously shown that small apertures are extremely sensitive to leaf positioning errors. For a 1 cm width field, 1 mm deviation or radiation field offset can cause a dose delivery error as large as 8%[61,62]. The accuracy of a planning system depends on the accuracy of commissioning data, which are either unavailable or inaccurate for fields smaller than 2 cm due to the lack of electron equilibrium, volume averaging and spectral variation[63,64]. Plans with a larger numbers of small fields are

more susceptible to the leaf end and tongue-and- groove effects, which were not included in the study but can be modeled using a stair function[65].

It is worth pointing out that although a simple level reduction method was used in the study for fluence stratification and MLC sequencing, the fluence regularization method was compatible with different MLC sequencing methods[40,66] for potentially greater improvements.

It is interesting to incorporate the fluence regularization in every iteration of the 4π optimization so the hardware constraints are incorporated in the beams selection as well as the fluence optimization. A possible approach to the problem is a process referred to as direct aperture optimization (DAO), which was commercially implemented as direct machine parameter optimization (DMPO[67]) on the Pinnacle planning system. However, DMPO in its current form is non-convex, highly non-linear and subject to many additional constraints. It is applicable to smaller scale problems where the beams are preselected and limited to a small number. DMPO would render the current 4π optimization framework to select and optimize beams from over 1000 candidate beams mathematically intractable. Therefore, a DMPO or DAO solution for the iterative combined beam orientation and fluence optimization problem is not currently available and will require a new approach. We are currently pursuing a solution to such problem and will compare it to DDR in the future.

2.1.5 Contribution to Deterministic DAO Approach

Dose domain regularization of fluence maps with the Chambolle-Pock algorithm was shown to be more robust to stratification and the removal of small apertures. This property made it optimal for plans that required many beam angles, such as 4π Radiotherapy, since it was beneficial to minimize the total number of segments and MU. The results indicated that, by solving dose domain fluence map regularization problem, it was possible to maintain a more homogenous PTV dose, with minimal dose degradation to most of the critical structures. This study was the first study to show the feasibility of applying the Chambolle-Pock algorithm to a fluence map optimization problem with total variation, which sets up the framework for the optimization problems that are handled in the subsequent studies.

2.2 Computerized triplet beam orientation optimization for MRI-guided Co-60 radiotherapy[68]

2.2.1 Introduction

A newly developed magnetic resonance imaging (MRI) guided intensity modulated radiation therapy (IMRT) system combines a 0.35 T magnetic resonance imaging scanner with therapeutic gamma-rays from Co-60 sources. The combination provides daily and intrafractional MRI soft tissue and functional imaging for improved target tracking and adaptive radiotherapy[69]. However, the use of cobalt sources instead of X-ray linac has significant implications including low output from a single source and subsequently long treatment time that worsens with Co-60 decay[70]. To compensate the low output, the system uses three Co-60 sources equidistantly spaced

120° apart to triple the output. However, this source arrangement complicates beam selection. In conventional treatment planning, a dosimetrist would intuitively select beams that better avoid the critical organs. The experience, however, does not apply to the triple source Co-60 platform (Tri-Co-60). As shown in Figure 2.8, beam 1 is ideally selected to avoid central organs that are sensitive to radiation but the other two beams in the triplet group would penetrate the spinal cord and the heart. The dosimetrist is often compelled to make an undesired tradeoff between the plan quality and delivery time. Understandably, the human beam selection becomes increasingly difficult with increasing number of beams. Therefore, there is a strong motivation to computerize beam orientation selection. Previously, we showed in non-coplanar treatment planning, computerized beam orientation optimization is a compelling solution in situation where the manual beam selection is no longer feasible[71-74]. Similarly, a computerized beam orientation optimization method may significantly improve the planning efficiency and quality for the tri-Co-60 platform.

A second important consideration for the study is that the deliverability of the computerized beam selection plans has to be equal to or better than the original clinical plan. Different from a linac that the X-ray can be switched on and off almost instantaneously, the tri-Co-60 sources need to be retracted into the safe between beam segments in the step-and-shoot IMRT delivery and slow down the treatment. In other words, the tri-Co-60 treatment plans need to be more "frugal" in the number of MLC segments. In the commercial MRI guided tri-Co-60 planning system, a plan efficiency parameter was implemented to reduce the number of MLC segments. The parameter is effective in attaining plans with a few segments but often at a cost of plan quality. In this study, to compare plans on a fair ground, the number of MLC segments needs to be part of the calculation. In the computerized beam orientation optimization approach, to avoid

dose degradation typically seen in conventional fluences map simplification and MLC segmentation[6,37-40], we introduce a second dose domain regularization (DDR) problem to reduce the number of MLC segments while maintaining the dosimetric quality. The DDR formulation is convex and can be efficiently solved with the Chambolle-Pock algorithm, a first-order primal-dual algorithm[20,21]. The motivation of the study can thus be summarized as a **F**luence **R**egularization and **O**ptimized **S**election of **T**riplets (FROST) method for tri-Co-60 beam orientation and the fluence optimization.

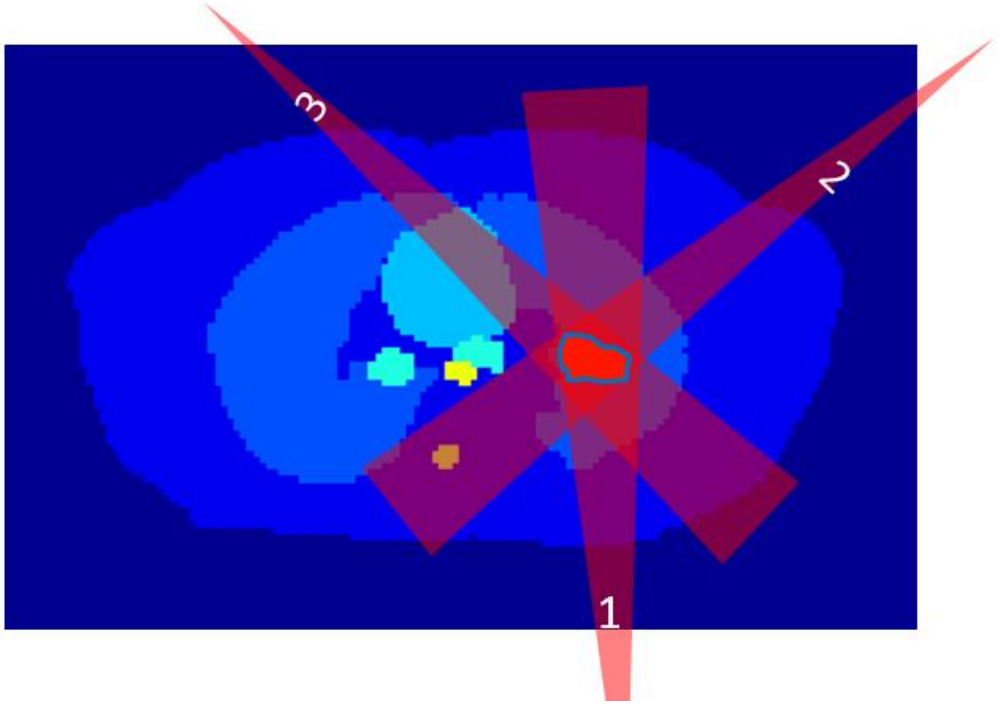


Figure 2.8: Example geometry for a lung tumor (outlined in blue) and critical structures (various colors). A triplet of Co-60 beams are shown to target the tumor. Beam 1 is ideal to treat the tumor and avoid central critical organs but the other two beams in the triplet are suboptimal.

2.2.2 Methods

As explained in the introduction, FROST planning is divided in *three* main sequential steps: (1) the computerized triplet angle optimization, (2) the dose domain regularization of the fluence map and (3) stratification and MLC segmentation. The details are described as follows.

2.2.2.1 Step 1: Computerized Triplet Angle and Fluence Map Optimization

The triplet angle selection and fluence map optimization is based on a column generation algorithm, whose variables are tabulated in Table 2.8.

Variable	Type	Description
$\boldsymbol{T_{all}}$	Set	Set of all available triplets.
$\boldsymbol{T_{select}}$	Set	Set of selected triplets.
$\boldsymbol{t}$	Index	Index for triplet.
$\boldsymbol{s}$	Index	Index for OAR.
$\boldsymbol{r}$	Index	Index for PTV.
$\boldsymbol{m}$	Index $m \in s, r$	Index for structure. Includes both OAR and PTV structures.
$\boldsymbol{\alpha_m}$	Scalar	Weight for m^{th}structure.
$\boldsymbol{G_m}(\cdot)$	Function	Cost function for m^{th} structure.

$\boldsymbol{x_t}$	Vector	Fluence of t^{th} triplet. Single element of x_t is referred to as "*beamlet*".
$\boldsymbol{A_t}$	Matrix	Fluence-to-dose transformation matrix for t^{th}triplet.
$\boldsymbol{d}$	Vector	Dose array containing all voxels in dose domain.
$\boldsymbol{d_m}$	Vector	Dose array containing only voxels for m^{th} structure.
$\boldsymbol{\gamma_s}$	Scalar $0 \leq \gamma_s \leq 1$	Tradeoff factor between mean dose and max dose for the s^{th} OAR.
$\boldsymbol{p_r}$	Vector	Prescription dose for r^{th} PTV. Entire vector typically set to one value.
$\boldsymbol{u}$	Vector	Lagrange multiplier corresponding to the constraint $d = \sum_{t \in T_{all}} A_t x_t$.
$\boldsymbol{v}$	Vector	Lagrange multiplier corresponding to the constraint $d \leq q$.
$\boldsymbol{w_t}$	Vector	Lagrange multiplier corresponding to the constraint $x_t \geq 0 \quad for\ t \in T_{all}$.

Table 2.8: List and description of variables involved with the column generation algorithm.

To solve the integrated triplet selection and fluence map optimization problem, we consider the following master problem:

$$\begin{aligned} \underset{d,\{x_t\}_{t\in T_{all}}}{argmin} \quad & \sum_{m\in s,r} \alpha_m G_m(d_m) \\ subject\ to \quad & d = \sum_{t\in T_{all}} A_t x_t \\ & d \le q \\ & x_t \ge 0 \qquad for\ t \in T_{all}, \end{aligned}$$

Equation 2.4

where

$$G_s(d) = \gamma_s mean(d_s) + (1-\gamma_s)\max(d_s) \qquad for\ the\ s^{th}\ OAR$$

$$G_r(d) = mean((p_r - d_r)_+) \qquad for\ the\ r^{th}\ PTV.$$

Equation 2.5

The optimization variables are the triplet fluence, x_t, and the projected dose, d. The fluence-to-dose transformation matrix, A_t, maps fluence f_t to dose d based on precomputed beamlets. The OAR objective function, G_s, penalizes the mean and max dose for the s^{th} OAR. The weighting factor, $\gamma_s \in (0,1)$ is determined by the radiobiological seriality of the organ. For parallel and serial organs, γ_s approaches 1 and 0 to weigh more heavily on the mean and max doses, respectively. In this study, γ_s was set to 0.75 for highly parallel structures, such as the lung, 0.25 for highly serial structures, such as the spinal cord and 0.5 for organs with seriality in between.

The PTV objective function, G_r penalizes underdosing to the PTV from the prescription dose, p_r. The non-negative operator, $(\cdot)_+$, projects the argument onto the non-negative orthant. The structure weight, α_m, determines the importance of individual structures. The upper bound constraint, q, limits the maximum dose to any voxel. In this study, q was set to be 110% of the prescription dose to hard-constrain the maximum dose.

With $t \in T_{all}$, the master problem is formulated to optimize the fluence of *all* candidate triplets. However, it is not of our interest to solve the master problem. Other than being computationally expensive, the solution is not practical to deliver. Instead, our goal is to select a *small subset* of triplets, T_{select}, that are most important to the master problem. We define a subproblem similar to the master problem:

$$\begin{aligned} \underset{d,\{x_t\}_{t\in T_{select}}}{argmin} \quad & \sum_{m\in s,r} \alpha_m G_m(d) \\ subject\ to \quad & d = \sum_{t\in T_{select}} A_t x_t \\ & d \le q \\ & x_t \ge 0 \qquad for\ t \in T_{select}. \end{aligned}$$

Equation 2.6

The only difference here is that the subproblem is solving over $t \in T_{select}$ instead of $t \in T_{all}$.

The goal of treatment planning is to find a suitable T_{select} and optimize the fluence using Equation 2.6. To add a triplet from $T_{all} \backslash T_{select}$ to T_{select}, we evaluate the Karush–Kuhn–Tucker (KKT) conditions[75] of the master problem.

The KKT conditions of the master problem are divided into 4 condition categories including the primal feasibility, dual feasibility, stationarity and complementary slackness, of which the dual feasibility and stationarity conditions are used to select the triplets:

$$\left.\begin{array}{ll} v \geq 0 & \\ w_t \geq 0 & for\ t \in T_{all} \end{array}\right\} \quad dual\ feasibility$$

$$\left.\begin{array}{ll} w_t = A_t^T u & for\ t \in T_{all} \\ u \in \sum_{m \in s,r} \left(\alpha_m \partial G_m(d_m)\right) + v_m & \end{array}\right\} \quad stationarity.$$

Equation 2.7

The variables u, v, and w_t are Lagrange multipliers associated to the subject to constraints of the master problem (see table 1). To determine the most valuable triplet that is not currently in the selected set, we are particularly interested in evaluating the Lagrange multiplier, w_t, which is the multiplier corresponding to the constraint $x_t \geq 0$. All w_t where $t \in T_{all} \backslash T_{select}$ are calculated using the first stationarity condition, $w_t = A_t^T u$. The Lagrange multiplier u is directly obtained as one of the dual variables from the primal-dual solution of the subproblem for a given T_{select}. Note that this applies to the initialization of optimization when T_{select} is an empty set. While w_t

meets the non-negativity dual feasibility condition for $t \in T_{select}$, the variable can be negative where $t \in T_{all} \backslash T_{select}$. Since w_t is a vector, all the negative values in w_t for a triplet are summed to produce a single value for that triplet. The triplet with the most negative value is responsible for the worst KKT violation, and thus will have the highest contribution in reducing the objective value of the master problem. This triplet is added to T_{select}, and the subproblem is minimized again with the updated T_{select}. The iteration solves what is known as the pricing problem[76]. The pseudo code of the column generation algorithm for triplet selection and fluence optimization is as follows:

- Initialize empty selection set T_{select}
- Repeat until desired number of triplets is achieved

  {
  - Solve subproblem in Equation 2.6
  - Directly obtain Lagrangian multiplier u
  - Calculate $w_t = A_t^T u$ for $t \in T_{all} \backslash T_{select}$
  - Find $t^* = \underset{t \in T_{all} \backslash T_{select}}{argmin} \; sum((w_t)_-)$
  - Add t^* to T_{select}

  }

Equation 2.8

The non-positive projection operator, $(\cdot)_-$, projects the argument onto the non-positive orthant. The formulation is solved using CPLEX (IBM, Academic Research Edition 12.2).

2.2.2.2 Step 2: Dose Domain Fluence Map Regularization Formulation

For tri-Co-60 planning, it is equally important to minimize the number of MLC segments via simplification of the fluence maps. We introduced fluence map regularization in the dose domain optimization problem to control the number of MLC segments while maintaining the optimization results[77]. This dose domain regularization (DDR) method is shown superior to other fluence map simplification methods that relied on smoothing the map while minimizing the difference between the original and simplified map[6,38,78-80], due to the fact that the fidelity term minimized error in the dose domain, rather than the fluence domain. The DDR method was applied to the problem at hand following the cost function:

$$\underset{\{\boldsymbol{f}_t\}_{t\in \boldsymbol{T}_{select}}}{\boldsymbol{argmin}} \quad \frac{\mathbf{1}}{\mathbf{2}}\left\| \boldsymbol{W}\left(\sum_{t\in \boldsymbol{T}_{select}} (\boldsymbol{A}_t\boldsymbol{x}_t) - \boldsymbol{d}\right)\right\|_{\mathbf{2}}^{\mathbf{2}} + \boldsymbol{\lambda}\sum_{t\in \boldsymbol{T}_{select}} (\|\boldsymbol{D}^{\parallel}\boldsymbol{x}_t\|_{\mathbf{1}} + \boldsymbol{\lambda}\|\boldsymbol{D}^{\perp}\boldsymbol{x}_t\|_{\mathbf{1}})$$

$$\boldsymbol{subject\ to} \qquad \boldsymbol{x}_t \geq \mathbf{0} \qquad \boldsymbol{for\ t} \in \boldsymbol{T}_{select},$$

Equation 2.9

where W is a diagonal weighting matrix that weights the structures of interest. $D^{\parallel}$ and $D^{\perp}$ take the derivatives of the fluence parallel and perpendicular to the MLC leaf movement direction. The weighting matrix W is initially selected based from the α_m in the triplet selection processes, and minor adjustments are made if the dosimetry is not acceptable. The weight λ controls how much influence the total variation term has on the cost function, and was set as $\lambda = \frac{1}{100}\|[A_{t=1} \quad \dots \quad A_{t=n}]\|_1$. For some vector, x, the p^{th} norm is defined as $\|x\|_p = \sqrt[p]{\sum_i (x_i)^p}$.

The optimization formulation, Equation 2.9, was solved utilizing the Chambolle-Pock algorithm, a proximal-class primal-dual algorithm[20,21]. This optimization formulation, with an L2-norm fidelity and an anisotropic total variation regularization, matches the formulation presented previously for dose domain regularization[28], which the only exception being the added weighting term, W. With the same base formulation, the operations performed for using the Chambolle-Pock algorithm are identical to the one presented previously, which was already described in great detail. The Chambolle-Pock algorithm efficiently handles problems of this form, since it does not need to solve a system of linear equations at each iteration. The majority of its computational expense comes from the straightforward multiplication of A_t, D_1, and D_2 on to the fluence and its corresponding dual variable at each iteration.

It is noted that the FROST approach is heuristic and does not guarantee global optimality. This algorithm for triplet selection is greedy in nature, and the dose domain regularization is a separate optimization with a different cost function, performed after the triplet selection process is entirely finished.

2.2.2.3 Step 3: Stratification and MLC Segmentation

In order to convert the optimized fluence maps into MLC deliverable fluence maps, an MLC segmentation algorithm, based on the reducing level method by Xia and Verhey[6], is implemented. This algorithm stratifies the fluences into in to a discrete equal step sizes, and then, using the reducing level method, breaks down the binned fluence to segments that are deliverable

along the leaf direction. Further details on the algorithm were presented previously[28]. A bisection algorithm that adjusts the stratification step size is utilized to find a specified number of deliverable segments.

2.2.2.4 Evaluation

Three patients, which includes 1 prostate (PRT), 1 lung (LNG), and 1 head and neck boost plan ($H\&N_{Boost}$), were evaluated in this study. Patient data is shown in Table 2.1, including the initial plan for the head and neck case ($H\&N_{Initial}$), which was also planned and treated with the MRI-guided tri-Co-60 platform. The beamlet dose was calculated using a Monte Carlo dose calculation engine derived from the well-verified Dose Planning Method (DPM)[81] with various improvements in efficiency, step size artifact, and variance reduction techniques. This implementation considers all electrons set free from interactions sites with a high weighting (i.e. considers them to be "fat"), which leads to an increased degree of uncertainty when a low number of statistics are used, especially in the out of field region were contaminant electron statistics are poor. The engine utilizes the patient electron density for dose calculation, and also accounts for the attenuation of the patient couch during simulation. The number of histories per beamlet was set to 10^6, as compared to the 2.5 x 10^6 particle histories used clinically. The magnetic field included in the simulation to accurately account for its effect on secondary electron scatter. The beamlet size is 1.05 cm x 0.5 cm, as defined by the MLC leaf width and the step size. For each beam, MLC positions were conformed to the PTV structure + 1 cm margin from beam's eye view to create a set of conformal fields. In order to reduce out-of-field dose that is primarily caused by Monte Carlo calculation noise in each beamlet, the beamlet dose outside the 9 cm diameter cylinder along the beamlet axis was set to zero. Figure 2.9 shows a

comparison of the beamlet dose with and without the cutoff applied. For each patient, the beamlet dose was calculated for 180 coplanar beams, spaced 2° apart, and the beams were grouped into 60 triplets. After dose calculation, the dose information is stored into the dose matrix A_t for each triplet, to allow for conversion from the beamlet intensity to the dose.

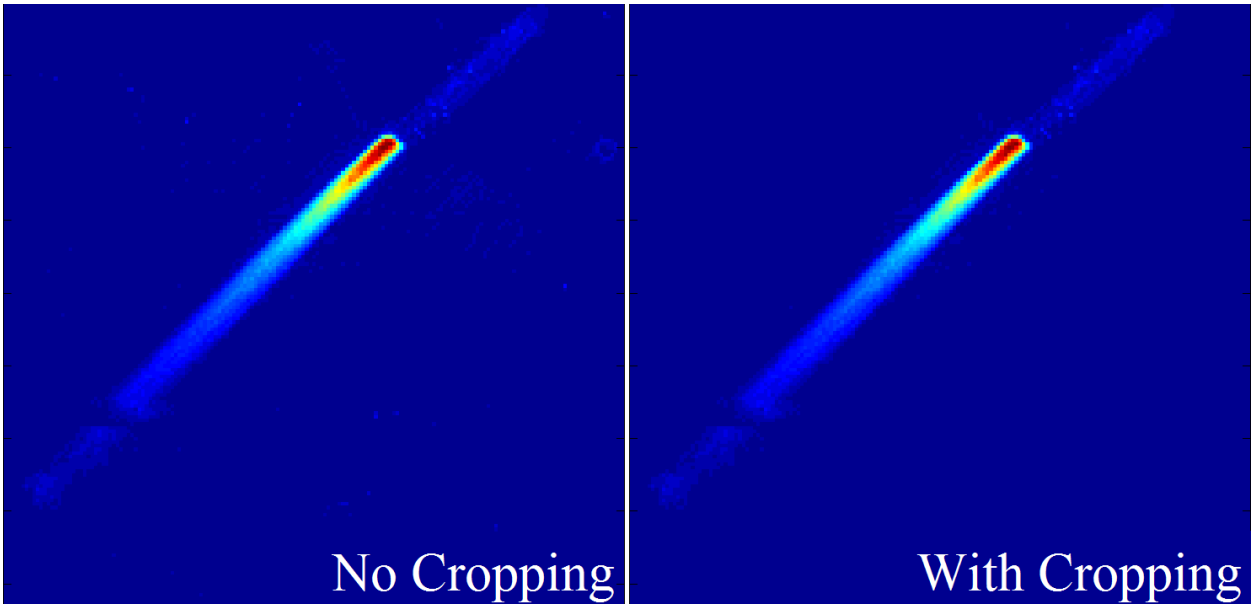


Figure 2.9: Beamlet dose comparison between the raw Monte Carlo dose without cropping (left) and the dose after cropping was applied (right). Cropping allows for the removal of the extra noise outside of the beamlet-line to reduce computational expense without affecting the beamlet's dose contribution.

	Prescription Dose (Gy)	PTV Volume (cc)
PRT	40	68.909
LNG	50	8.094

$H\&N_{Boost}$	14	4.363
$H\&N_{Initial}$	66	28.190

Table 2.9: Prescription dose, and PTV volume for each patient case

For evaluation, the FROST plans were compared against the clinical plans (CLN), which were produced by an experienced dosimetrist. The $H\&N_{Boost}$ plan was compared standalone, as well as a comparison as a plan sum with the initial H&N plan ($H\&N_{PlanSum}$). The plan sums for the FROST and CLN methods both use the same $H\&N_{Initial}$ plan, and the only differing aspect is the FROST or CLN boost plan. PTV homogeneity, max dose, mean dose, D95, D98, and D99 were evaluated. OAR max and mean doses were also evaluated and compared. PTV homogeneity is defined as $\frac{D95}{D5}$, The maximum dose, following the recommendations of ICRU-83[58], is defined as the dose to 2% of the structure volume, D2. R50, a measure of high dose spillage and defined as the ratio of the 50% isodose volume over the planning target volume, was also determined.

2.2.3 Results

The Monte Carlo beamlet dose calculation took an average of 5 hours per patient for all 180 coplanar beams using a CPU. This is an additional one time cost compared to the commercial planning system using a proprietary analytical model to calculate beamlets only for the few manually selected beams. However, the time is expected to reduce with GPU parallelization and adopting a similar analytical beamlet model. The dose matrix size ranged from 20 MB to 40 MB per beam—approximately 100 MB per triplet. The beam angle selection process took 20 min to

30 min to complete, and the dose domain fluence map regularization process took approximately 5 minutes. The process and was repeated if OAR and PTV weighting parameters needed to be retuned. On average, the weighting parameters were retuned 2 to 3 times for an acceptable dose distribution. It should be noted that no interaction is needed with the optimizers while it is running, so the time needed for active user intervention is minimal. In comparison, the laborious manual planning by the dosimetrists took several hours or longer to complete a plan.

Patient Case		Number of:					Average number of segments per beam
		Beam Groups	Triplets	Doublets	Singlets	"On" Beams	
PRT	FROST	6	6	0	0	18	8.77
	CLN	6	6	0	0	18	8.77
LNG	FROST	6	1	5	0	13	2.31
	CLN	6	3	2	1	14	2.29
$H\&N_{Boost}$	FROST	7	7	0	0	21	3.00
	CLN	7	2	4	1	15	2.73

Table 2.10: Beam triplet and number of segments per beam data for each case

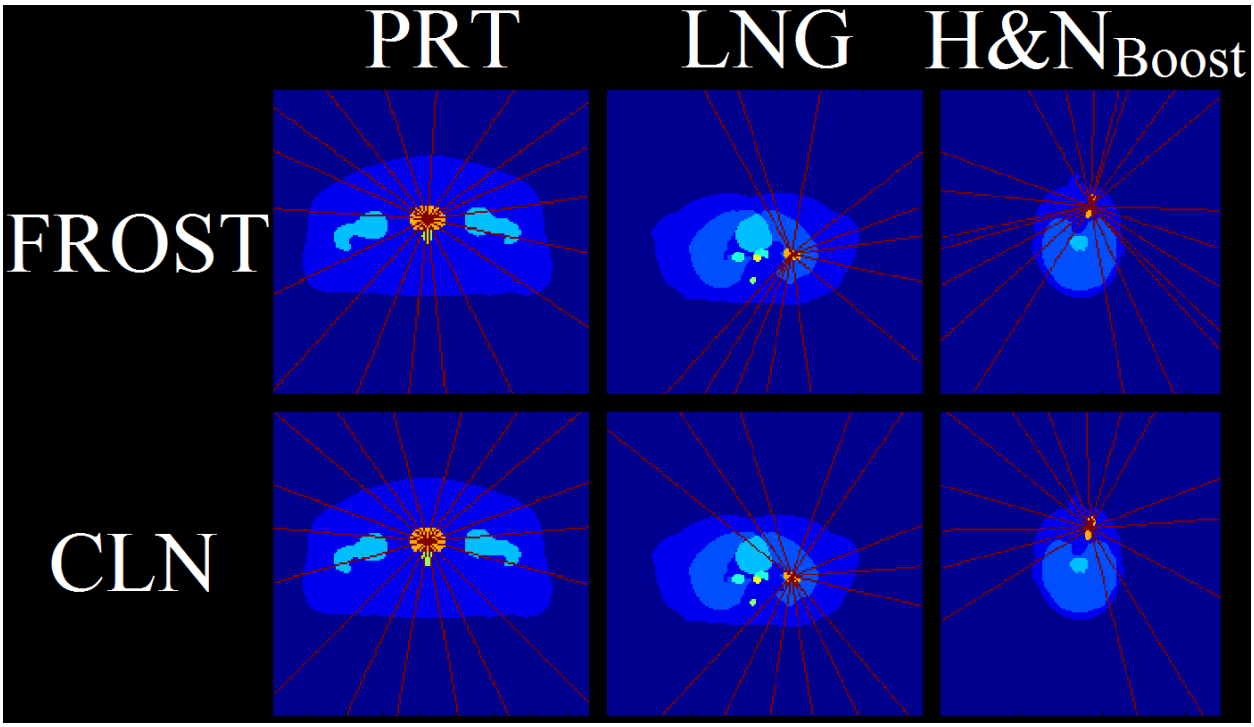


Figure 2.10: Schematic of the beam angle arrangements for the on beams for each case. The beam angles are indicated in red and the PTV in orange. All other OARs are displayed in various colors.

The FROST plans either resulted in same or fewer beam groups than the CLN cases. Because of the capability of simultaneous delivery of the grouped beams, the delivery time is heavily influenced by the number of triplets rather than the number of "on" beams. The number of beam groups and its breakdown into triplets, doublets, and singlets is shown in Table 2.10. The PRT case used similar beam groups and arrangements between both FROST and CLN plans. However the LNG and $H\&N_{Boost}$ cases used markedly different beam orientations in FROST than in CLN, shown in the schematic in Figure 2.10. Although the total number of beams is the same in FROST and CLN for the LNG case, there are fewer beam groups in FROST than that in CLN and all beam groups are triplets in FROST vs. two triplets and 3 doublets in CLN. The $H\&N_{Boost}$

case used the same number of beam groups in both FROST and CLN, but the selected FROST beam groups are all triplets, in contrast to the CLN plan that used two triplets, four doublets and one singlet.

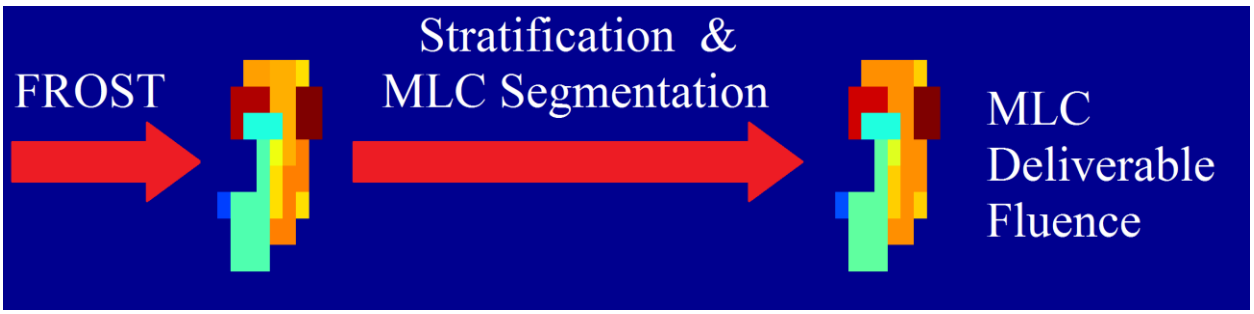
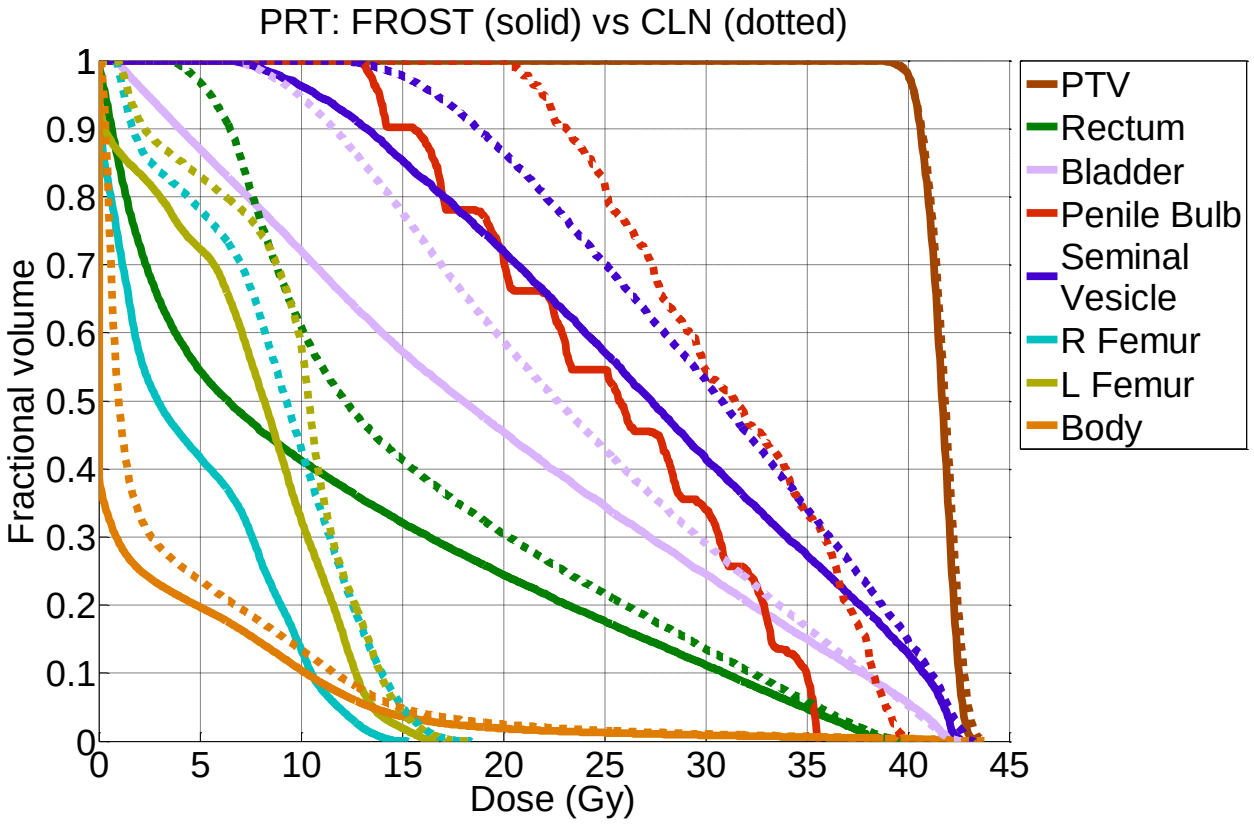
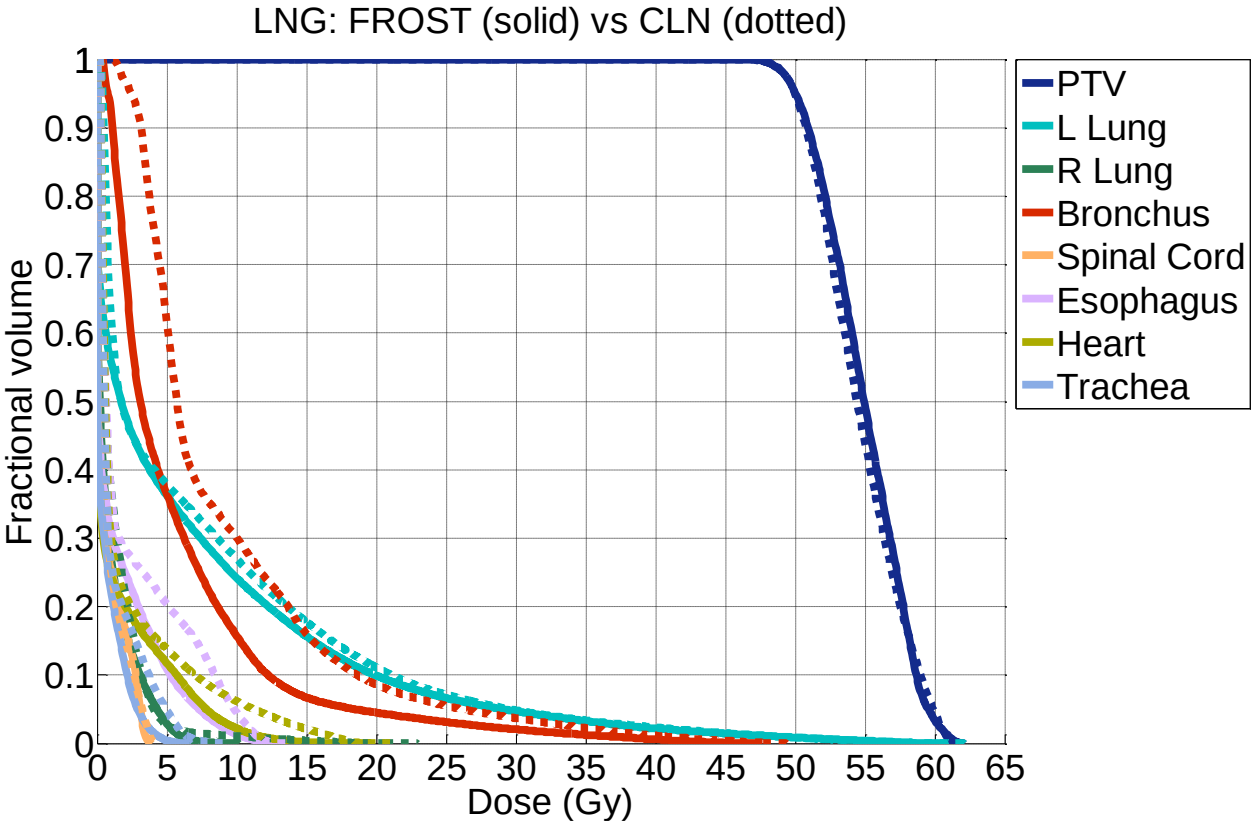
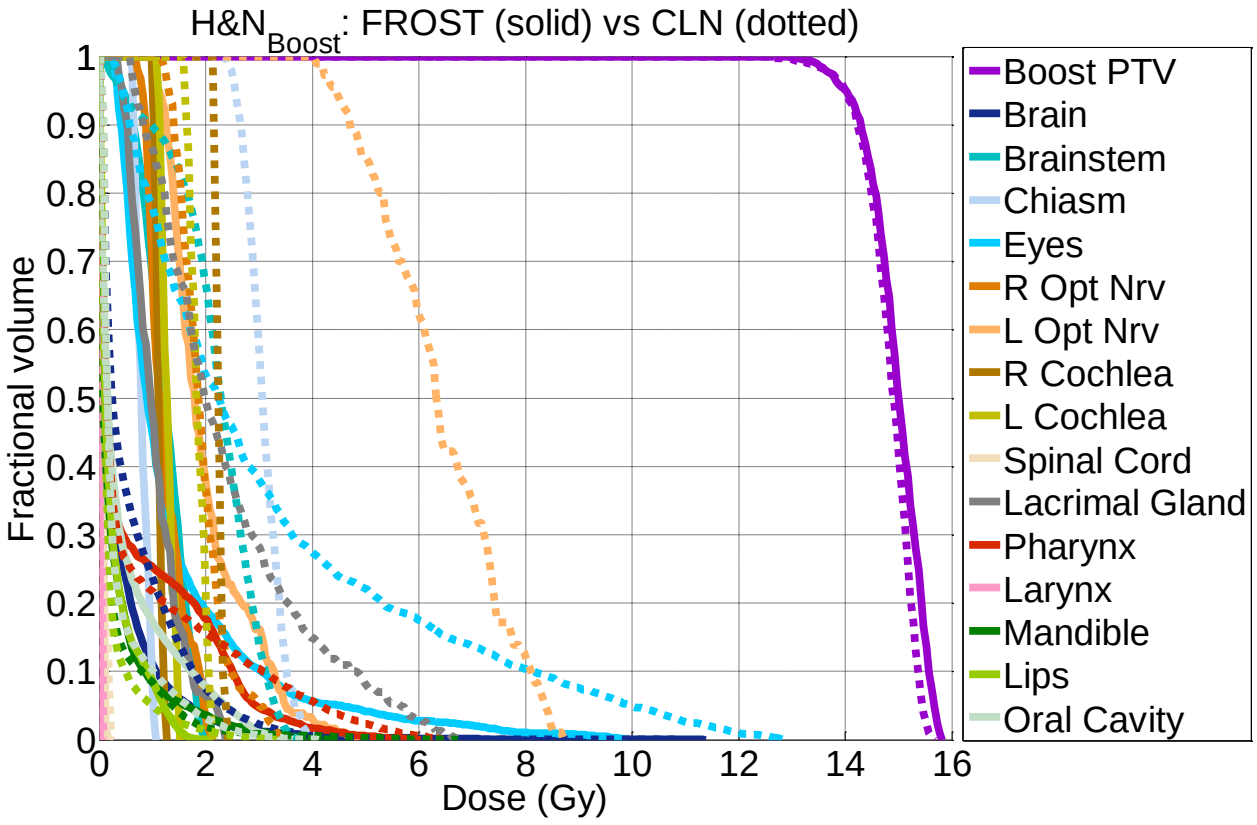


Figure 2.11: Schematic of a PRT beam fluence optimized with FROST, and then undergoing post-processing steps, described previously[28], to make it deliverable.

Figure 2.11 shows a FROST fluence map that is subject to stratification and MLC segmentation. The total variation regularization term in the FROST formulation encourages piecewise smoothness in the fluence maps, giving the fluences a blocky pattern that is robust to the MLC segmentation step. As shown in Figure 2.11, the fluence map was minimally changed with stratification and the MLC segmentation to preserve the optimized plan quality.


PRT: FROST (solid) vs CLN (dotted)
Fractional volume
Dose (Gy)
PTV
Rectum
Bladder
Penile Bulb
Seminal Vesicle
R Femur
L Femur
Body


LNG: FROST (solid) vs CLN (dotted)
Fractional volume
Dose (Gy)
PTV
L Lung
R Lung
Bronchus
Spinal Cord
Esophagus
Heart
Trachea


H&N Boost: FROST (solid) vs CLN (dotted)
Fractional volume
Dose (Gy)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
2
4
6
8
10
12
14
16
Boost PTV
Brain
Brainstem
Chiasm
Eyes
R Opt Nrv
L Opt Nrv
R Cochlea
L Cochlea
Spinal Cord
Lacrimal Gland
Pharynx
Larynx
Mandible
Lips
Oral Cavity

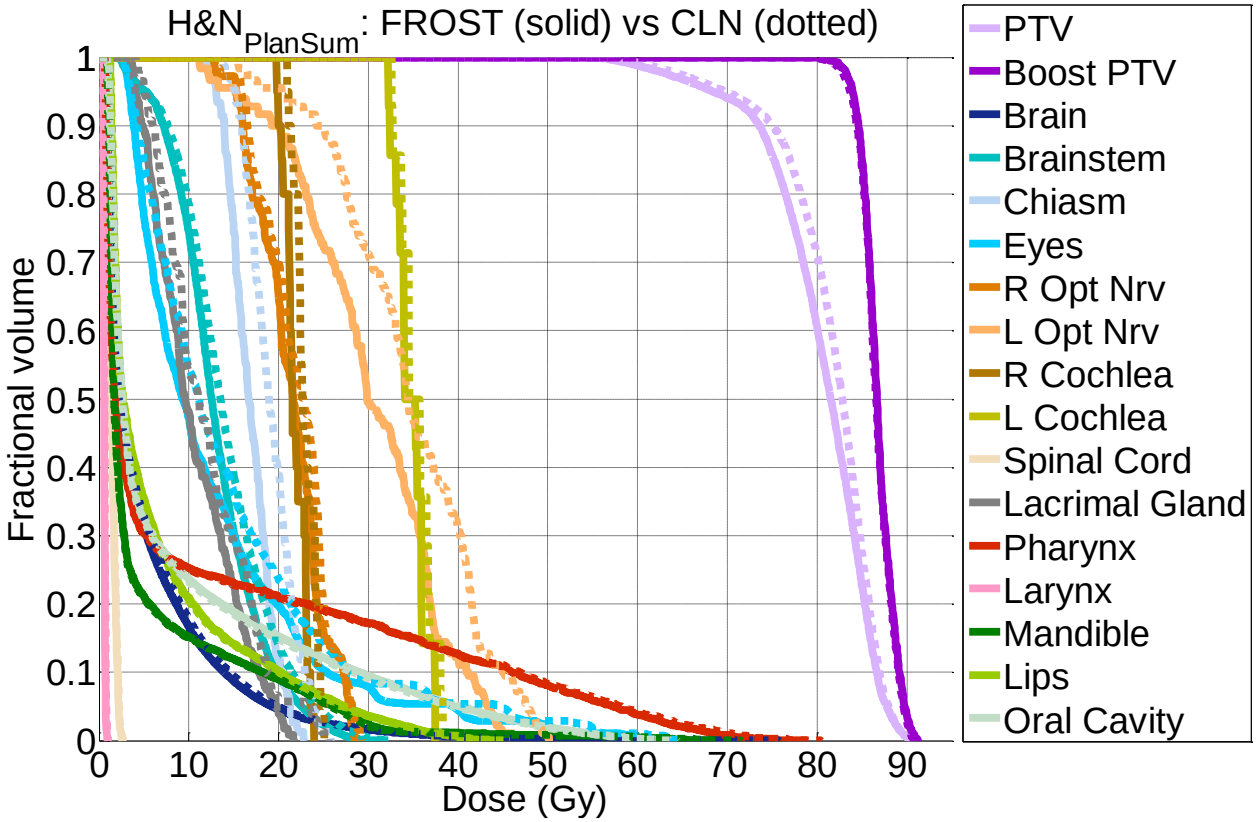

H&N PlanSum: FROST (solid) vs CLN (dotted)
Fractional volume
Dose (Gy)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
PTV
Boost PTV
Brain
Brainstem
Chiasm
Eyes
R Opt Nrv
L Opt Nrv
R Cochlea
L Cochlea
Spinal Cord
Lacrimal Gland
Pharynx
Larynx
Mandible
Lips
Oral Cavity

Figure 2.12: DVHs of all the patient cases, including the plan sum.

Figure 2.12 shows the DVHs for the 3 patients and the plan sum of the H&N case. Qualitatively, the FROST plans better spared the OARs while maintaining a comparable PTV dosimetry. For PRT, all OARs doses and the PTV dose homogeneity were improved with FROST. The $H\&N_{Boost}$ case shows essentially the same PTV dose with markedly improved dose sparing in the critical structures, including a 4 Gy reduction to the left optical nerve using the FROST plan.

Patient Case		PTV Statistics						R50	
		Homogeneity		D95	D98	D99	Dmax		
		FROST	CLN	$D_{FROST} - D_{CLN}$ (Gy)				FROST	CLN
PRT		0.945	0.938	+0.02	+0.02	–0.07	–0.34	5.954	7.909
LNG		0.840	0.834	+0.04	–0.02	+0.10	–0.01	11.467	10.587
$H\&N_{Boost}$		0.894	0.906	+0.00	+0.03	+0.16	+0.20	17.551	18.500
$H\&N_{PlanSum}$	PTV	0.779	0.794	–1.48	–1.36	–1.30	–0.03	8.829	8.631
	BoostPTV	0.936	0.934	+0.28	+0.43	+0.43	+0.18		

Table 2.11: Statistics for R50, PTV homogeneity, dose coverage (D95, D98, and D99), and Dmax.

Dose	Dmax	Dmean

Difference $D_{FROST} - D_{CLN}$ (Gy)	Largest Value	Smallest Value	Average Value	Largest Value	Smallest Value	Average Value
PRT	+0.13 Bladder	–3.96 PenileBulb	–1.42	–1.98 L Femur	–5.88 PenileBulb	–4.01
LNG	–0.29 Cord	–5.61 Bronchus	–2.27	–0.23 Cord	–3.30 Bronchus	–1.00
$H\&N_{Boost}$	+0.55 OralCavity	–4.38 Eyes	–1.41	+0.14 OralCavity	–4.37 L Opt Nrv	–0.90
$H\&N_{PlanSum}$	+0.52 Mandible	–4.48 Eyes	–1.36	+0.13 OralCavity	–4.38 L Opt Nrv	–0.90

Table 2.12: Largest, smallest, and average values found for (FROST – CLN) dose differences for Dmax and Dmean.

On average, D95, D98, and D99 between FROST and CLN methods differed by +0.04%, +0.07%, and +0.25% of the prescription dose, showing that dose coverage is virtually the same for all of the cases. The mean PTV homogeneity, between all the cases, also showed to be nearly identical with values at 0.8788 (FROST) and 0.8812 (CLN). R50, on average, decreased by 0.67 with FROST over CLN. The largest increase in max dose was the oral cavity by 0.55 Gy in the $H\&N_{Boost}$ case, but this was still a minor change compared to the 4.38 Gy that the FROST plan was able to spare the eyes. On average, FROST was able to spare Dmax and Dmean from the

OARs by 7.30% and 6.08%, respectively. More detailed statistics for the PTV and R50 are shown in Table 2.11, and for the OARs are shown in Table 2.12.

The $H\&N_{PlanSum}$ case saw similar dose sparing to the $H\&N_{Boost}$ plan. The FROST plan had better boost PTV coverage in the plan sum. Although the original PTV had a lower D95, D98, and D99 in FROST, it should be noted that the $H\&N_{Boost}$ plan did not have the original PTV in their objectives, and that the initial plan had met all of the dose coverage criteria.

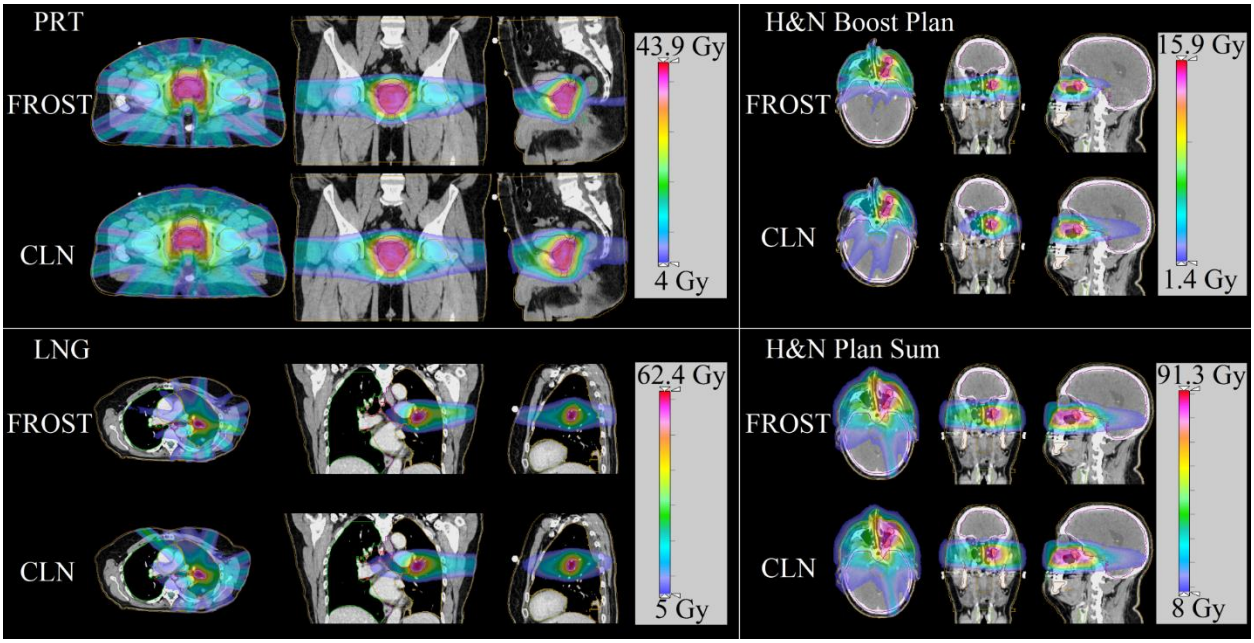


Figure 2.13: Dose washes of the 3 cases as well as well as the $H\&N_{PlanSum}$. Dose cutoff for viewing is 10% of the prescription dose.

Figure 2.13 shows the dose color wash for all of the patients. The dose distributions produced by FROST are visually different from those produced by CLN, particularly in the LNG and $H\&N_{Boost}$ cases. The LNG FROST plan spared the right lung volume from 5 Gy or greater doses, as opposed to the large dosing the anterior tip of the right lung volume in the CLN plan. Likewise, the $H\&N_{Boost}$ FROST plan clearly better spared the brain than the CLN plan. The PRT case has

fairly similar dosimetry between both the FROST and CLN plans, but improved dose conformity can be appreciated for doses above 20 Gy showing in green. The dose conformity improvement was also confirmed by the lower R50 for the FROST plan in Table 2.11.

2.2.4 Discussion

Because of the superior soft tissue contrast, MRI guided radiotherapy has the promise of improving image guided radiation therapy. It may also pave the path to longitudinal monitoring of the tumor and normal tissue response to radiotherapy using multiparametric imaging such as the diffusion MRI for cellularity measurement[69]. However, it is a significant engineering challenge to combine the therapy and imaging modalities in the same platform. The MRI-guided tri-Co-60 system uses a simpler radiation source and was able to gain early clinical access that is extremely valuable for accumulating knowledge about MRI guided RT[82-91]. There are several known limitations with Co-60 sources, including lower energy and penetration, compared to the 6 MV and above X-ray energies used by a conventional linac and large source size (2 cm) compared to the linac source (~2 mm). Relevant to the current study, in addition to these known limitations in using the Co-60 sources, treatment planning on the tri-Co-60 system is counterintuitive due to the triplet source arrangement designed to increase the dose rate, making approaching a clinically acceptable plan difficult even after tedious manual searching of beam orientations.

Another difficulty in tri-Co-60 planning is to balance the plan quality against the number of MLC segments, which substantially influence the plan delivery time due to the mechanical motion needed to shield the source between MLC segments. Conventional methods to smooth

the fluence map inevitably degrade the plan quality but the degradation is not as obvious when the number of MLC segments is not as restrictive.

The contribution of the current study is to show that the last two issues are manageable. The computerized beam triplet selection chose the optimal triplets via column generation, and FMO formulation utilized a L2-norm fidelity term to minimize the dose distribution of the prescription dose, and an anisotropic total variation regularization to encourage piecewise smoothness in the fluence maps. The column generation algorithm was inspired by the column generation method for direct aperture optimization outlined by Romeijn et al[17]. While being applied to a different problem, both algorithms use the KKT information to determine the next action to perform. The column generation approach was successfully retooled for non-coplanar beam angle selection[31,32,92-95]. The advantage of column generation, besides being efficient to solve large scale optimization problems, is to be able to integrate fluence optimization into the triplet selection. Instead of relying on human operators to select the triplet orientations, the computerized beam orientation identified optimal triplets that may seem unintuitive to human operators. The tri-Co-60 planning problem further benefits from the dose domain optimization with fluence map regularization that is able to control the dose degradation while minimizing the number of MLC segments.

Our method can be applied to other planning problems involving multiple sources at fixed geometry, such as the GammaKnife[96] and the GammaPod[97]. Although these devices are not equipped with MLC but inverse optimization utilizing available cones can be performed. The demonstrated beam orientation optimization method can be used to improve the delivery efficiency by using as many sources as possible at the same time.

2.2.5 Contribution to Deterministic DAO Approach

A **F**luence **R**egularization and **O**ptimized **S**election of **T**riplets (FROST) method for tri-Co-60 beam orientation and the fluence optimization was developed to overcome the planning challenges imposed by the multiple source geometry. Although the process is heuristic and does not guarantee optimality, the results showed that the more efficient FROST plans also yielded universally superior quality plans than manually created plans. This study's contribution to the deterministic DAO approach is in the modification to the dose fidelity term in Equation 2.9. Specifically, the structure weighting term, W, was added to control the dose distribution to the OARs. This eliminates the need to have to have an optimal and realistic dose distribution that was required in the first dose domain regularization paper.

2.3 A novel software and conceptual design of the hardware platform for intensity modulated radiation therapy[98]

2.3.1 Introduction

Intensity modulated radiation therapy (IMRT) is a cornerstone of modern radiation therapy physics. By modulating the incident X-ray intensities, IMRT has provided unprecedented control over X-ray dose distributions to target tumors and spare normal tissues. IMRT was afforded by the development of inverse optimization algorithms and enabling hardware. With a few exceptions, such as the compensator based IMRT[99-103], the multileaf collimator (MLC) that consists of a large number of thin moving tungsten leaves has been the dedicated device to modulate the X-ray fluence. Due to considerations including the mechanical complexity, cost,

accessibility and reliability of early MLCs, attempts have been made to use only the orthogonal jaws on conventional linacs for IMRT[104-106]. Jaws-only IMRT requires the entire IMRT plan to be delivered using exclusively rectangular apertures. Different methods were reported to generate these apertures, including fluence stripping[107] and direct aperture optimization (DAO)[105]. In the former method, a fluence optimization was first performed and the resulting fluence was subsequently stripped to rectangles for jaws-only delivery. The method inevitably suffered from dosimetric quality compromise and low delivery efficiency due to sequential simplification of the fluence maps. To address these difficulties, additional collimator devices such as masks[106] and rotational and dynamic jaws[108] have been theorized. The DAO approach was incorporated into a commercial planning system[105]. This approach utilized a simulated annealing process to search and select rectangular apertures that minimized an objective function. It was shown that for simple cases, the jaws-only IMRT can achieve superior plan quality to 3D conformal plans[109]. In a follow up study, Mu and Xia showed that even for more complex head and neck IMRT plans, the jaws-only approach can result in acceptable dosimetry, at a cost of longer delivery time[104].

With the maturation of MLC technology, the problems that plagued early MLC devices, such as reliability and cost, have been mitigated. However, there is still a need to miniaturize the dose modulator for purposes including small animal irradiator, robotic linac and high resolution dose modulation where reducing the leaf thickness for the application is no longer feasible, or the reduction requires significant compromise in field size, reliability and interleaf leakage. For these applications, the jaws-only IMRT may still be viable but its viability clearly depends on whether its limitations in dosimetry quality and delivery efficiency can be overcome by new

algorithms, hardware development and the presence of proper applications. Recently, we described a formulation of the inverse optimization problem that consists of an L2 fidelity terms and an L1 regularization term[28]. The solution algorithm based on the first-order primal-dual approach gave remarkable freedom to modify and simplify the optimized fluence maps. Because the optimization was performed in the dose domain, substantial modification and simplification of the fluence maps is possible without degrading the dosimetric quality. This development paves the path to a new algorithm that optimally implements the jaws-only IMRT concept on a generalized collimator.

2.3.2 Methods

2.3.2.1 Sparse Orthogonal Collimators

The SOC design consists of two orthogonally oriented collimator systems with N number of leaves in each collimator bank, where N is a small number. The term "sparse" was chosen in contrast to the "dense" leaf arrangement in a conventional MLC. For this study, we estimated the delivery efficiency of SOC having N=1, N=2, and N=4 leaves in each bank, shown in Figure 2.14. Evidently, jaws-only is a special case of SOC when $N=1$. In theory, the intensity modulation resolution of SOC is determined by the leaf step size instead of the leaf width.

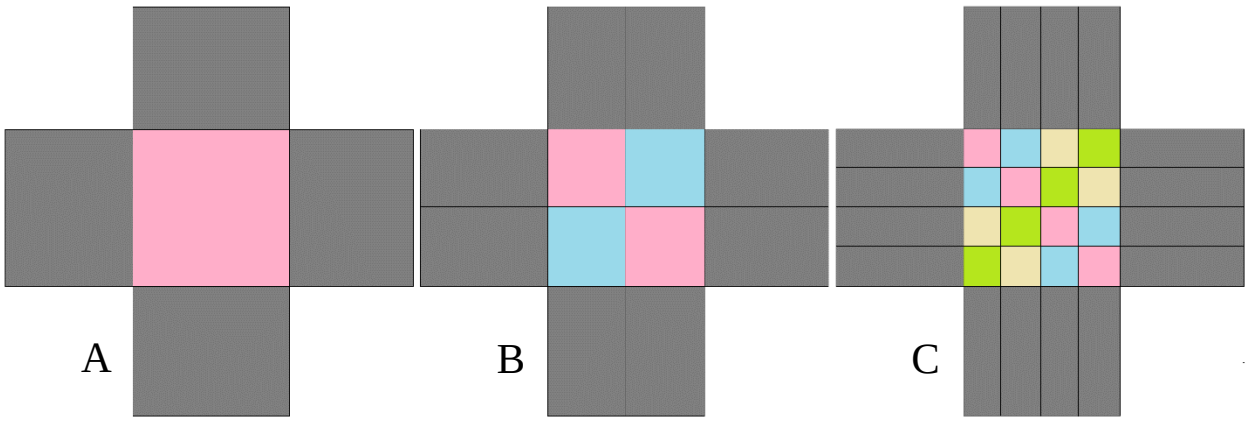


Figure 2.14: Schematic of the SOC design with *N*=1 (A), *N*=2 (B), and *N*=4(C). Gray regions are the collimator leaves. Matching colored regions indicate areas where the fluence field can be delivered in parallel.

2.3.2.2 Rectangular Basis Transform

Same as jaws-only IMRT, the apertures deliverable by SOC are rectangular. As such, development of a rectangular basis to solve the optimization was necessary. We developed a scaling function for rectangular representation, R_m. This scaling function has two main appealing properties. First, it produces a single region that is rectangular and contains one value. Second, the value of the coefficient in the rectangular basis is equal to the value of the rectangle that is produced from the coefficient when it is transformed, making easy to breakdown the final fluence to its constituent rectangles. R_m uses a coefficient set, notated as α_m. For a 2^n x 2^n fluence matrix, the coefficient matrix α_m has dimensions $(2^{n+1}-1)$ x $(2^{n+1}-1)$, and R_m has dimensions $(2^{n+1}-1)$ x 2^n. For example, if the fluence matrix is a 4 x 4 matrix then the rectangular basis transform is

$$R_m = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

Equation 2.10

and the fluence can be constructed from the basis coefficients via the equation

$$R_m^T \alpha_m R_m = x_m.$$

Equation 2.11

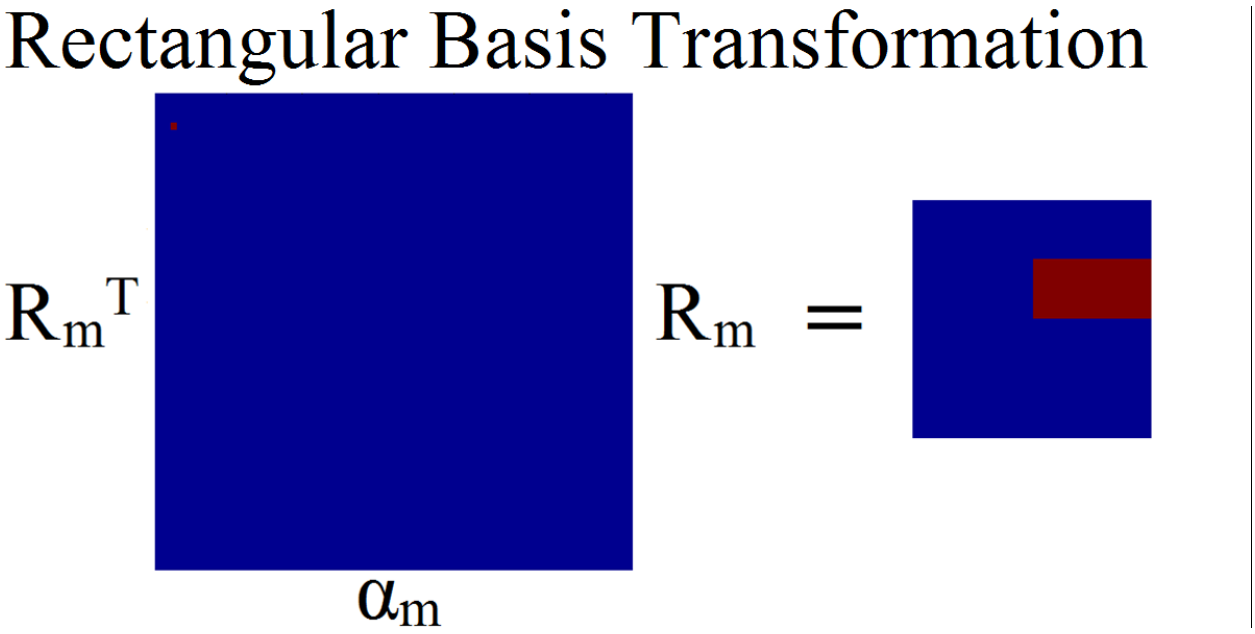


Figure 2.15: Schematic of the rectangular basis transform used in this study.

In order to prepare for optimization, the coefficient matrix α_m and fluence map x_m were vectorized, denoted as α_v and x_v respectively. To handle the vectorized notation, a special

transform matrix, R_v, that incorporates both the vertical and horizontal operations produced by R_m^T and R_m, was created. An expression to relate these variables together is

$$R_v\alpha_v = x_v = vec(x_m) = vec(R_m^T\alpha_m R_m),$$

Equation 2.12

where $vec(X)$ is the vectorization operator.

2.3.2.3 Optimization Formulation and Chambolle-Pock Algorithm

The optimization based on rectangular representation is as follows:

$$\begin{aligned} &\boldsymbol{\underset{\alpha_v}{argmin}} && \frac{1}{2}\|\boldsymbol{W}(\boldsymbol{AR}_v\boldsymbol{\alpha}_v - \boldsymbol{d})\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{\alpha}_v\|_1 \\ &\boldsymbol{subject\ to} && \boldsymbol{\alpha}_v \geq \mathbf{0}, \end{aligned}$$

Equation 2.13

where $\alpha_v = vec(\alpha_m)$ is the optimization variable. R_v is the rectangular transform matrix for the coefficient vector α_v. W is a weighting factor for the structures of interest, A is the fluence to dose transformation matrix, and d is the desired dose, which is set to the prescription dose for the PTV and zero for the OARs. The fluence to dose transformation matrix is calculated using a convolution/superposition code using a 6 MV x-ray polyenergetic kernel.

The L2-norm data fidelity term penalizes any deviations in the dose from d, and the L1-norm sparsity term encourages a low number of non-zero coefficients. The definition of norm, for some vector x of length n, is $\|x\|_p = \sqrt[p]{\sum_{i=1}^{n}|x_i|^p}$. The number of non-zero coefficients is exactly equal to the number of fluence segments for delivery. The weighting parameter λ is used to regulate the number of non-zero coefficients by weighting the importance of the L1-norm in the cost function.

Constraining α_v to stay positive is possible since the rectangular transform was constructed such that the value of the coefficients equals the value of the fluence rectangle. The optimization problem was solved utilizing the Chambolle-Pock algorithm, a first-order primal-dual algorithm[20], which is found in the Appendix. Due the convex nature of the optimization problem, the starting values α_v^0 and z^0 can be set as anything within the boundary constraints. For this study, $\alpha_v^0 = 0$ and z^0=0.

2.3.2.4 Expanding the Rectangular Basis for Optimal Delivery Efficiency

It is recognized that the current formulation of R_v does not contain of the every possible rectangle for the matrix. The affect the dosimetry result is minimal, since the current rectangular basis is overcomplete, and every possible rectangle can be expressed as a linear combination of the existing basis rectangles. However, this can affect the delivery efficiency, causing a longer delivery time with more apertures. The increased computational cost of including every possible rectangle is enormous. A 64 x 64 fluence grid would have 86528000 variables to solve for 20 beams, which is much larger than the 322580 variables found with the reduced basis. To address

this issue, a method for adding new rectangles to the current rectangle library has been implemented. This method allows for the selection of rectangles that do not currently exist in the original rectangular basis, while keeping increases to computational to a minimum by only adding in new rectangles that are probable to be to chosen. The method is as follows:

1) Start with original rectangular basis R_v
2) Optimize equation 6 with Chambolle-Pock algorithm
3) Find new rectangular apertures based on the selected apertures from step 2
    a. Take every pairwise combination of selected apertures
    b. For each pair, perform a union of the aperture shapes to generate a new aperture
    c. New aperture passes if it fits 2 criteria
        i. New aperture is rectangular
        ii. New aperture is unique to current library of apertures
4) Update R_v for all new apertures that pass in part 3
5) Repeat steps 2 through 4 until no new rectangular apertures are found in step 3

Due to the fact that the original rectangular basis was already overcomplete and had access to the individual beamlets, adding the new step will not affect the dosimetry, but instead it will reduce the number of selected apertures for delivery and increase delivery efficiency.

2.3.2.5 Evaluation

2.3.2.5.1 Patient studies

Seven previously treated patients consisting of 2 glioblastoma multiforme patients (GBM), 3 head & neck patients (H&N), including 2 SBRT patients and 1 conventional fractionated patient with 3 PTVs ($H\&N_{3PTV}$), and 2 lung patients (LNG) were included in the study. The prescription doses and PTV volumes are shown in Table 2.13. The patients were first planned on the 4π radiotherapy platform to optimize both beam orientation and fluence maps using previously described column generation and pricing approach[76,110]. During optimization, a 5 cm ring structure was added around the PTV to penalize the dose spillage to normal tissue. The platform has been shown to achieve superior dosimetry to coplanar volumetric modulated arc therapy (VMAT)[71-73]. As a result of the optimization, 20 non-coplanar beams were selected from a candidate pool of 1162 equally spaced non-coplanar beams for each patient.

	Prescription Dose (Gy)	PTV Volume (cc)
GBM #1	25	6.23
GBM #2	30	57.77
H&N #1	40	23.76
H&N #2	40	18.86
Lung #1	50	138.75
Lung #2	50	47.78

$H\&N_{3PTV}$	54.00	197.54
	59.40	432.56
	69.96	254.98

Table 2.13: Prescription doses and PTV volumes for each of the cases.

The dose distributions were compared after generating deliverable segments on the two different platforms: 1) the Rectangular Aperture Optimization (RAO) method and 2) the direct MLC segmentation (DMS) method.

For the RAO plans, the optimized fluences from the 4π radiotherapy plan were not used while adopting the optimized beam angles. Rectangular apertures were calculated using equation 6 and the Chambolle-Pock algorithm. λ was varied until the average number of segments per beam for each case was roughly 15.

For the DMS method, the raw fluence from the 4π radiotherapy plan was stratified and an MLC segmentation algorithm was applied to calculate the deliverable MLC segments. The MLC segmentation algorithm was based on a reduction level method by Xia and Verhey[6], described previously in detail[28]. The stratification step size was adjusted through a bisection algorithm so that the number of calculated MLC segments in DMS equaled to that of RAO, and the MLC segments were calculated to be deliverable along the direction of leaf motion without any collimator rotations.

For all of the cases except the $H\&N_{3PTV}$ RAO plan, the beamlet resolution was 0.5 cm^2 and the dose matrix resolution was 0.25 cm^3. Due to computational complexities of the optimization, the $H\&N_{3PTV}$ RAO plan was evaluated with a beamlet resolution of 1 cm^2 and dose matrix resolution of 0.5 cm^3.

All treatment plans were normalized such that the prescription dose was delivered to 95% of the PTV. As dosimetric endpoints for comparison, R50 and PTV homogeneity, D98, D99, and Dmax, as well as OAR mean and max doses, were evaluated. R50 is a measure of high dose spillage, and is defined as the 50% isodose volume divided by the PTV volume. Homogeneity is defined as $\frac{D95}{D5}$, and maximum dose is defined from ICRU-83[58], where Dmax is defined as D2, the dose to 2% of the structure's volume.

Aperture size, reported as number of bixels, is evaluated and compared between the RAO and DMS methods for all 7 cases. A bixel is a basic square unit on an IMRT fluence map, and aperture continuity is defined as the four directly neighboring bixels. The RAO method, by design, is limited to 1 aperture per delivery segment, while the DMS method can have multiple apertures per segment.

2.3.2.5.2 Sparse orthogonal collimator travel time estimation

In order to evaluate the potential delivery time for using SOCs with N=1, N=2 and N=4 leaves per bank, the delivery order of the segments must first be optimized. The problem was formulated as a modified form of the open traveling salesman problem (TSP), and then solved using a basic genetic algorithm, an open source MATLAB code provided by Kirk[111].

The general TSP attempts to find the shortest path through a set of points in space, traveling to each point only once. The open variation of the problem allows for the end point to be different than the start point, rather than forming a closed loop. As a genetic algorithm, the open source code solves the TSP by first generating random populations, which are different potential routes that visit all points exactly once. The genetic algorithm then groups the paths into random groups of 4. For each group, it takes the population with the shortest route, randomly selects 2 of the points along the route, and performs 3 types of mutations on the sub-route between the 2 points: 1) Swap: The order of the 2 points is swapped along the sub-route 2) Flip: The sub-route existing between the 2 selected points is reversed 3) Slide: All points between and including the 2 selected points are shifted by 1, with the first point becoming the last point on the sub-route. These 3 mutations replace the 3 worst solutions of the group of 4. This is performed for every grouping and repeated for a set number of iterations, selecting a different random grouping of 4 each time. After a set number of iterations, an optimal or near-optimal solution is found.

For our problem, each rectangular segment can be defined as 4 numbers describing the location of the collimator edge. In a sense, each rectangular segment can be mathematically described as

a 4-dimensional point. Subtracting two of these points tells us how far each collimator has to travel from one segment to another. The open source code by Kirk can solve the TSP in any number of dimensions by finding the Euclidean norm distance between two points. Since the limiting factor in travel time between two segments is defined by the one collimator leaf that has to travel the furthest, the code was modified such that instead of using the Euclidean distance between two points, it used the single largest distance a collimator leaf had to travel for calculation.

In this study, 100 populations and 10000 iterations were used in the genetic algorithm to solve the TSP.

Once the order of the segments is solved, the collimator travel time was calculated with an estimated jaw acceleration of 5 cm/s^2 and maximal speed of 2 cm/s, which is typical for a C-arm linac.

The N=2 and N=4 SOC designs have color coded regions shown in figure 1 that can be delivered in parallel. To account for this, the segments are grouped into regions and an individual collimator travel time for each region is optimized and calculated separately. Between regions that can be simultaneously delivered, the longest travel time is recorded. When using SOC with N>1 to deliver a larger aperture, adjacent leaves move together as a single leaf.

The same aperture ordering scheme and time calculation, using the same jaw acceleration and maximal speed, was performed for DMS as well, with each aperture described as the location of every MLC location.

The total travel time for delivery and efficiency is evaluated for the 7 patients and compared between DMS and RAO with the N=1, N=2 and N=4 leaf collimator designs.

2.3.3 Result

2.3.3.1 Fluence Map Evaluation

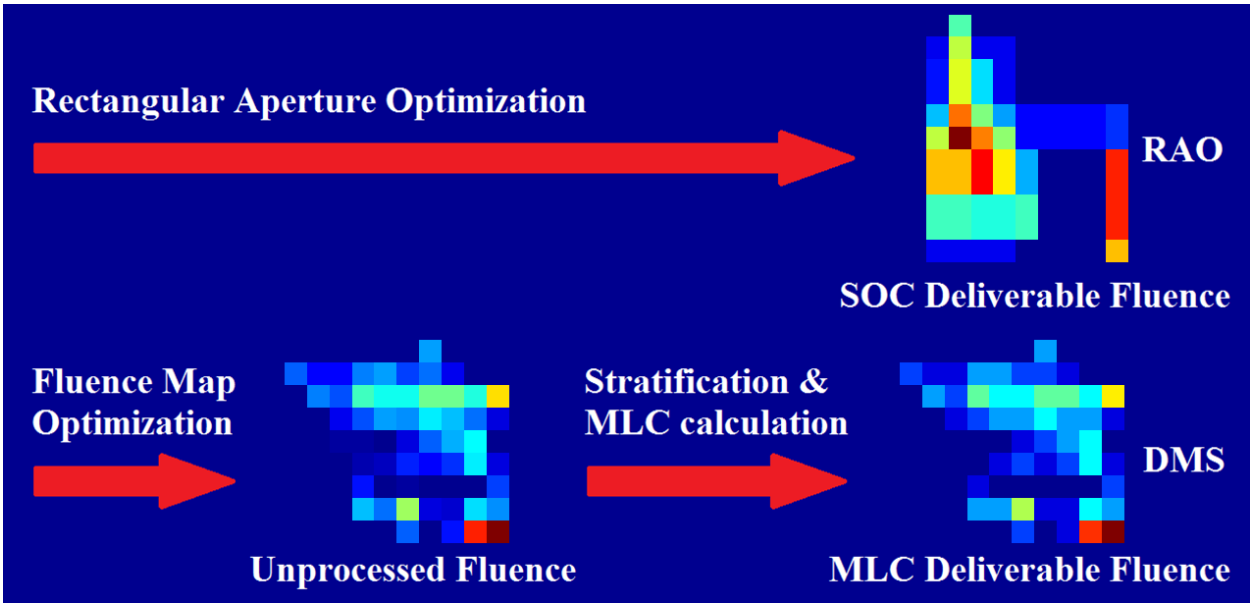


Figure 2.16: Schematic of fluence maps produced by RAO and DMS method for the same beam angle.

Figure 2.16 is a schematic of the fluences from the same beam angle and the MLC segmentation steps needed to create deliverable fluences. The DMS method requires extra post-optimization processing to stratify the fluence, which changes the optimized fluence and degrades the dosimetry. The RAO method, on the other hand, creates rectangles that can be delivered without any further processing of the fluence. Despite delivery from the same beam angle, the fluence patterns from the two methods are substantially different for most beam angles. This difference can be explained by the fact that RAO is a direct aperture regularization approach centered on the dose domain optimization. It utilizes the sparsity term to limit the total number of coefficients, and therefore, limit the total number of apertures. RAO plans included the aperture constraints *during* the optimization stage while DMS incorporate the MLC constraints *after* the optimization.

Figure 2.17A illustrates the transform from rectangular basis coefficients, our optimization variable, to the fluence domain, and Figure 2.17B shows the breakdown of the fluence map into its deliverable rectangles, which correspond directly to the coefficients. On average, approximately 16 delivery segments per beam were used.

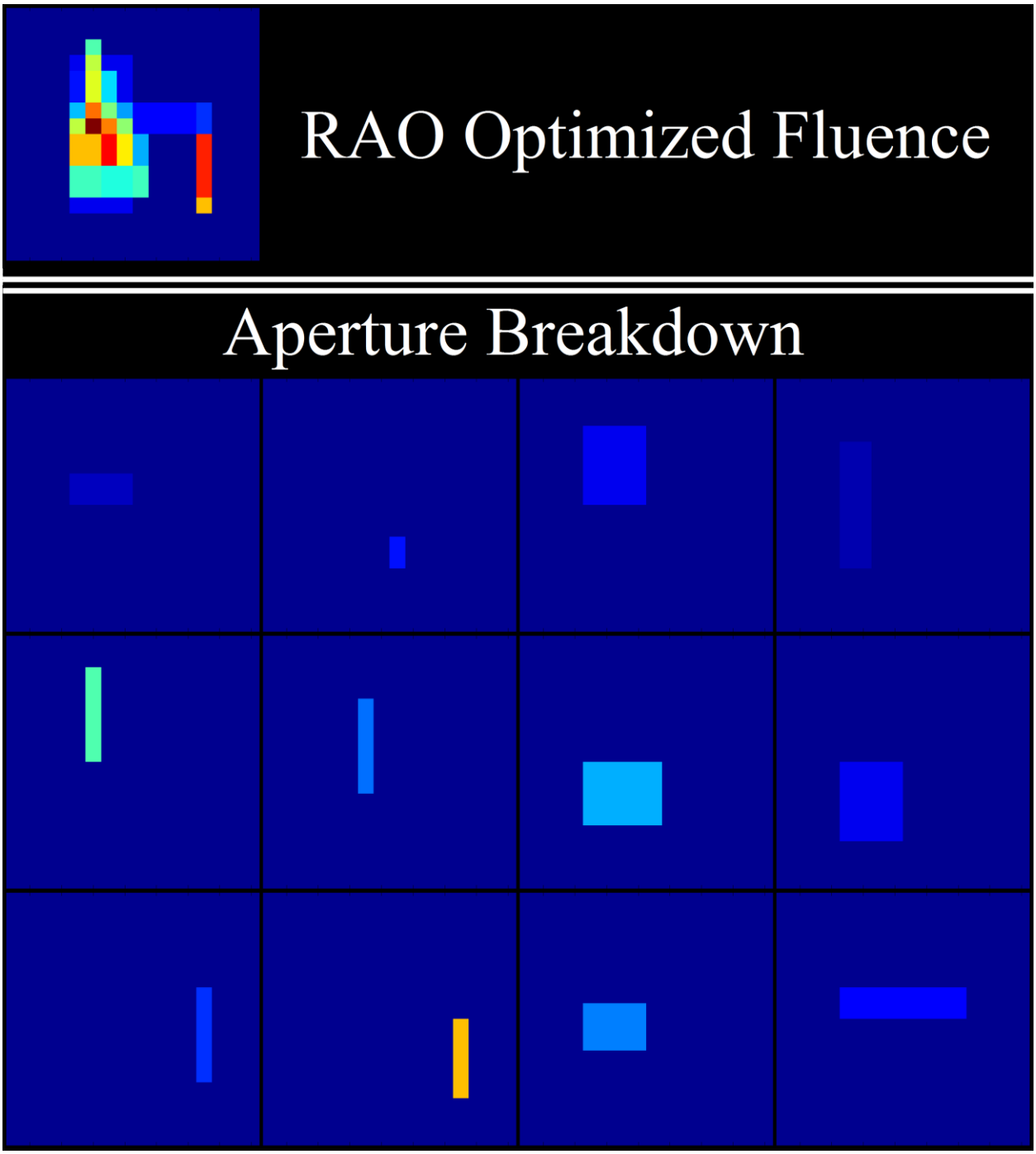
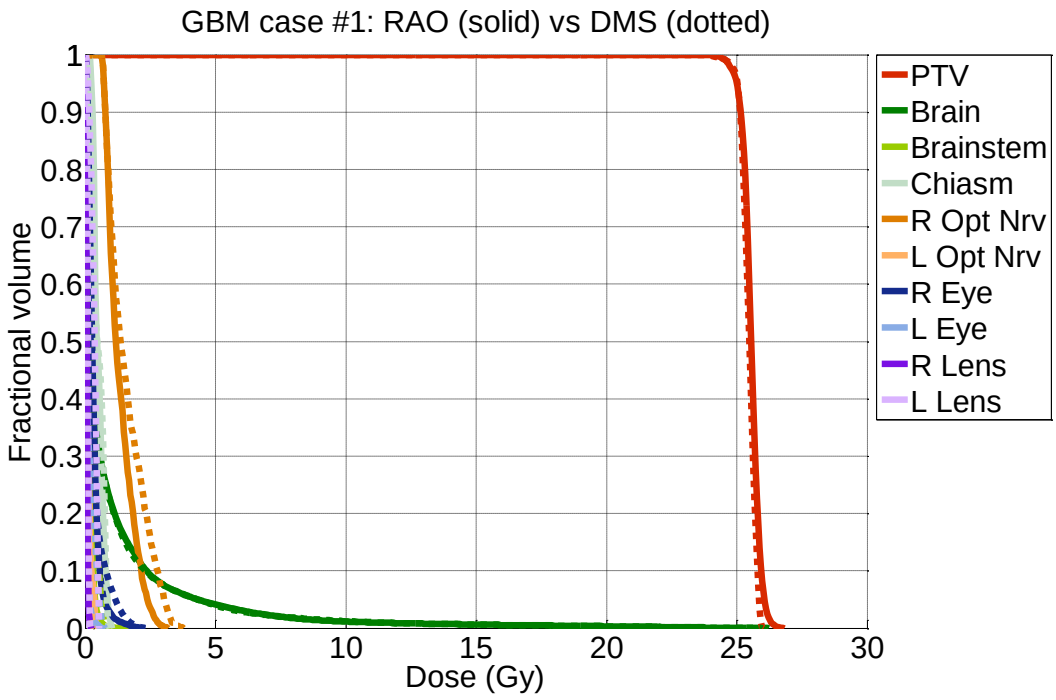
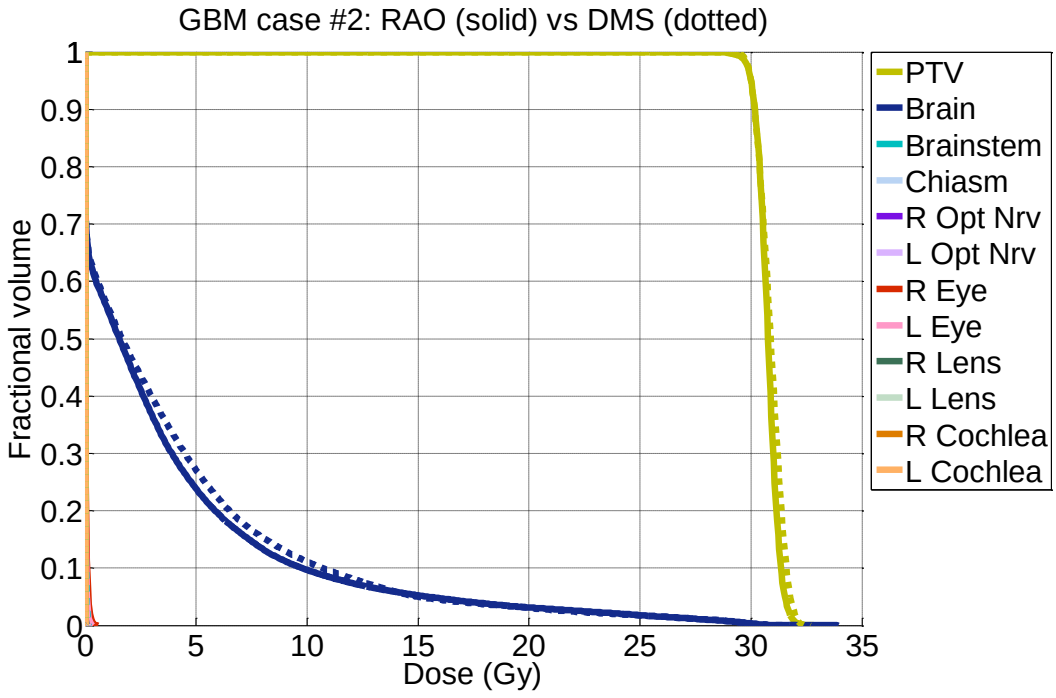
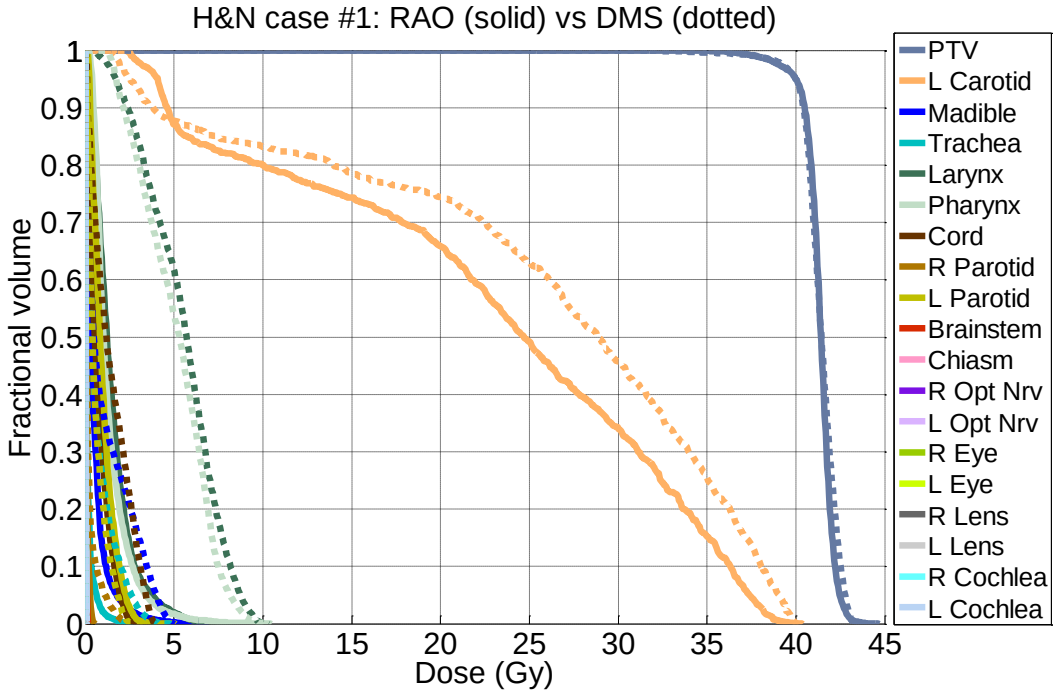


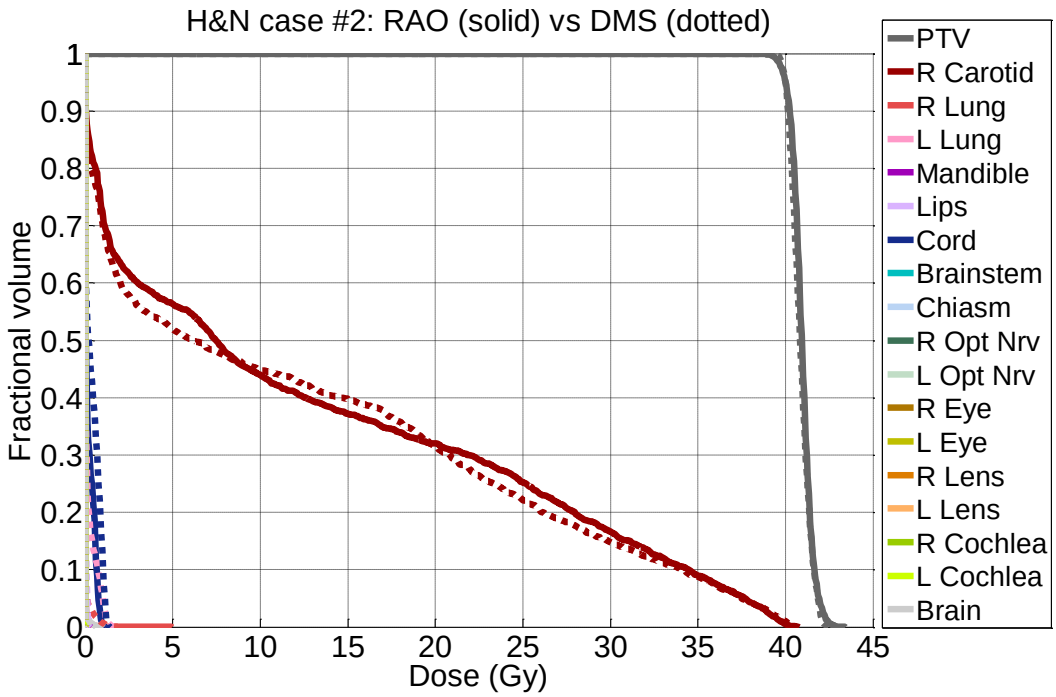
Figure 2.17: Schematic of an optimized fluence and the breakdown into its deliverable rectangular apertures.

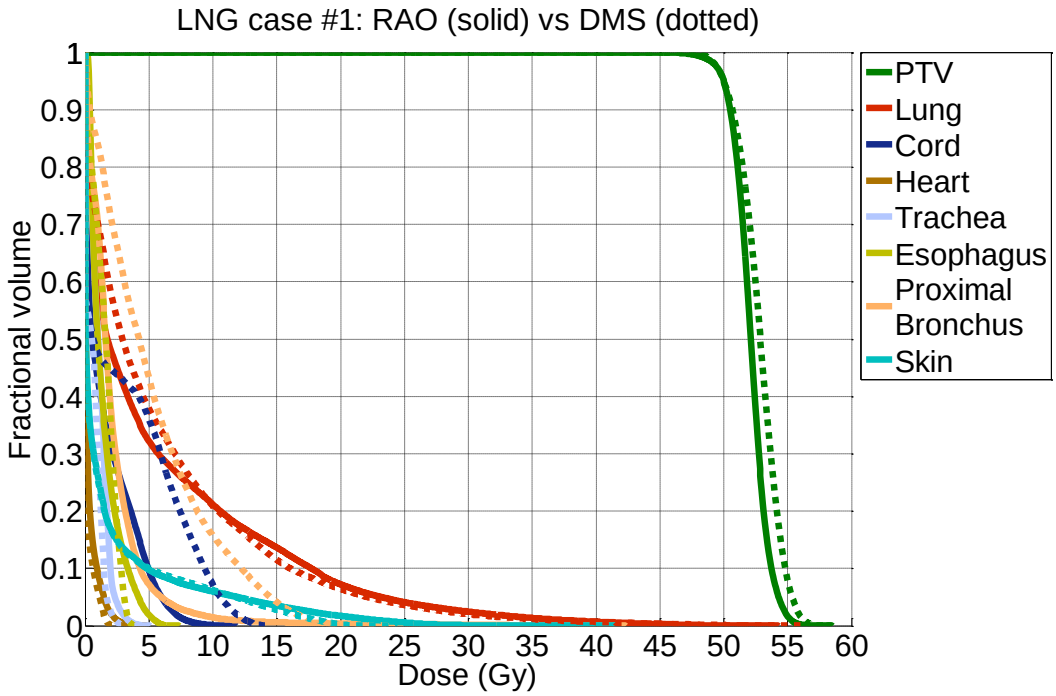
2.3.3.2 Patient Results

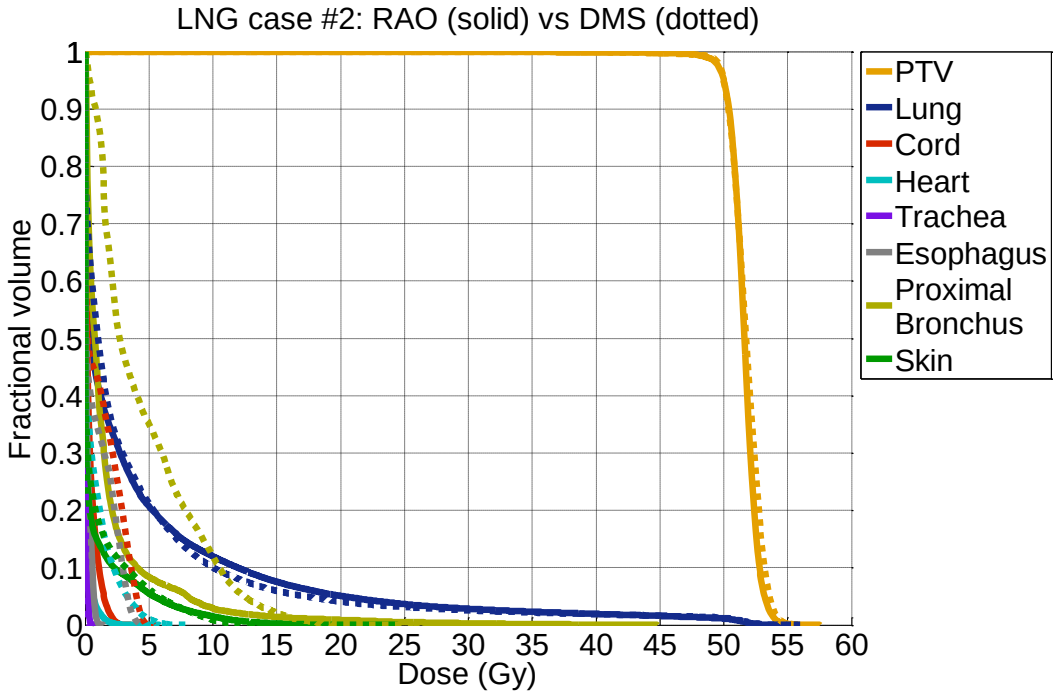





H&N case #1: RAO (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
PTV
L Carotid
Madible
Trachea
Larynx
Pharynx
Cord
R Parotid
L Parotid
Brainstem
Chiasm
R Opt Nrv
L Opt Nrv
R Eye
L Eye
R Lens
L Lens
R Cochlea
L Cochlea


H&N case #2: RAO (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
PTV
R Carotid
R Lung
L Lung
Mandible
Lips
Cord
Brainstem
Chiasm
R Opt Nrv
L Opt Nrv
R Eye
L Eye
R Lens
L Lens
R Cochlea
L Cochlea
Brain


LNG case #1: RAO (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
PTV
Lung
Cord
Heart
Trachea
Esophagus
Proximal Bronchus
Skin


LNG case #2: RAO (solid) vs DMS (dotted)
Fractional volume
Dose (Gy)
PTV
Lung
Cord
Heart
Trachea
Esophagus
Proximal Bronchus
Skin

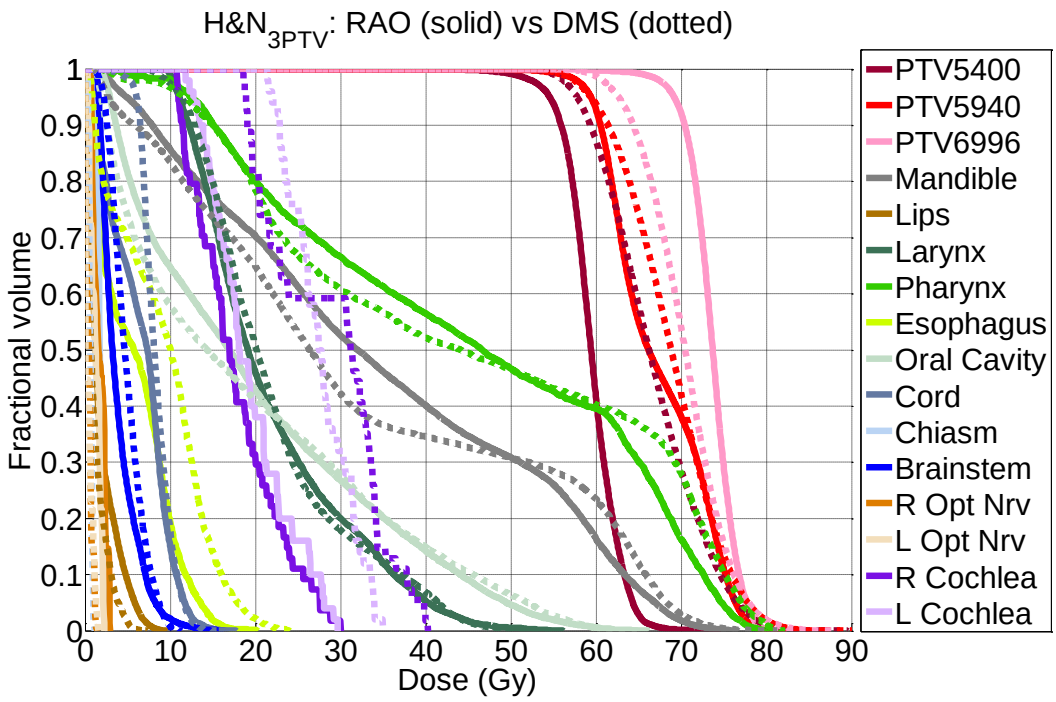


Figure 2.18: DVH comparisons of the GBM, H&N, LNG, and H&N$_{3PTV}$ cases.

Patient Case	Average number of segments per beam	PTV Statistics					R50	
		Homogeneity		D98	D99	Dmax		
		RAO	DMS	RAO – DMS (Gy)			RAO	DMS
GBM #1	13.55	0.959	0.968	–0.072	–0.029	+0.309	3.418	3.505
GBM #2	12.70	0.952	0.944	–0.029	–0.100	–0.223	2.636	2.517
H&N #1	12.45	0.942	0.935	–0.219	–0.083	–0.167	3.857	3.088
H&N #2	10.90	0.954	0.955	–0.119	–0.235	+0.213	3.961	3.549

Lung #1	11.85	0.921	0.904	–0.053	–0.309	–1.041	4.192	3.466
Lung #2	8.35	0.940	0.933	–0.100	–0.060	–0.272	3.961	3.381
Average excluding $H\&N_{3PTV}$	11.63	0.945	0.940	–0.099	–0.136	–0.197	3.671	3.251
$H\&N_{3PTV}$	54.00	18.95	0.848	0.754	–3.832	–4.665	–13.309	10.699	9.798
	59.40		0.778	0.763	+0.294	+0.079	–2.270		
	69.96		0.891	0.804	+6.662	+6.077	–1.436		
Total Average	12.68	0.909	0.884	+0.281	+0.075	–2.022	4.675	4.186

Table 2.14: Comparison of PTV homogeneity, D98, D99, and Dmax as well as R50 and average number of delivery segments. Total average includes the PTVs from the 6 single target cases and the 59.4 Gy PTV from the $H\&N_{3PTV}$ case. R50 from the $H\&N_{3PTV}$ case is calculated based on the 59.4 Gy prescription dose and the total PTV volume contributed by all 3.

Dose difference	Dmax			Dmean		
	Largest	Smallest	Average	Largest value	Smallest	Average

RAO–DMS (Gy)	value	value	value		value	value
GBM #1	+0.123 L Opt Nrv	–0.609 R Opt Nrv	–0.216	+0.023 L Opt Nrv	–0.245 R Opt Nrv	–0.080
GBM #2	+0.200 R Eye	–0.005 Brainstem	+0.051	+0.023 R Eye	+0.0002 L Opt Nrv	+0.006
H&N #1	+0.546 L Paroid	–4.325 Larynx	–1.570	+0.368 L Parotid	–3.959 Larynx	–1.197
H&N #2	+0.070 Brain	–1.097 L Lung	–0.288	+0.364 R Carotid	–0.189 Spinal Cord	–0.002
Lung #1	+2.698 Skin	–7.737 ProxBronch	–0.695	+0.489 Trachea	–3.117 ProxBronch	–0.640
Lung #2	+1.188 Lung	–3.208 Heart	–1.360	+0.076 Lung	–2.681 ProxBronch	–0.687
$H\&N_{3PTV}$	+1.992 Lips	–11.150 R Cochlea	–1.486	+1.801 Mandible	–10.8611 R Cochlea	–1.475

Table 2.15: Largest and smallest values found for (RAO – DMS) dose differences for the Dmax and Dmean. The average value of the dose differences between OARs for each case is included. OARs that received 0 Gy in both the RAO and DMS cases are excluded in the evaluation.

Figure 2.18 shows the DVHs for all cases. The average difference (RAO – DMS) of D98, D99, and Dmax for the PTV between all cases, using only the 59.4 Gy PTV from the $H\&N_{3PTV}$ case, was –0.149%, –0.264%, and –0.834%, respectively, as a percent of the prescription dose, indicating equivalent PTV coverage. The PTV homogeneity, on average, is 0.025 higher with RAO than with DMS. On the other hand, RAO increased R50 by 0.489, indicating a slightly increased high dose spillage to the body.

The $H\&N_{3PTV}$ case had severe dose degradation in the DMS case after MLC segmentation. Before this post processing step, the plan was able to achieve a D95 of 55.3 Gy, 59.4 Gy, and 68.8 Gy for the 54 Gy PTV, 59.4 Gy PTV, and 69.96 Gy PTV, respectively. However, segmenting the large fluence maps into a relatively small number of deliverable segments caused the dose to degrade, and after normalization to the 59.4 Gy PTV, the D95 changed to 57.8 Gy, 59.4 Gy, and 62.6 Gy, respectively. The higher prescription level PTVs had large dose degradation, and the dose normalization caused overdosing to the 54 Gy PTV by an extra 3.8 Gy.

Due to the unacceptable dosimetry to the target volumes of the $H\&N_{3PTV}$ caused by the DMS method, another plan, termed the dose domain regularized (DDR) plan[28], was included for comparison. DDR piecewise smooths the fluence using an anisotropic total variation regularization term while penalizing deviations from the optimal dose calculated from the 4π radiotherapy plan but before any post processing from the DMS plan. The DDR plan then underwent the same stratification and segmentation steps as the DMS plan. Even though the DDR plans and the DMS plans originated from the same exact 4π optimized plan, the DDR plan suffers much less from the stratification and segmentation steps, allowing for better dose

coverage and homogeneity to the 69.96 Gy PTV than the RAO plan. A DVH comparison between RAO and DDR is shown in Figure 2.19.

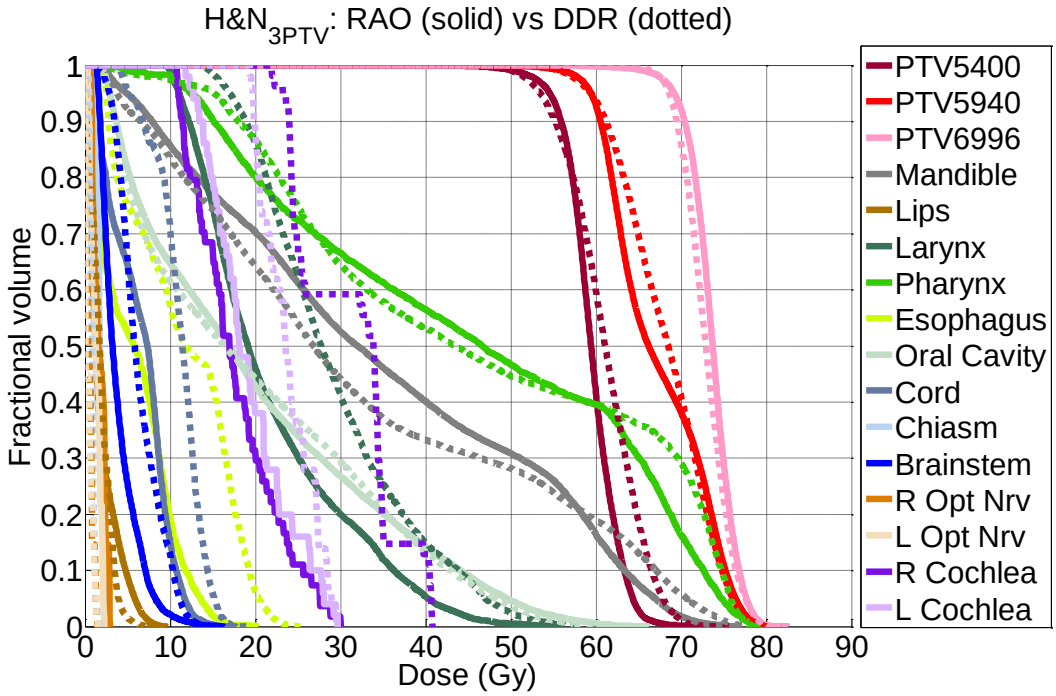
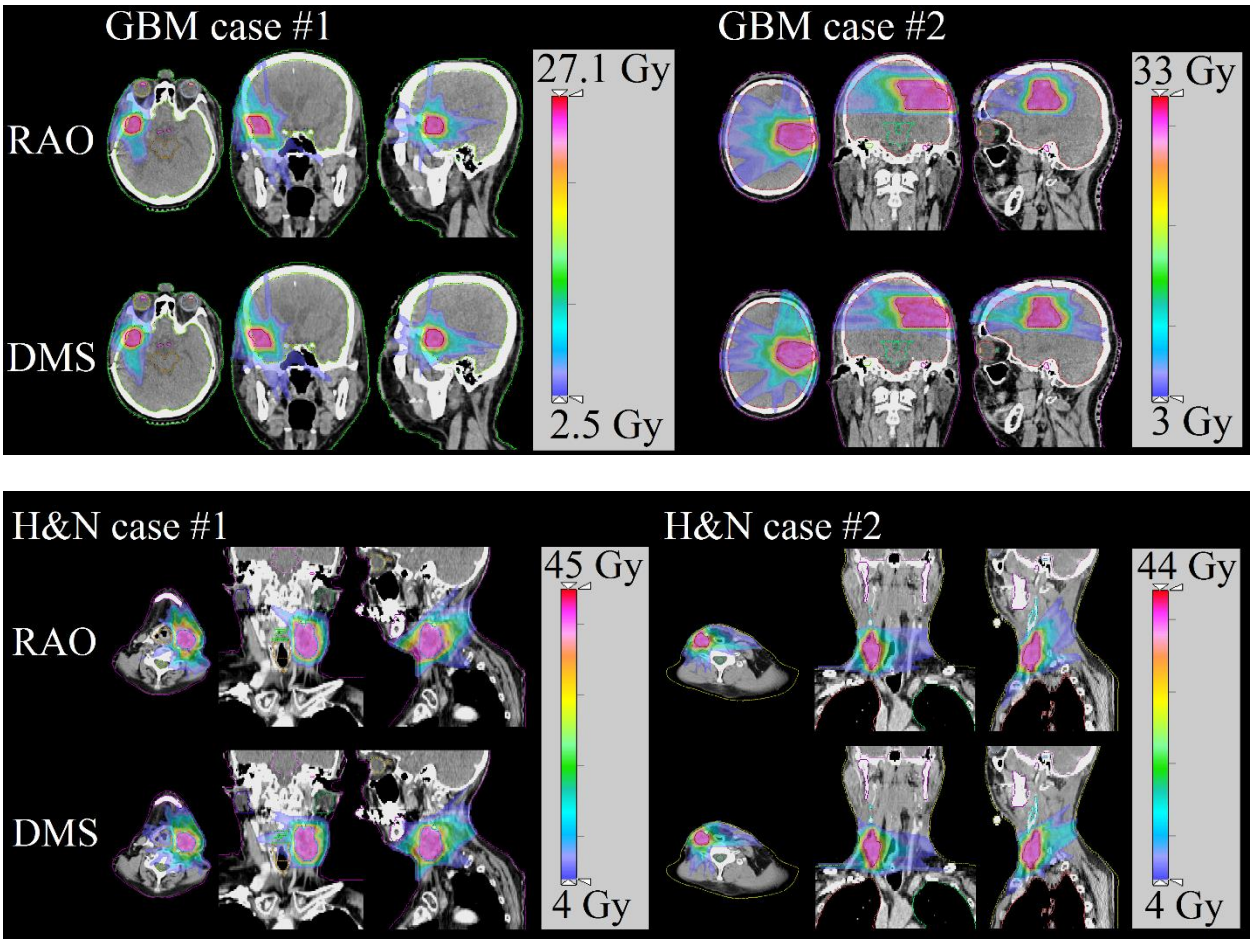
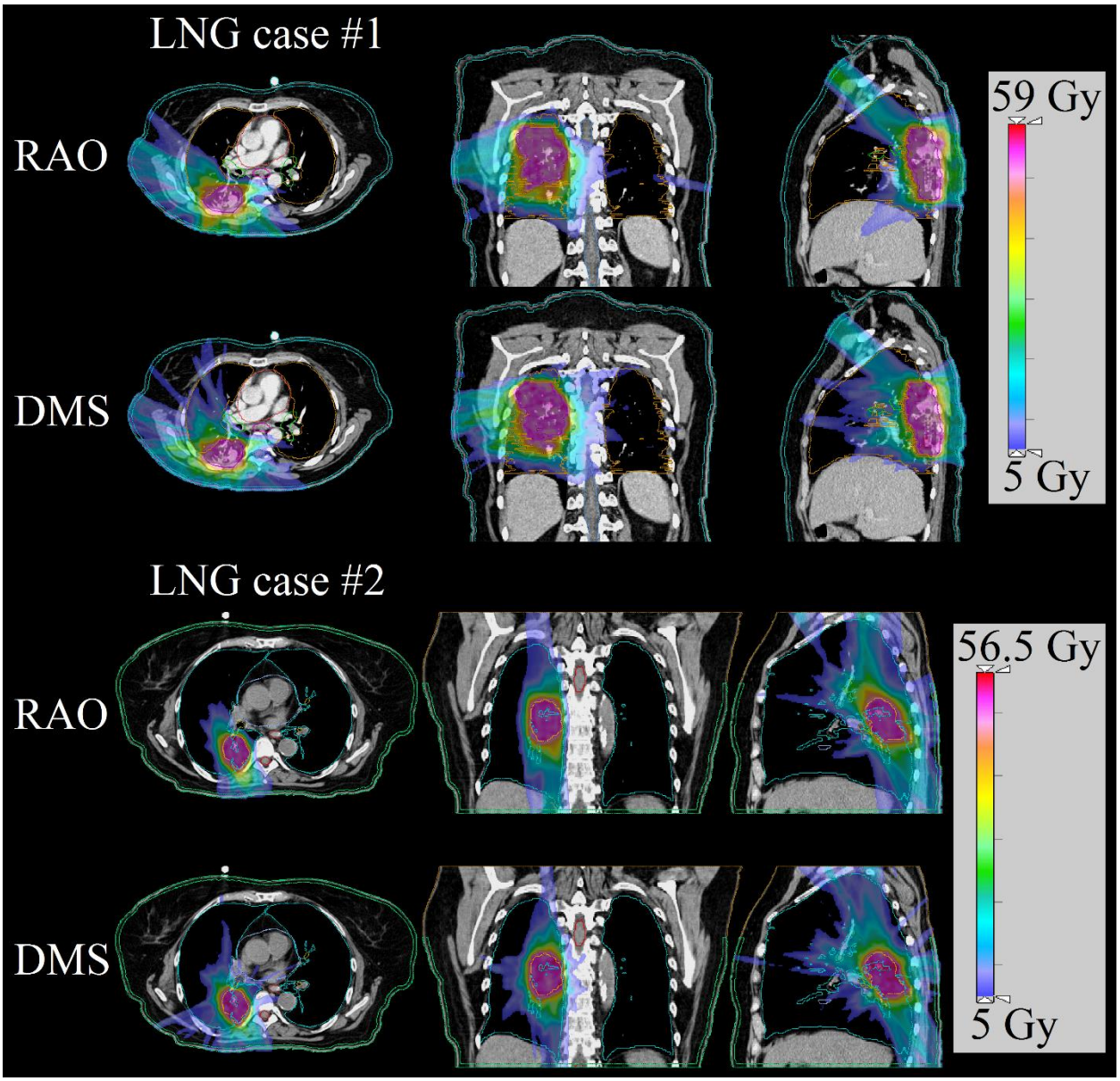


Figure 2.19: DVH comparison between RAO and DDR methods for H&N$_{3PTV}$ case.

This average difference for maximum and mean dose for all OARs between all plans, excluding structures that received zero dose from both plans, are –1.94% and –1.44%, respectively, as a percent of the prescription dose, meaning that for typical radiotherapy plans that have a prescription dose ranging from 30 to 60 Gy, we can expect to spare from 0.582 Gy to 1.164Gy of max dose and from 0.432 Gy to 0.864 Gy of mean dose to OARs.

Figure 2.20 shows the dose color washes for 7 patient cases. Both methods achieved high dose conformality with comparable PTV coverage and slightly different normal organ dose distributions.




LNG case #1
RAO
DMS
59 Gy
5 Gy
LNG case #2
RAO
DMS
56.5 Gy
5 Gy

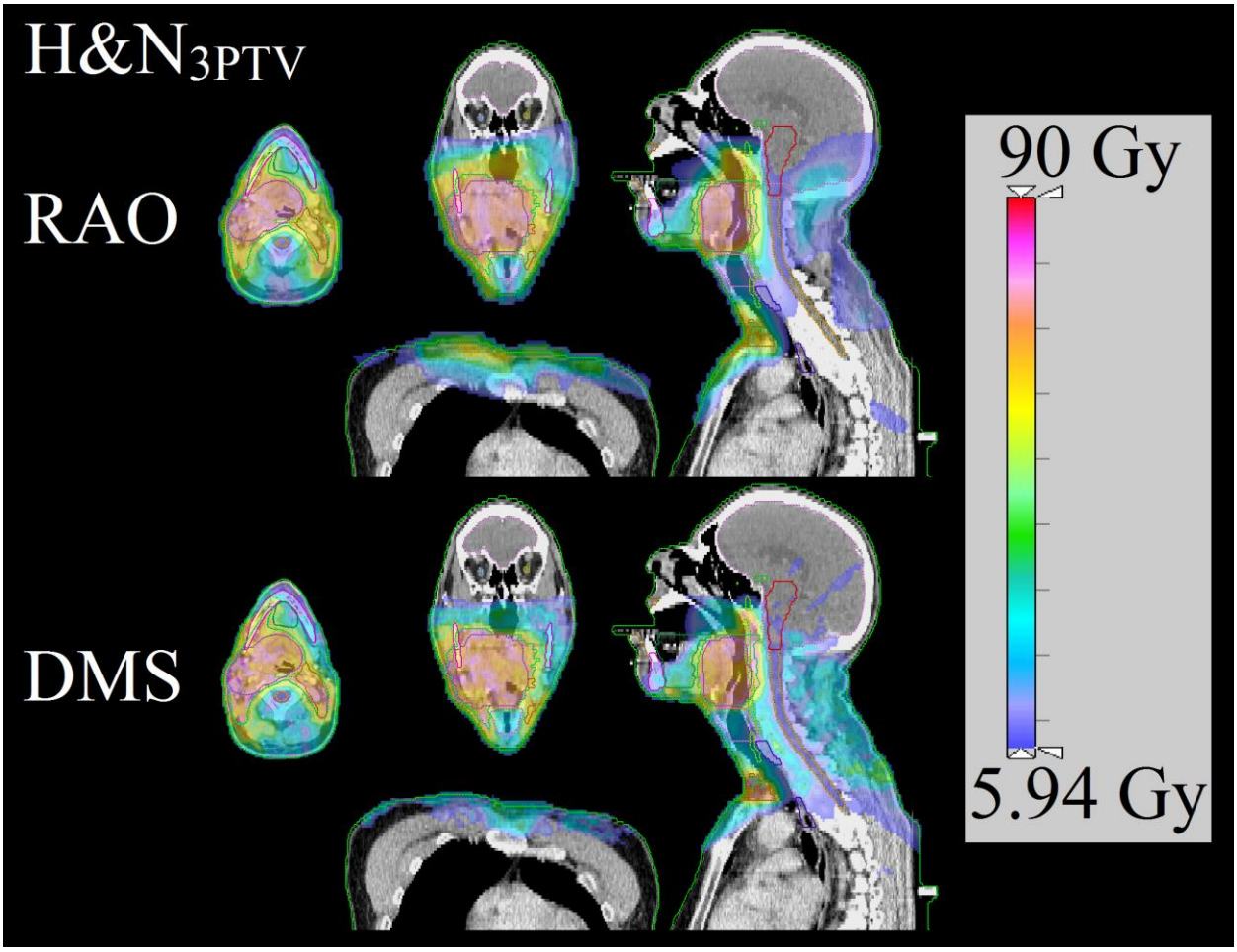


Figure 2.20: Dose color washes of A) GBM, B) H&N, C) LNG, D) H&N$_{3PTV}$ patients. The dose cutoff for viewing was chosen to be 10% of the prescription dose.

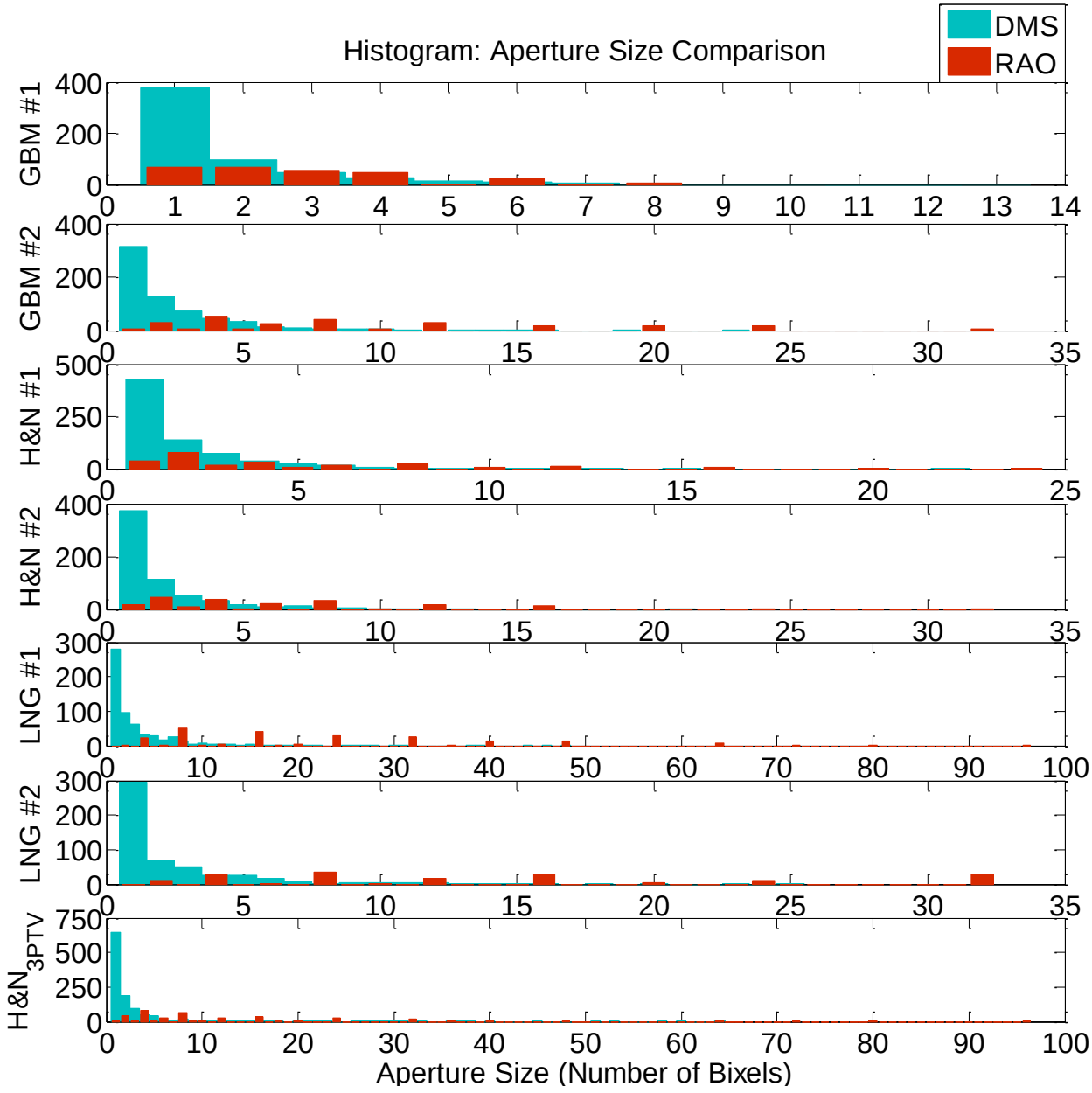


Figure 2.21: Histograms comparing the aperture sizes of the RAO and DMS method for each of the 7 cases.

	Total Number of Apertures	Mean Aperture Size (Number of Bixels)	Max Aperture Size (Number of Bixels)

	RAO	DMS	RAO	DMS	RAO	DMS
GBM #1	271	587	2.79	1.86	8	13
GBM #2	254	643	9.37	2.45	32	23
H&N #1	249	743	5.06	2.06	24	22
H&N #2	218	650	5.88	2.22	32	21
LNG #1	237	619	22.52	3.96	96	48
LNG #2	167	525	14.29	2.71	32	25
$H\&N_{3PTV}$	379	1144	12.96	3.24	96	64
Average	253.57	701.57	10.41	2.64	45.71	30.86

Table 2.16: Aperture statistics for the RAO and DMS methods for the 7 cases

Limited to one aperture per delivery segment, the RAO plan produced far fewer apertures than the DMS method, which had, on average, about 3 apertures per segment. The mean aperture size for RAO is roughly 3.9 times larger than the mean aperture size for DMS. The maximum aperture size is about 15 bixels larger using RAO comparing to DMS. The aperture statistics are shown in Table 2.16.

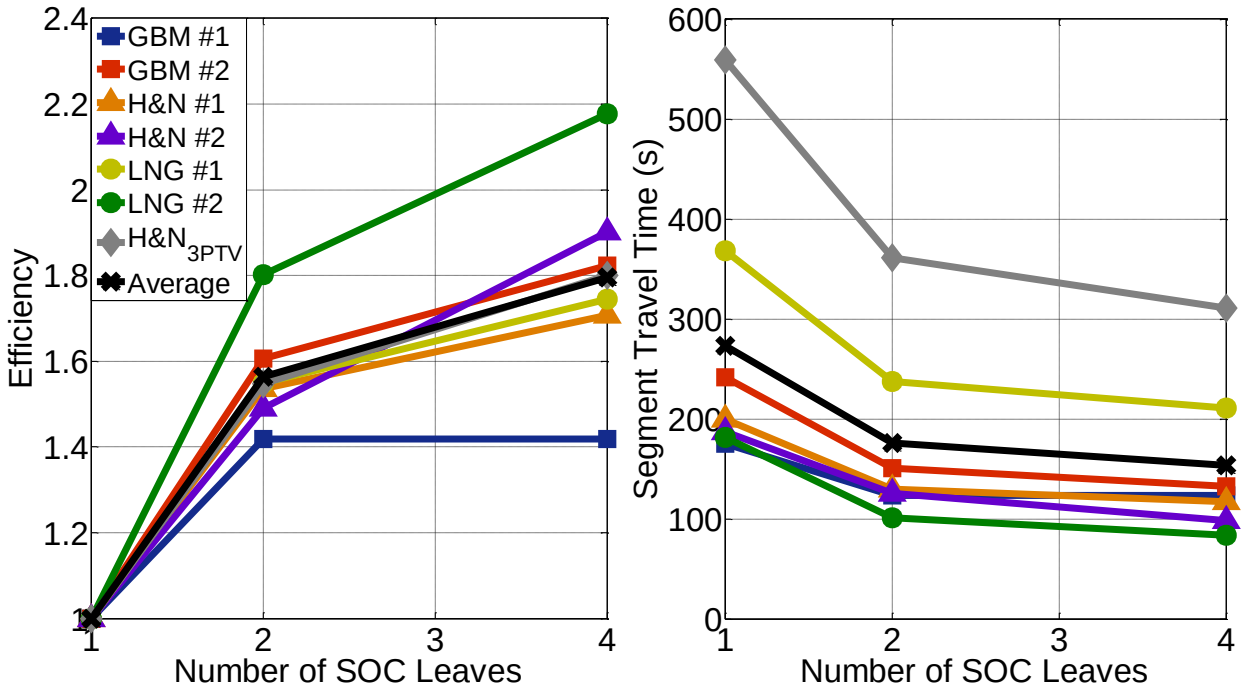


Figure 2.22: Plots of Efficiency and Segment Travel Time against the number of SOC leaves. $\boldsymbol{Efficiency = \frac{1\ leaf\ SOC\ delivery\ time}{x\ leaf\ SOC\ delivery\ time}}$

Segment Travel Time Comparison (s)				
	DMS	RAO N=1	RAO N=2	RAO N=4
GBM #1	173.95	175.30	123.62	123.62
GBM #2	235.11	241.83	150.51	132.59
H&N #1	168.10	199.44	129.89	116.80
H&N #2	175.68	187.13	125.60	98.42
LNG #1	346.60	368.08	237.56	210.92
LNG #2	218.13	181.68	100.79	83.43

$H\&N_{3PTV}$	629.22	558.97	361.30	310.60
Average	278.11	273.20	175.61	153.77

Table 2.17: Comparison of the segment travel time between DMS and RAO with the various SOC designs.

Figure 2.22 shows the total segment travel time for all 20 beams for each patient case. The average efficiency for the *N*=1, *N*=2 and *N*=4 leaf SOC designs is 1 (by definition), 1.56, and 1.80, respectively, and the average time is 273.20 s, 175.61 s, and 153.77 s, respectively. The largest gain in efficiency was observed when the SOC design was changed from *N*=1 to *N*=2. The *N*=4 design does further increases the efficiency, but a trend of diminishing returns is apparent. The average segment travel time between all patients for the DMS case was estimated to be 278.11 s, which is approximately equal to the N=1 SOC design. The individual travel times are shown in Table 2.17. Overall, the total leaf travel times are consistent with 20 field step-and-shoot IMRT delivery.

2.3.4 Discussion

In the study, we presented a novel method to generate rectangular apertures for SOC IMRT delivery. This method is based on regularization of the rectangular basis coefficient in the dose domain to minimize the number of apertures while maintaining the dosimetric quality. Compared to the MLC-based plans, despite the remarkable simplification of fluence maps into deliverable rectangles, the direct aperture optimization nature allows RAO to stay competitive. Our study

was based on a non-coplanar beam orientation optimization platform that we previously showed to be superior to coplanar arc plans. However, the same RAO method is readily applicable to coplanar IMRT without modification.

Although the Chambolle-Pock algorithm is remarkably efficient in solving the dose domain optimization problem, the computational cost of RAO plans remains substantial, particularly for larger PTVs. The length and width of the fluence grid are discretized to be a power of 2 to work properly. For instance, using our beamlet size of 0.5 cm^2, any plan with a PTV larger than 4 cm but smaller than 8 cm in diameter must use a 16 x 16 fluence grid, even if it is just slightly larger than 4 cm. The $H\&N_{3PTV}$ plan had the largest PTV dimension measuring approximately 20 cm. If RAO was performed using the regular beamlet size, this plan would have required a 64 x 64 fluence grid for each beam. Our coefficient space then has a resolution of 127 x 127 pixels, making the optimizer solve 322580 variables simultaneously for all 20 beams. The inclusion of a dose domain transformation matrix, which contains tens of millions of non-zero entries, in the optimization further increases the computational complexity. The amount of data that must be handled simultaneously in the memory would exceed the available 512GB RAM on our workstation for larger PTVs and high dose calculation resolution. Therefore, the RAO method for the $H\&N_{3PTV}$ plan was recalculated at half the dose matrix resolution and half of the beamlet resolution to achieve a reasonable cost optimization although superior dosimetric quality is expected had the higher dose resolution been used.

The scaling function R_m is neither orthogonal nor invertible. However, this does not affect plan optimality because sufficient basis were used to cover the entire aperture. We then added a

second step to capture all those basis that were not included in the initial optimization but are more efficient to deliver. Moreoever, the optimization method only requires a one way transform from the coefficient space, α_m, to the fluence domain, f_m, and does not ever need to decompose the fluence back to the rectangular wavelet domain. Therefore, orthogonality and invertibility is irrelevant in this problem.

Jaws-only IMRT was initially developed as an alternative method to MLC based IMRT, which was costly and unreliable in its early stage. However, with the maturation of MLC technology, the need has considerably declined for general purpose IMRT on state of the art C-arm machines. The SOC IMRT, however, is appealing in several unique aspects. First, the delivery time and dosimetric quality do not need to be compromised because of the new hardware design and optimization algorithm. Second, even for N=4, the SOC system still has far fewer and thicker moving leaves than the conventional MLC, allowing it to be further miniaturized to for compact linac head designs[112] and small animal irradiators. An effective way to reduce the linac head size is by moving the beam intensity modulator closer to the X-ray source. Moving the MLC closer to the X-ray source while maintaining the same intensity modulation resolution is increasingly difficult since a shorter source-to-collimator distance would require reduction of the already thin MLC leaf width, making fabrication more challenging, reducing mechanical reliability and increasing the interleaf leakage from the loss of the tongue and groove. In comparison, the resolution of SOC IMRT is not dependent on the leaf width, but rather on the motor accuracy capable of being in the micron range. For the same reason, the undesirable tradeoff between large field size and high resolution MLCs can be avoided with SOC. Third, as shown in the aperture size comparison, the SOC plan apertures are on average 3.9 times larger than those of

the MLC plans. This would allow a significantly shorter beam-on time, reduced leakage dose and potentially improved IMRT QA results.

The dosimetric improvement was driven by the novel RAO algorithm. To overcome the other major deficiency of jaw-only IMRT, the SOC utilizing increasing number of leaves per bank can significantly improve delivery efficiency. The number of leaves in the proposed SOC is still far fewer than the typical number of leaves in a MLC, thus maintaining the ability to miniaturize and achieve a higher modulation resolution. Based on our estimation, increasing N beyond 4 may still increase the delivery efficiency, but the returns start to diminish. Although the same colored areas can be delivered in parallel, these areas are smaller and more fragmented with increasing N. Multiple leaves need to move in synchrony to deliver X-rays in the rest of the area, reducing the return for more leaves per bank. Considering the increasing mechanical complexity, an N=2-N=4 leaf/bank design may be an optimal balance between complexity and delivery efficiency.

In this study, the potential dosimetric advantages of using higher achievable resolution were not explored, limited by the finite beamlet calculation resolution, the discretized nature of the rectangular representation and available computer resources. To explore this potential, we will investigate an adaptive grid and multi-resolution approach in future research.

2.3.5 Contribution to Deterministic DAO Approach

The IMRT problem was formulated into a direct aperture optimization problem minimizing the dose error while solving for the fluences using rectangular representation. This resulted in rectangular apertures that can be directly delivered with jaws only. The delivery efficiency may further be enhanced using modified sparse orthogonal collimators (SOC) utilizing 2-4 leaves per collimator bank. The potential gains from the greater resolution capabilities of jaws and SOCs have yet to be fully explored because of current computational limits in handling the number of discretized beamlets. The direct aperture nature of the rectangular aperture optimization was eye opening, as the fluence maps produced were completely different than the fluence maps produced without DAO. This study was our first attempt at the DAO problem with rectangular apertures accessible at the time, and revealed the importance of DAO for radiation therapy. However this study constricts itself to creating fluence maps using only rectangular apertures. In order to have a complete description of DAO, a reformulation of the optimization problem to include any shaped segment is required.

2.4 A comprehensive formulation for volumetric modulated arc therapy[113]

2.4.1 Introduction

Volumetric modulated arc therapy (VMAT) is a widely adopted radiation therapy technique. The adoption was supported by dosimetric studies showing that with comparable dose distributions[114], VMAT is significantly more efficient in both treatment time and total monitor units (MU) than static beam intensity modulated radiation therapy (IMRT)[115-117]. The theoretical

framework of VMAT was originally introduced in 1995 by Yu[118] as intensity modulated arc therapy (IMAT), which generated multiple MLC segments per beam angle and requires multiple arcs to deliver[8,119,120]. More practical single arc VMAT algorithms were subsequently developed[121-123] including a representative publication by Otto[124].

Compared to the static beam IMRT problem, the arc optimization problem was considered significantly more complex, due both to the substantially increased beam orientations and the additional machine mechanical constraints such as gantry and MLC mechanical limits. Using a multi-resolution approach[121-124], several methods progressively inserted new beams between sparsely sampled beams using interpolation and then randomly sampled MLC aperture shapes and weights using simulated annealing. These methods were effective to reduce the optimization problem complexity and achieve aperture continuity between adjacent apertures. However, such greedy methods do not guarantee optimality. To mitigate the local minimum problem, in practice, two or more arcs are still commonly required to introduce different initial conditions and achieve the desired dosimetry[125], despite the original promise of using the single arc. Also due to the multi-resolution approach, optimization weights and penalties applied at earlier stages of optimization tend to carry a greater influence. In addition to the stochastic nature of simulated annealing optimization, the optimization results highly depend on the order and timing that the optimization parameters are applied, making reproducing an existing plan difficult, if not impossible. Craft et al[126] attempted to avoid the progressive sampling issue by starting an IMRT optimization at every 2˚, and then created the apertures by merging and simplifying adjacent fluence maps. This leads to another problem that is common in inverse treatment planning,

which is the heuristic conversion from fluence map to MLC segments, which typically introduces noticeable varying degrees of dosimetric quality degradations[40,127,128].

To avoid the stochastic simulated annealing method used in previous direct aperture VMAT implementations and directly optimize based on beam apertures, Peng et al developed a column-generation-based VMAT method algorithm[129]. This method iteratively selects a new aperture for densely sampled arc beams from an aperture set based on its contribution to the objective function. Once the aperture is selected, the optimization proceeds to the next beam and selects the next aperture, imposing potential mechanical limitations based on the previous aperture shape. This method has obvious limitations from being a greedy heuristic algorithm, as it solves a subproblem in each step that does not simultaneously optimize all possible beam angles. Furthermore, the number of possible aperture shapes increases combinatorially. Using a complete aperture set for large tumor or high resolution dose modulation quickly becomes mathematically intractable. Li-Tien Cheng et al introduced a binary level-set shape optimization model for VMAT, illustrating the efficacy of level-set methods for radiotherapy[130]. While efficient, the disadvantage to using a binary level-set lies in that fact the level-set function is discontinuous. Derivatives of the level-set at the boundary do not exist and this can lead to poor accuracy in the variational approach. Because of the pivotal role of VMAT in today's radiotherapy practice, there is a strong need to overcome the existing limitations and develop a new level-set based VMAT framework that formulates the full DAO program using a continuous level-set function.

2.4.2 Methods

2.4.2.1 Optimization Formulation

The proposed comprehensive VMAT (comVMAT) optimization formulation takes the following form:

$$\underset{\{x_\theta\,,\,c_\theta,\,\Phi_\theta\}_{\theta=0}^{n}}{\boldsymbol{argmin}} \quad \underbrace{\frac{1}{2}\left\|W\left(\sum_\theta (A_\theta x_\theta) - d\right)\right\|_2^2}_{dose\ fidelity\ term} + \sum_\theta \underbrace{\left(\lambda_1 \left\|D_\theta^{\parallel} x_\theta\right\|_1 + \lambda_2 \left\|D_\theta^{\perp} x_\theta\right\|_1\right)}_{term\ set\ 1}$$

$$+ \sum_\theta \sum_{i,j} \left(\underbrace{\frac{\gamma}{2}\left[\left(x_{\theta ij} - c_\theta\right)^2 H\left(\Phi_\theta(i,j)\right) + x_{\theta ij}^2 \left(1 - H\left(\Phi_\theta(i,j)\right)\right)\right]}_{term\ set\ 2} \right.$$

$$\left. + \underbrace{\frac{k}{2}\left[\left(H\left(\Phi_\theta(i,j)\right) - H\left(\Phi_{\theta-1}(i,j)\right)\right)^2 + \left(H\left(\Phi_\theta(i,j)\right) - H\left(\Phi_{\theta+1}(i,j)\right)\right)^2\right]}_{term\ set\ 3} \right)$$

$$\boldsymbol{subject\ to} \qquad x \geq 0,$$

Equation 2.14

where x_θ, c_θ, and Φ_θ are the optimization variables. x_θ is the vectorized fluence map, c_θ is a value that x approaches within an aperture, and Φ_θ is the level set function[131], defined as positive where the aperture exists and negative elsewhere. The level set $\{(i,j)|\Phi_\theta(i,j) = 0\}$ describes the aperture boundary. Beam angles are indexed by θ, which ranges from 1 to n, and i and j are indices for a beamlet at a given θ. The fluence to dose transformation matrix is denoted by A, and the desired dose, d_0, is set as the prescription dose at the PTV and zero elsewhere. The

diagonal weighting matrix, W, weights the structures of interest. The derivative matrices, $D^{\parallel}$ and $D^{\perp}$, take the derivative of the fluence in both directions parallel and orthogonal to the MLC leaf movement. H is the Heaviside function

$$H(v) = \begin{cases} 1 & if\ v \geq 0 \\ 0 & if\ v < 0 \end{cases}.$$

Equation 2.15

Essentially, $H\left(\Phi_{\theta}(i,j)\right)$ equals one inside the aperture and zero elsewhere. $x_{\theta ij}$ is a scalar value representing a single beamlet at a given beam angle θ and an i^{th} and j^{th} location on the beam, while x_{θ} is a vector of all the fluences at a specific beam angles. c_{θ} is a scalar quantity and only has one value per beam at a given time.

Intuitively, the dose fidelity term, in Equation 2.14, attempts to push the final dose as close as possible to the desired dose. Term set 1 is the anisotropic total variation (TV) regularization, which has been shown to successfully encourage piecewise continuity on the fluence maps[28,132]. The TV regularization term considers the entire fluence map of the beam, so the term ultimately controls the segment size and shape, abating irregularities and holes in the aperture shape. Soft regulation of the minimal leaf gap and the max leaf interdigitation can be accomplished by independently adjusting the weightings λ_1 and λ_2, respectively. Term set 2 is pushing x towards c where the aperture is defined and zero elsewhere. Term set 3 encourages adjacent beam angles

to be similar to regulate leaf movement between beam angles. For the 1st and nth θ, the $\Phi_{\theta-1}(i,j)$ and $\Phi_{\theta+1}(i,j)$ are equal to their respective $\Phi_{\theta}(i,j)$.

2.4.2.2 Algorithm

We use a block minimization algorithm to solve the minimization problem in Equation 2.14 by alternatingly updating the fluence x_θ, aperture intensityc_θ, and aperture shapeΦ_θ, while holding the other two constant. The algorithm is broken down into 3 modules described below. Each iteration of the algorithm runs module 1 through 3, and the process is repeated until a satisfactory convergence rate is achieved. Convergence of the alternating approach is guaranteed, as long as each module is able to find a minimum for its respective variable, while holding the other variables constant. A proof of the convergence is provided by Gorski et al.[133]

2.4.2.2.1 Algorithm Module 1: Update $\boldsymbol{x_\theta}$

Module 1 minimizes Equation 2.14 w.r.t. x_θ while holding c_θ and Φ_θ constant. This subproblem can be rewritten as

$$\begin{aligned} \underset{\{x_\theta\}_{\theta=0}^{n}}{argmin} \quad & \frac{1}{2}\left\| W\left(\sum_\theta (A_\theta x_\theta) - d\right)\right\|_2^2 + \sum_\theta \left(\lambda_1 \left\| D_1^{\parallel} x_\theta \right\|_1 + \lambda_2 \| D_2^{\perp} x_\theta \|_1\right) \\ & + \frac{\gamma}{2}\sum_\theta \left(\left\| H_{\Phi_\theta}\left(x_\theta - c_\theta \vec{1}\right)\right\|_2^2 + \left\| \left(I - H_{\Phi_\theta}\right) x_\theta \right\|_2^2\right) \\ subject\ to \quad & x \geq 0, \end{aligned}$$

Equation 2.16

where I is the identity matrix, and H_{Φ_θ} in this notation is a diagonal matrix that has the information of $H\big(\Phi_\theta(i,j)\big)$ along its diagonal for all i and j. Essentially, the diagonal of H_{Φ_θ} has a value of 1 if the corresponding $\Phi_\theta(i,j)$ is positive and zero otherwise. The formulation is solved using the Chambolle-Pock algorithm[20], a proximal-class primal-dual algorithm, which is described in the Appendix.

2.4.2.2.2 Algorithm Module 2: Update $\boldsymbol{c_\theta}$

Step 2 minimizes Equation 2.14 w.r.t c_θ given Φ_θ and x_θ constant, which is provided by the closed-form solution

$$c_\theta = \frac{\sum_{i,j} x_{\theta ij} H\big(\Phi_\theta(i,j)\big)}{\sum_{i,j} H\big(\Phi_\theta(i,j)\big)} \qquad for\ \theta = 1,\ldots,n.$$

Equation 2.17

This calculation takes an average of the beamlet intensities that are defined as part of the aperture for each beam angle.

2.4.2.2.3 Algorithm Module 3: Update $\boldsymbol{\Phi_\theta}$

Step 3 minimizes Equation 2.14 w.r.t. Φ_θ while holding x_θ and c_θ constant:

$$\underset{\{\Phi_\theta\}_{\theta=0}^{n}}{argmin} \quad \sum_{\theta}\sum_{i,j}\left(\frac{\gamma}{2}\left[\left(x_{\theta ij}-c_\theta\right)^2 H\left(\Phi_\theta(i,j)\right)+x_{\theta ij}^2\left(1-H\left(\Phi_\theta(i,j)\right)\right)\right]\right.$$

$$\left.+\frac{k}{2}\left[\left(H\left(\Phi_\theta(i,j)\right)-H(\Phi_{\theta-1}(i,j)\,)\right)^2+\left(H\left(\Phi_\theta(i,j)\right)-H(\Phi_{\theta+1}(i,j)\,)\right)^2\right]\right).$$

Equation 2.18

The level set function, Φ_θ, is iteratively updated by the expression

$$\Phi_\theta^{k+1}(i,j)=\Phi_\theta^k(i,j)+\frac{d\Phi_\theta^k(i,j)}{dt}dt,$$

Equation 2.19

where $\frac{d\Phi_\theta(i,j)}{dt}$ was derived as

$$\frac{d\Phi_\theta(i,j)}{dt}=\frac{\gamma}{2}\left(2c_\theta x_{\theta ij}-c_\theta^2\right)\delta\left(\Phi_\theta(i,j)\right)dt$$

$$+k\left(H\left(\Phi_{\theta-1}(i.j)\right)+H\left(\Phi_{\theta+1}(i,j)\right)-2H\left(\Phi_\theta(i,j)\right)\right)\delta\left(\Phi_\theta(i,j)\right)dt.$$

Equation 2.20

Practically, we use the sigmoid function and its derivative to approximate the Heaviside and the Dirac delta function:

$$H(\Phi) \cong Sigmoid(\mathrm{q}\Phi) = \frac{1}{1+e^{-q\Phi}}$$

$$\delta(\Phi) = \frac{dH(\Phi)}{d\Phi} \cong \frac{dSigmoid(q\Phi)}{d\Phi} = \frac{qe^{q\Phi}}{(1+e^{q\Phi})^{2}},$$

Equation 2.21

where q is some constant. A larger value of q allows for the sigmoid function to more closely resemble the Heaviside function.

2.4.2.3 Ensuring optimal plan quality

Once the algorithm has converged, and the apertures shapes no longer change, a final polishing step is taken to ensure superior plan quality. The formulation locks in the solved aperture shapes and solves for the fluence of each beam angle without the aperture regularization constraints, which is expressed as

$$\begin{aligned} &\underset{b}{argmin} && \frac{1}{2}\|W(AFb-d)\|_2^2 \\ &subject\ to && b \geq 0. \end{aligned}$$

Equation 2.22

The optimization variable, b, contains one intensity value for each beam angle. F is a binary matrix containing all of the aperture information from $H(\Phi_\theta(i,j))$ for all of the beam angles. These two variables are related to the fluence via the equation $x = Fb$. This optimization can be easily solved with the Chambolle-Pock algorithm. By solving the optimization in Equation 2.22 as the last step, only the dose difference is penalized, ensuring that the regularization and aperture constraints are not hindering the final dosimetric outcome.

2.4.2.4 Evaluation

To assess the efficacy of the optimization, 4 patients were chosen for this study: A glioblastoma multiforme patient (GBM), a lung patient (LNG), and 2 head and neck patients, one with 3 PTVs ($H\&N_{3PTV}$) and one with 4 PTVs ($H\&N_{4PTV}$). Table 2.18 shows the 4 patients with their respective prescription doses and PTV volumes.

	Prescription Dose (Gy)	**PTV Volume (cc)**
GBM	25	6.23
LNG	50	47.78
$H\&N_{3PTV}$	54	197.54
	59.4	432.56
	69.96	254.98
$H\&N_{4PTV}$	54	58.98

	60	149.32
	66	242.23
	70	175.20

Table 2.18: Prescription doses and PTV volumes.

Using a convolution/superposition code with a 6 MV x-ray polyenergetic kernel, the beamlet dose was calculated for 180 equally spaced coplanar beam angles around the patient. The dose calculation method was described in our previous publications[71,134]. The chosen beamlet size was 0.5×0.5 cm^2, and the dose matrix resolution was 0.25×0.25×0.25 cm^3. The resulting dose was stored in the dose matrix A for optimization. A 5 cm ring structure was added to the optimization to minimize dose spillage. Each patient was then optimized using the comprehensive VMAT (comVMAT) algorithm, and W was adjusted until a desirable dose was achieved.

The comVMAT plans were compared to the patients' respective clinical VMAT plan (clnVMAT). The clnVMAT plans were planned on the Eclipse treatment planning system using 2 superimposing 360^o coplanar arcs with 90^o collimator rotation. The PTV D95, D98, D99, Dmax, and PTV homogeneity, defined as $\frac{D95}{D5}$, were evaluated. The organs-at-risk (OAR) Dmax and Dmean were also assessed. Max dose is defined as the dose at 2% of the structure volume, D2, which is recommended by the ICRU-83 report[58].

2.4.3 Result

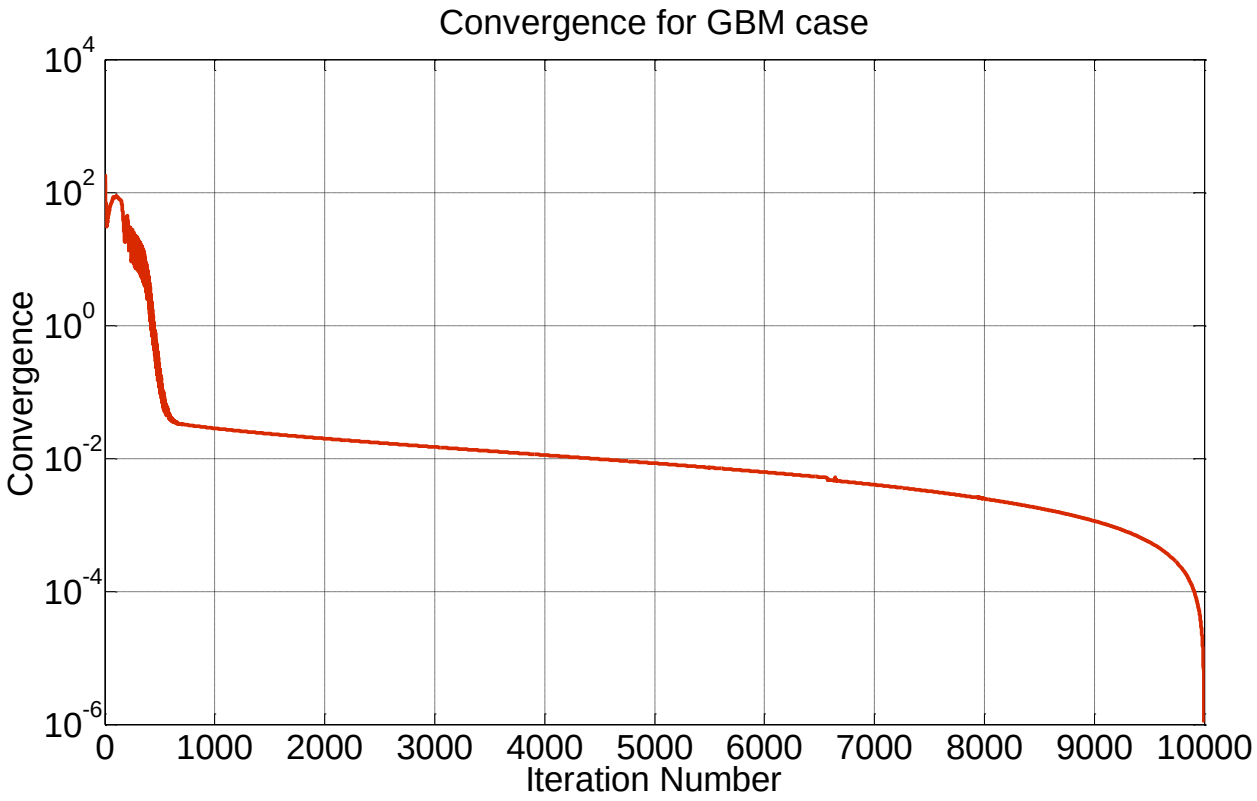


Figure 2.23: Convergence plot for the GBM case. Convergence is defined as $convergence(k) = \frac{(objval(k) - minobj)}{minobj}$, where $objval(k)$ is the objective value at the k^{th} iteration and $minobj$ is the minimum objective value, taken at the 10000th iteration. Objective value is based off of the objective function defined in equation (1), and is recorded after each block iteration of modules 1, 2, and 3.

The aperture shapes converged, with a relative convergence of 10^{-1}, within 600 iterations of the optimization for each case. This degree of convergence has been shown to produce plans that are dosimetrically equivalent to other plans that have tighter convergences[28]. A convergence plot for the GBM case is shown in Figure 2.23, showing the convergence relative to the optimal value taken at the 10000th iteration. The oscillatory convergence pattern in the beginning before 600

iterations comes from the alternating optimizations to solve for f, c, and Φ. The variables are each taking small steps towards optimality, but may make the other variables temporarily and slightly less optimal in a given iteration. However this pattern diminishes after 600 iterations. Depending on various factors—such as case complexity, tumor volume, and body volume—total computational time varied from 5 minutes for the GBM case to 40 minutes for the H&N cases, per optimization run. Unlike the clinical implementation, no human involvement is required during the optimization process after the weights have been set. While the overall planning time is within acceptable range, the algorithm was written and tested in MATLAB for proof of principle. Its performance should improve considerably using faster programming language.

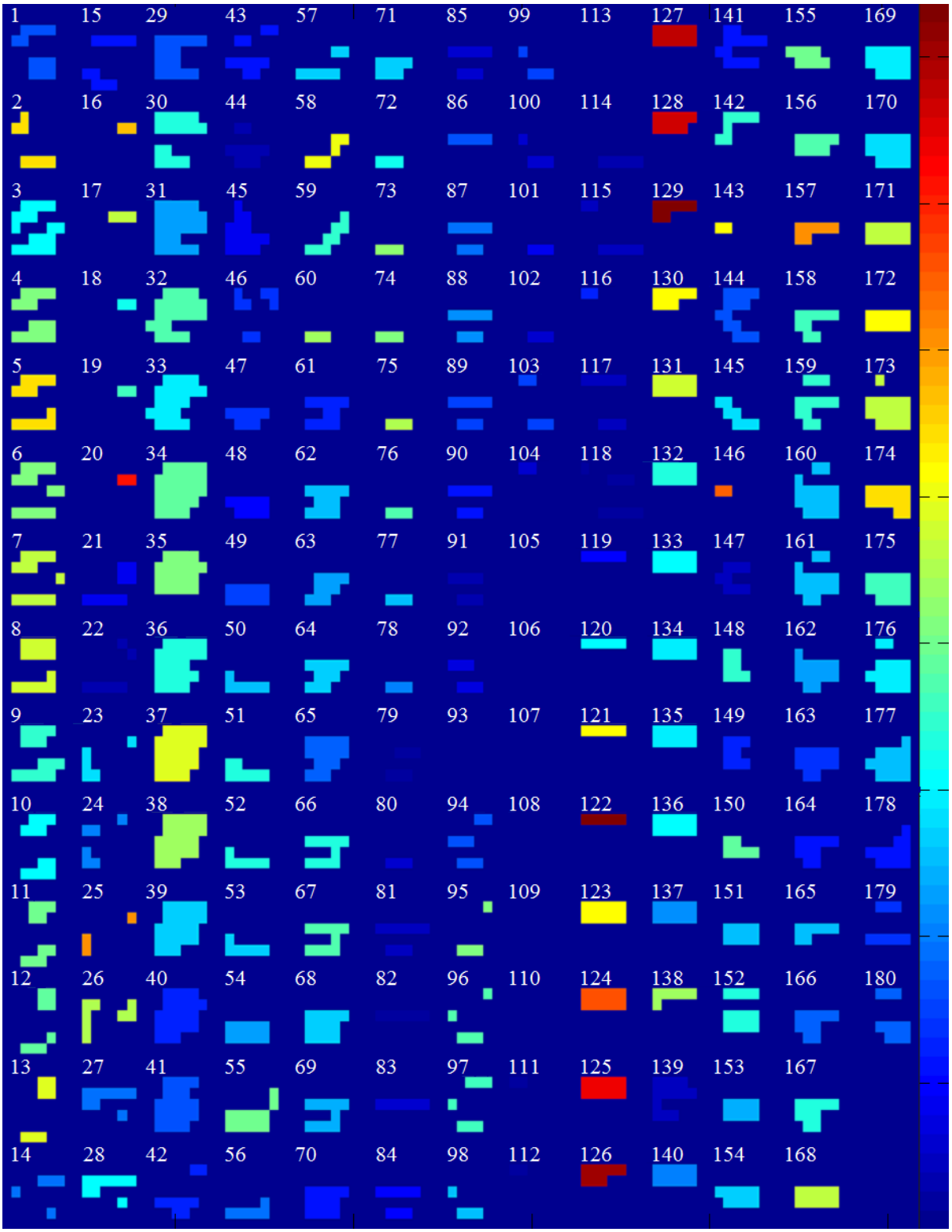
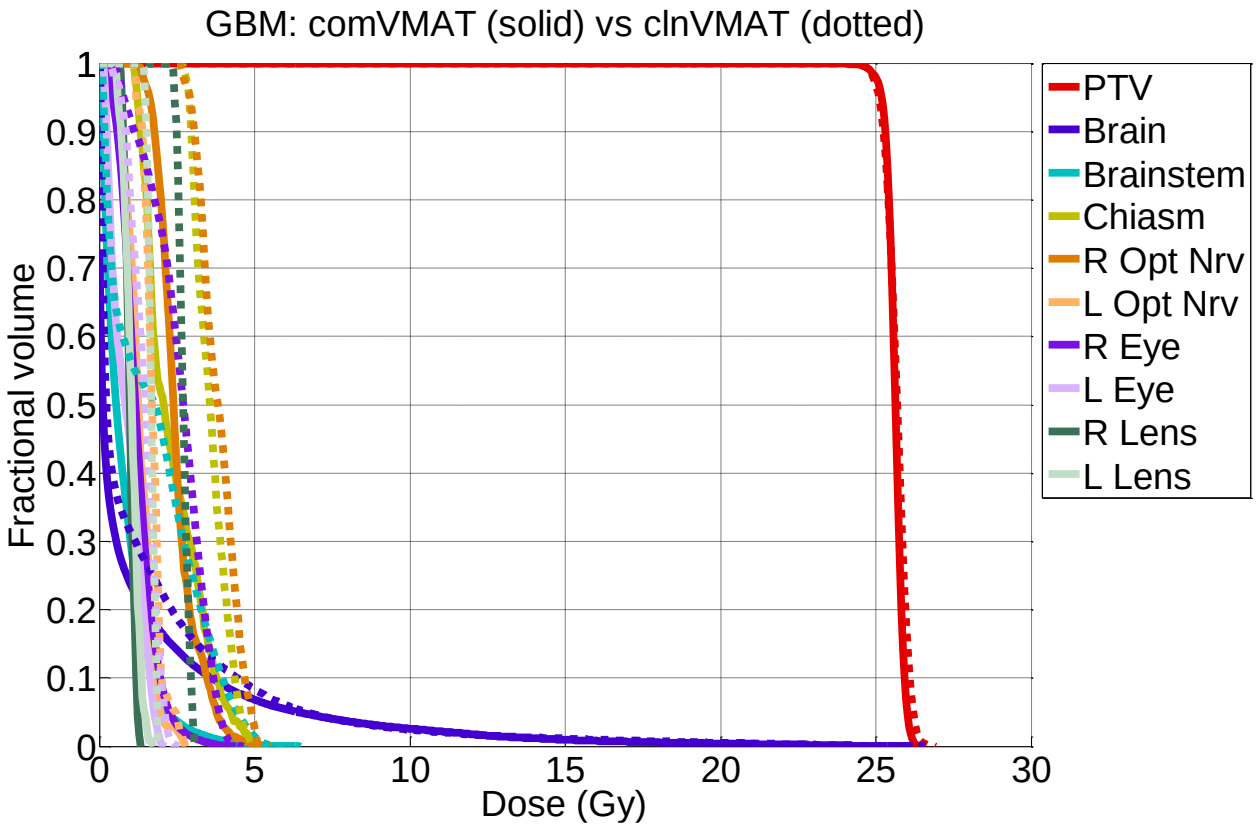
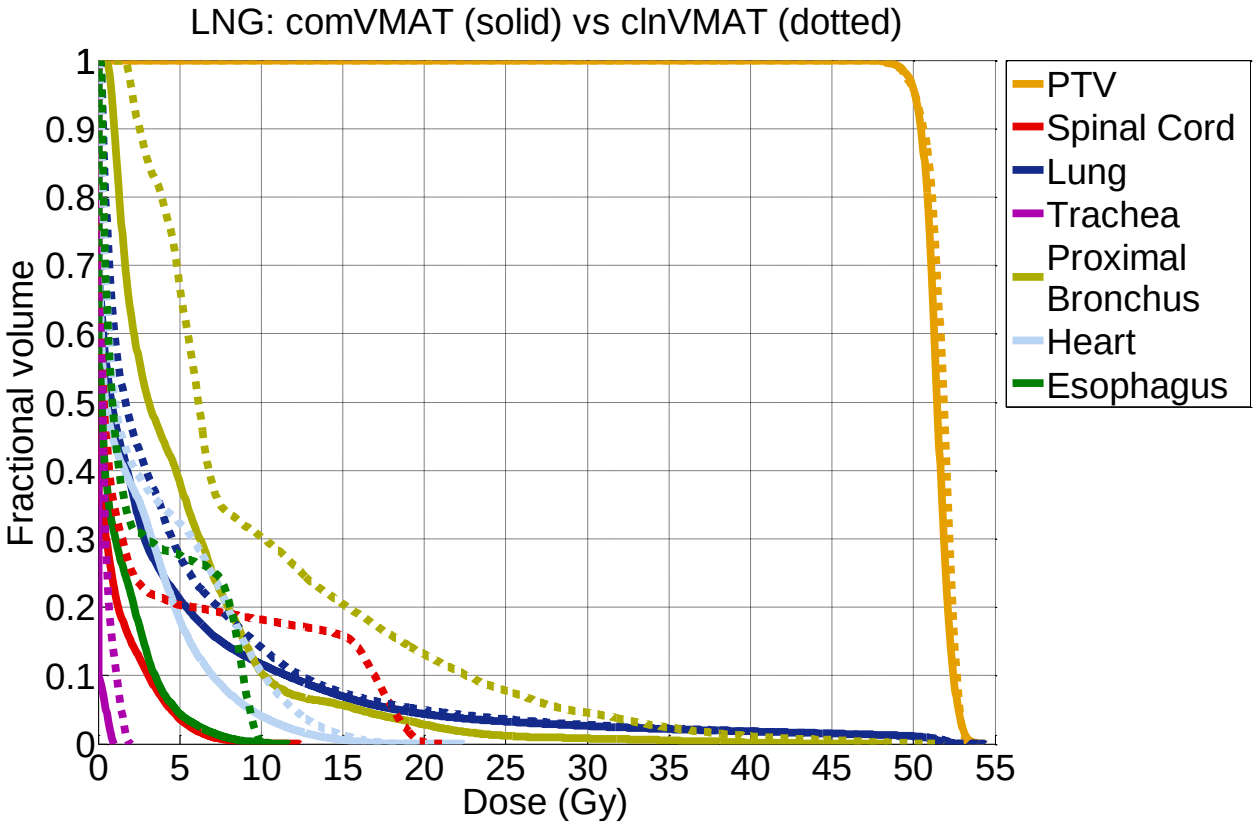
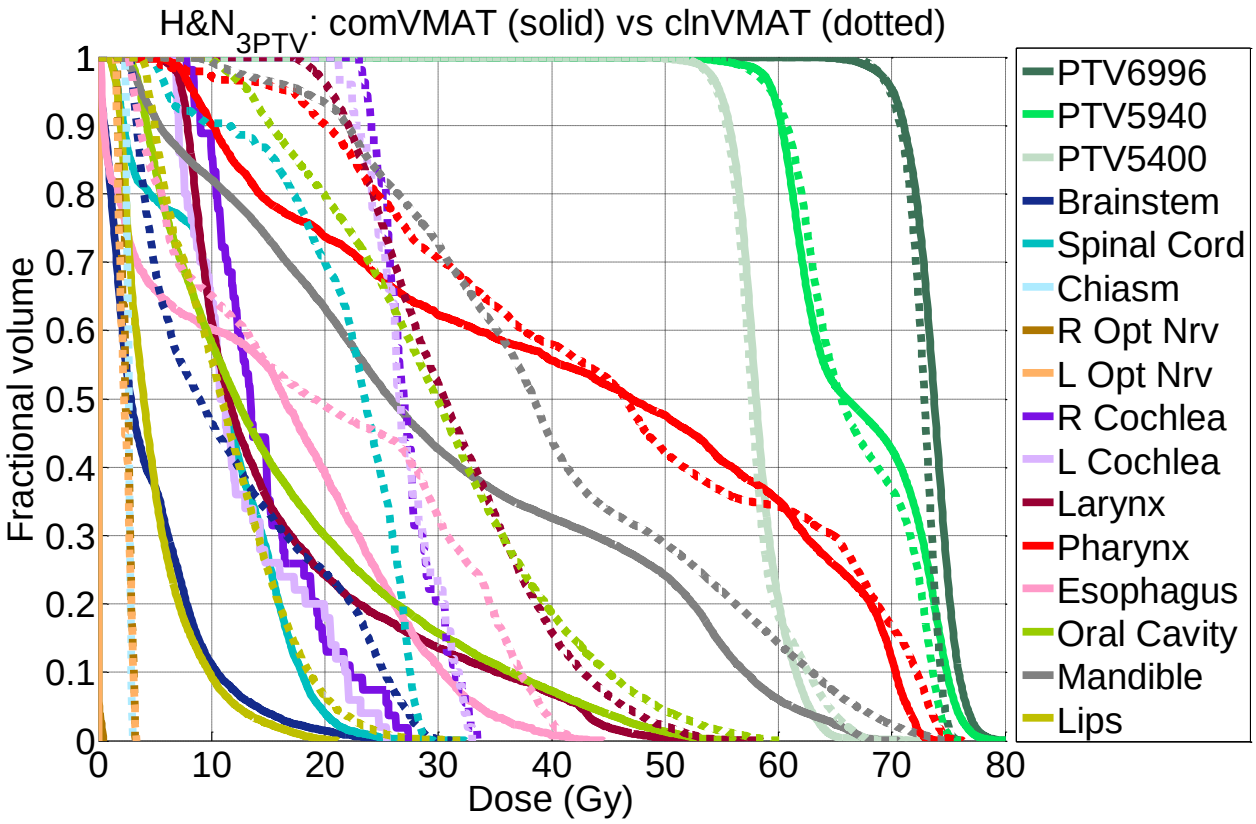


Figure 2.24: Schematic of apertures for all 180 beams, spaced 2° apart, for GBM case. MLC leaf direction is horizontal for this diagram. Color scales show fluence intensities.

The comVMAT method managed to optimize all 180 beams in the coplanar arc simultaneously for all tested cases. Figure 2.24 shows the 180 apertures from the GBM case. The MLC leaf motion direction is horizontal in the schematic. It is observed that beam aperture shapes are similar to their neighbors. A small fraction of beams, such as beam 3 and beam 26, require two segments to deliver. Since every beam is spaced apart by 2 degrees, in practice, these beams can be split into two beams spaced 1 degree apart, each delivering one of the apertures. The gantry speed may be modulated in order to deliver the apertures.




LNG: comVMAT (solid) vs clnVMAT (dotted)
Fractional volume
Dose (Gy)
PTV
Spinal Cord
Lung
Trachea
Proximal Bronchus
Heart
Esophagus


H&N3PTV: comVMAT (solid) vs clnVMAT (dotted)
Fractional volume
Dose (Gy)
PTV6996
PTV5940
PTV5400
Brainstem
Spinal Cord
Chiasm
R Opt Nrv
L Opt Nrv
R Cochlea
L Cochlea
Larynx
Pharynx
Esophagus
Oral Cavity
Mandible
Lips

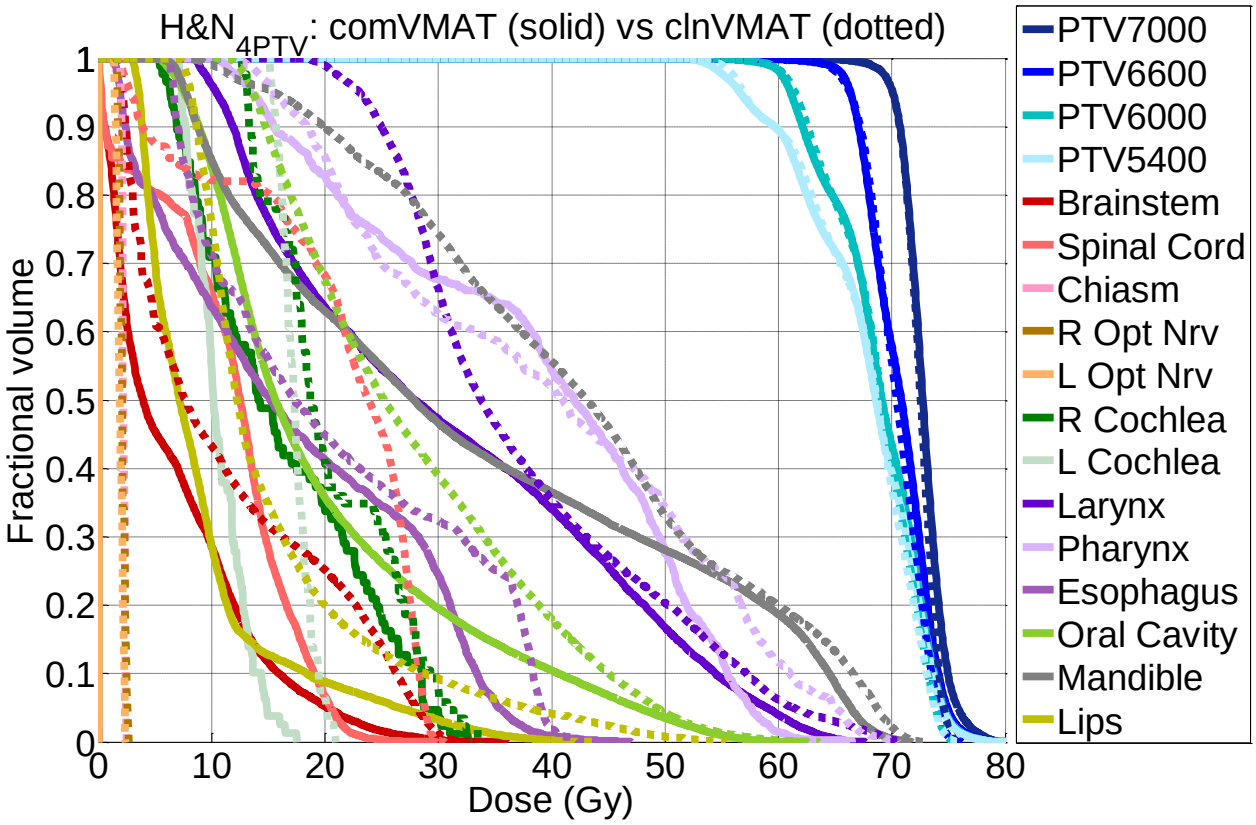


Figure 2.25: DVHs of the GBM, LNG, H&N$_{3PTV}$, and H&N$_{4PTV}$ patients comparing the comVMAT and clnVMAT plans.

Figure 2.25 shows the 4 DVHs from each of the patients, comparing the comVMAT plans against the clnVMAT plans. Qualitatively, it can be observed that comVMAT is able to better spare the OARs while maintaining a competitive PTV dosimetry. The two H&N cases with multiple PTVs, while matching the dose coverage to 95% of the PTVs, had slightly hotter tails to the PTVs by a few Gy with the comVMAT plan. PTVs are shown in green shades for the H&N$_{3PTV}$ case and blue shades for the H&N$_{4PTV}$ case in Figure 2.25. However this marginal increase is outweighed by the substantial sparing in all of the OARs for each plan. The GBM and LNG cases, which had only 1 PTV, were superior in all aspects with the comVMAT plans.

Patient Case		**PTV Statistics**					
		Homogeneity		D95	D98	D99	Dmax
		comVMAT	clnVMAT	comVMAT – clnVMAT (Gy)			
GBM		0.968	0.958	+0.08	+0.09	+0.02	–0.21
LNG		0.949	0.948	+0.00	+0.13	+0.09	+0.04
$H\&N_{3PTV}$	54	0.874	0.847	+0.23	+0.19	+0.06	–2.30
	59.4	0.786	0.801	+0.02	+0.73	+0.15	+1.96
	69.96	0.915	0.935	+0.29	–0.12	–0.63	+2.36
$H\&N_{4PTV}$	54	0.760	0.771	–0.30	–0.97	–1.06	+1.11
	60	0.818	0.832	–0.40	–0.18	–0.11	+1.17
	66	0.885	0.895	+0.05	+0.36	+0.26	+1.40
	70	0.924	0.940	+0.00	–0.05	–0.10	+2.08

Table 2.19: PTV homogeneity, dose coverage (D95, D98, and D99), and Dmax.

Dose Difference comVMAT – clnVMAT (Gy)	**Dmax**			**Dmean**		
	Largest Value	Smallest Value	Average Value	Largest Value	Smallest Value	Average Value

GBM	–0.02 Chiasm	–1.76 R Lens	–0.80	–0.43 L Opt Nrv	–1.73 R Lens	–1.04
LNG	–0.38 Lung	–14.53 ProxBronch	–5.63	–0.37 Trachea	–4.61 ProxBronch	–1.98
$H\&N_{3PTV}$	–1.95 Pharynx	–9.74 Brainstem	–4.91	–2.43 R Opt Nrv	–15.31 Larynx	–8.26
$H\&N_{4PTV}$	–0.56 OralCavity	–13.38 Lips	–4.33	–1.16 Pharynx	–9.47 Mandible	–5.13

Table 2.20: Largest, smallest, and average values found for (comVMAT – clnVMAT) dose differences for Dmax and Dmean.

On average, the PTV D95, D98, and D99 changed by –0.01%, +0.02% and –0.23 % of the prescription dose, indicating virtually identical dose coverage between comVMAT and clnVMAT. However PTV Dmax increased, on average, by 1.40 % of the prescription dose. This change is associated to the two H&N cases with multiple PTVs. The GBM case actually had reduced max dose to the PTV, while the LNG case minimally increased the PTV max dose by 0.08% of the prescription dose. The average calculations include all of the PTVs from the H&N cases. The comVMAT optimizer was able to decrease all of the OARs from all of the cases, shown in Table 2.20, where the largest valued dose differences were still negative. On average, comVMAT plan spared the OARs max and mean dose by 6.59% and 7.45% of the prescription dose, respectively. Comparing all the cases, the LNG case had the single largest sparing in max

dose in an OAR, and spared the proximal bronchus by 14.5 Gy of max dose. Likewise, the single largest reduction in the mean dose to an OAR was from the $H\&N_{3PTV}$ case, sparing the larynx by 15.3 Gy of mean dose.

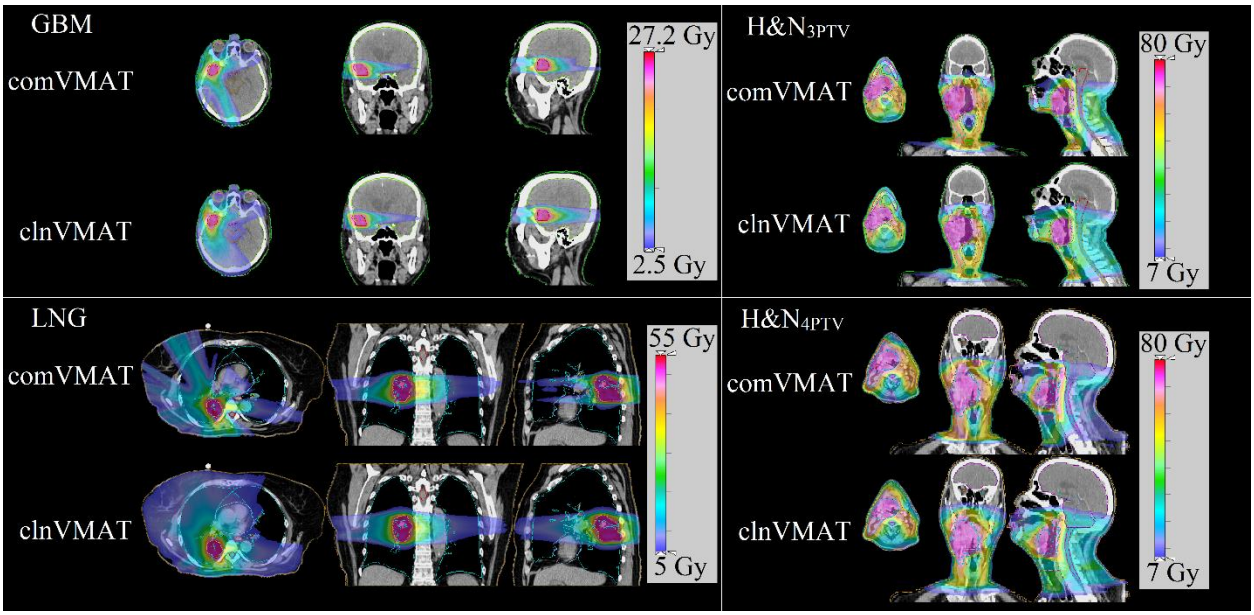


Figure 2.26: Dose washes for all patients. The low dose cutoff for viewing is set at 10% of the prescription dose.

Figure 2.26 shows the dose wash for all of the patients. From a qualitative perspective, it can be observed that comVMAT distributes the dose very differently than clnVMAT. The GBM and LNG cases more clearly illustrate that comVMAT plans gave a much heavier weighting to some selective beams, giving rise to a dose distribution that bears some resemblance to a coplanar IMRT plan, even though there is still only 1 fluence value per beam angle. The clnVMAT overall spreads the distribution of fluence intensities more evenly among the beams, giving rise to a less angular modulated dose distribution pattern and greater dose to OARs. For example, for

the GBM plan in Figure 2.26, the comVMAT plan was able to entirely avoid the brainstem, while the clnVMAT plan covers the brainstem with at least 2.5 Gy of dose.

2.4.4 Discussion

When solved using an algorithm based on a proximal-class primal-dual algorithm, the Chambolle-Pock algorithm[20], we found a new methods, based on L2-norm fidelity terms and L1-norm regularization terms[28,98], to incorporate fluence maps simplification into the dose domain optimization. In this study, we further developed the methods to solve the VMAT problem, which previously only had greedy heuristic solutions. The new method optimizes all VMAT beams at the same time without progressive sampling. This translates into a number of theoretical and practical advantages.

At the theoretical level, the optimization cost function provides a simple yet complete description of the physical problem. An L2-norm fidelity term is used to minimize the dose distribution of the prescription dose, and an anisotropic total variation regularization term to piecewise smooth the fluence map. The following level set function term shapes the fluence to fit the aperture and then encourages the continuity in the aperture shapes between adjacent beams. As we demonstrated, solving the optimization problem results in a final solution that requires no additional “patches” that are commonly observed in heuristic solutions. Because of the optimization problem size and the fact that the cost function is not differentiable, methods such as interior points and gradient descent are not suitable to solve the optimization problem. We instead used the Chambolle-Pock algorithm[20] to efficiently manage the optimization problem.

The algorithm is remarkably fast at solving this type of optimization problem because it does not require solving system of linear equations involving the fluence to dose transformation matrix, A, at every iteration, unlike other first order methods such as alternating direction method of multipliers[55]. Instead, Chambolle-Pock simply requires just the multiplication of the matrix and its transpose at each iteration.

The comVMAT method is superior to existing VMAT methods in the following aspects. comVMAT optimizes all beam apertures and beam intensities together, providing greater flexibility to approach the ideal dose. A difference in isodose distribution can be appreciated in Figure 2.26, where clnVMAT resulted in a more uniform dose spillage and comVMAT gave heavy weights to a narrow range of beams, resulting in dose distribution resembling beam orientation optimized static beam IMRT or hybrid IMRT VMAT plans[30]. In terms of the optimization solver, the analytical solution used in comVMAT is more robust that the stochastic or greedy heuristic algorithms used in existing VMAT methods. Moreover, comVMAT solves the direct aperture problem. Previously, the aperture was generated either in an additional step that degraded the optimization results, or was limited to use a small subset of available apertures[15,16] due to non-polynomial computational cost to include all possible apertures. In contrast, comVMAT can arrive at any aperture shape without being limited to a pre-set library or the neighborhood of conformal apertures.

A remaining issue in comVMAT is that the resultant aperture per beam is not explicitly guaranteed to be deliverable in one segment. While the total variation regularization term has limited the number of apertures for most beams to 1, there are a small fraction of beams that

require more apertures. For example, in figure 2, beams 3, 26 and 46 have segments that must be delivered in 2 apertures. While increasing the regularization weighting may eliminate multiple apertures per beam, there may be a compromise in the dosimetry if the TV penalty is too high. There are multiple solutions in handling the delivery of these particular beams. Since each aperture delivery is spread across 2°, the first solution is to split the beam into two 1°-apart beams and deliver the two segments sequentially. Alternatively, the two segments may be approximated by the closest single segment. The alternative would result in change in dose distribution but the change is expected to be minimal because only a small fraction of beams need more than 1 MLC segment. On the other hand, the issue may become an opportunity for generating hybrid static beam IMRT and VMAT plans. By further relaxing the TV regularization, we expect dosimetric improvement and more beams that require 2 or more MLC segments. The gantry speed and linac output would be modulated to deliver these beams, adding more static beam flavor to comVMAT for superior dosimetry.

Delivery time of the plans are not explicitly controlled by the objective function, but instead are indirectly maintained by the total variation term, term set 1, to limit the number of deliverable segments, and the aperture similarity term, term set 3, to limit the MLC leaf motion between each angle. Relaxing the weights on these terms to allow more segments and greater changes between apertures will increase treatment time and offset the benefit of using a single arc. The tradeoff between plan quality and delivery time is a topic of further investigation.

At the practical level, the single arc comVMAT is shown superior to the current commercial implementation of clnVMAT using two superimposing arcs. comVMAT may be potentially

advantageous, pending further research, in knowledge based planning because it is more robust to the optimization history, while clnVMAT depends on the entire history of optimization parameter set up, which is impossible to track and incorporate in knowledge based planning. The comVMAT plans are reproducible, providing a single set of optimization parameters for future learning.

The focus of the study is to present a new VMAT formulism that does not depend on greedy heuristics and results in dosimetry that is at least comparable to the multiple arc VMAT plans using a single arc. While we have carefully fine-tuned our dose calculation code to match the PDD and penumbra of the Eclipse system beams, the TPC commissioning process is not entirely transparent for us to exactly reproduce. The TPS vendor will have to implement our algorithm for an exact comparison.

As a VMAT approach, the particular DAO formulation that was presented was specifically tailored to produce one segment per beam in most cases. Generalization of the method for applications including static beam IMRT requires multiple level of segmentation that requires different mathematical tools. We will explore the potential of an approach to a generalized library-less DAO formulation for both static beam and VMAT IMRT problems.

2.4.5 Contribution to Deterministic DAO Approach

A new approach for the comprehensive VMAT optimization was demonstrated. The new approach formulates the VMAT problem as a single optimization function, including a level set

function to regularize the MLC aperture shapes without relying on a preset aperture library. The optimization function was solved using a proximal-class primal-dual algorithm, which is more robust that stochastic method used in the existing VMAT solutions. The results showed that the new comVMAT using a single arc was consistently superior in OAR sparing to the clinical VMAT using two arcs, while keeping a similar PTV dosimetry. The contribution of this study to the deterministic DAO was in developing a method to create any shaped aperture for fluence map optimization. Currently, the comVMAT expression was a specialized form of a DAO that aims at reducing the number of segments per beam to exactly one. For the generalized DAO, a method for having more than one aperture is needed, which is developed and addressed in the Deterministic DAO study.

3 DETERMINISTIC DIRECT APERTURE OPTIMIZATION USING MULTIPHASE PIECEWISE CONSTANT SEGMENTATION

3.1 Introduction

Intensity modulated radiation therapy (IMRT) was theorized in the 1980s[2] and subsequent developed in several landmark papers[3-5]. Since then, IMRT has been widely accepted into radiation therapy clinics as the staple approach to radiotherapy. While IMRT has been proved exceptional in controlling of dose distributions, one major weakness lies in the fluence map optimization (FMO), which does not consider machine constraints in the inverse optimization and produces optimal fluence maps that need to be converted to deliverable multileaf collimators (MLC) sequences. While several clever sequencing methods have been developed, such as sliding window or the reducing level method[6], this additional post processing step separate from initial FMO inverse optimization leads to varying degrees of degradation in the dose quality.

Direct Aperture Optimization (DAO) was invented to solve the problem by incorporating machine constraints in the optimization to eliminate the post-processing step that follows FMO. One significant challenge with DAO is that the number of possible apertures as the combination of beamlets is mathematically intractable. To manage the challenge, stochastic and greedy approaches have been implemented. The stochastic DAO method utilizes a simulated annealing process[7] to iteratively update the aperture shapes and intensity values[8-14]. In the greedy approach, DAO plans are created by optimizing a predetermined library of apertures, and a column generation method to expand the aperture library until an acceptable treatment plan is acquired[15-17]. These DAO methods result in plans that are directly deliverable by the machine without the need for a separate MLC segmentation step but there are significant limitations. Due to the computational cost, both methods can only search a small fraction of the possible apertures, which may be insufficient for complex IMRT plans.

In this study, we aim to develop a new deterministic DAO method that is not limited to a narrow range of apertures in the optimization. Instead, the MLC segmentation problem is formulated as a piecewise constant segmentation problem that can be efficiently solved using methods such as Mumford-Shah[18-21]. Piecewise-constant segmentation solves an optimization problem that partitions a scaler image into piecewise constant regions that the original image. Following the original Mumford-Shah formulation, recent developments have led to multiphase piecewise-constant segmentations[19,20,22-25]. Pertinent to the IMRT optimization problem, because the piecewise-constant segmented regions are arbitrarily shaped, it provides a mathematically complete description of any possible MLC aperture for a specific fluence map.

In this study, we integrate the multiphase piecewise-constant Mumford-Shah function with the fluence map optimization problem into a multiconvex formulation[26,27]—a non-convex problem that can be solved by alternating among convex modules. We then test the feasibility of this optimization on several patient cases and compare its performance to the DAO method using simulated annealing.

3.2 Methods

3.2.1 Optimization Formulation and Algorithm

The master optimization formulation is written as

$$
\begin{aligned}
\underset{x,u,c}{argmin} \quad & \underbrace{\frac{1}{2}\left\| W\left(\sum_{b=1}^{n_b}(A_b x_b) - d\right)\right\|_2^2}_{Term\ 1:\ Dose\ Fidelity} + \underbrace{\sum_{b=1}^{n_b}\left(\lambda_1 \left\|D_b^{\parallel} x_b\right\|_1 + \lambda_2 \|D_b^{\perp} x_b\|_1\right)}_{Term\ Set\ 2:\ Anisotropic\ TV\ on\ Fluence} \\
& + \gamma \sum_{b=1}^{n_b} \Bigg(\underbrace{\frac{1}{2}\sum_{p=1}^{n_p}\sum_{s=0}^{n_s} u_{bps}\left|x_{bp} - c_{bs}\right|^2}_{Term\ Set\ 3} + \underbrace{\sum_{s=0}^{n_s}\left(\omega_1 \left\|D_b^{\parallel} u_{bs}\right\|_1 + \omega_2 \|D_b^{\perp} u_{bs}\|_1\right)}_{Term\ Set\ 4} \Bigg)_{Multiphase\ Piecewise\ Constant\ \ Mumford-Shah\ Formulation} \\
subject\ to \quad & x \geq 0 \\
& u \geq 0 \\
& \vec{1}^T u_{bp} = 1,
\end{aligned}
$$

Equation 3.1

where the notations for the data and variables used in the DAO formulation are described in Table 3.1.

Notation	Type	Description
Indices		
b	Index	Index for beam. $b \in \{1,2,\dots,n_b\}$
p	Index	Index for beamlet. $p \in \{1,2,\dots,n_p\}$
s	Index	Index for segment. $s \in \{0,1,2,\dots,n_s\}$
Optimization variables		
x	Vector	All beamlet intensities from all selected beams.
x_b	Vector	All beamlet intensities for the b^{th} beam. $x = \begin{bmatrix} x_{b=1} \\ \vdots \\ x_{b=n_b} \end{bmatrix}$
x_{bp}	Scalar	Intensity value for the p^{th} beamlet of the b^{th} beam. $x_b = \begin{bmatrix} x_{b,p=1} \\ \vdots \\ x_{b,p=n_p} \end{bmatrix}$
c	Matrix	Intensity values for all segments and beams.
c_b	Vector	Intensity values for all the segments of the b^{th} beam.

		$c = \begin{bmatrix} c_{b=1}^T \\ \vdots \\ c_{b=n_b}^T \end{bmatrix}$
c_{bs}	Scalar	Intensity value for the s^{th} segment of the b^{th} beam. $c = \begin{bmatrix} c_{b=1,s=1} & \cdots & c_{b=1,s=n_s} \\ \vdots & \ddots & \vdots \\ c_{b=n_b,s=1} & \cdots & c_{b=n_b,s=n_s} \end{bmatrix}$ $c_b = \begin{bmatrix} c_{b,s=1} \\ \vdots \\ c_{b,s=n_s} \end{bmatrix}$
u	$Matrix$	Segmentation labeling array for all selected beams.
u_b	$Matrix$	Segmentation labeling array for the b^{th} beam. $u = \begin{bmatrix} u_{b=1} \\ \vdots \\ u_{b=n_b} \end{bmatrix}$
u_{bs}	Vector	Probability that the beamlets from the b^{th} beam belong to the s^{th} segment. $u_b = [u_{b,s=1} \quad \cdots \quad u_{b,s=n_s}]$

u_{bp}	Vector	Segment probabilities for the p^{th} beamlet of the b^{th} beam. Lives in probability simplex. $$u_b = \begin{bmatrix} u_{b,p=1}^T \\ \vdots \\ u_{b,p=n_p}^T \end{bmatrix}$$
u_{bps}	Scalar	Probability that the p^{th} beamlet of the b^{th} beam belongs to the s^{th} segment. $$u_b = \begin{bmatrix} u_{b,p=1,s=1} & \cdots & u_{b,p=1,s=n_s} \\ \vdots & \ddots & \vdots \\ u_{b,p=n_p,s=1} & \cdots & u_{b,p=n_p,s=n_s} \end{bmatrix}$$ $$u_{bs} = \begin{bmatrix} u_{b,p=1,s} \\ \vdots \\ u_{b,p=n_p,s} \end{bmatrix}$$ $$u_{bp} = \begin{bmatrix} u_{b,p,s=1} \\ \vdots \\ u_{b,p,s=n_s} \end{bmatrix}$$
Other Data		
A	Matrix	Fluence-to-dose transformation matrix from all selected beams.

A_b	Matrix	Fluence-to-dose transformation matrix for the b^{th} beam. $A = [A_{b=1} \quad \cdots \quad A_{b=n_b}]$
W	Matrix	Weighting matrix for structures of interest.
d	Vector	Prescription dose the PTV and zero elsewhere.
$D^{\parallel}$	Matrix	Derivative matrix for all fluence maps from all beams. Takes derivative parallel to the MLC leaf motion.
$D_b^{\parallel}$	Matrix	Derivative matrix for fluence map of b^{th} beam. Takes derivative parallel to the MLC leaf motion. $D^{\parallel} = \begin{bmatrix} D_{b=1}^{\parallel} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_{b=n_b}^{\parallel} \end{bmatrix}$
$D^{\perp}$	Matrix	Derivative matrix for all fluence maps from all beams. Takes derivative perpendicular to the MLC leaf motion.
$D_b^{\perp}$	Matrix	Derivative matrix for fluence map of b^{th} beam. Takes derivative perpendicular to the MLC leaf motion.

		$D^{\perp} = \begin{bmatrix} D^{\perp}_{b=1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D^{\perp}_{b=n_b} \end{bmatrix}$
q	Matrix	An intermediate variable that has the same structure and subscript notation as u . Is used in the appendix, section A.3. $q_{bps} = \frac{1}{2}\left\|x_{bp} - c_{bs}\right\|^2$

Table 3.1: Important notations and data structures used in the study.

Term 1 is the dose fidelity term which attempts to optimize the fluence, x, such that the projected dose, Ax, is penalized for deviations from the prescription dose, d. The structures of interest are weighted by the diagonal matrix, W. Term set 2 is the anisotropic total variation regularization on the fluence maps to promote piecewise continuous fluence maps[28]. The matrices, $D^{\parallel}$ and $D^{\perp}$, take derivatives parallel and orthogonal to the MLC leaf direction.

Term sets 3 and 4 belong to the multiphase piecewise-constant Mumford-Shah formulation[18-20,22-25], which partitions each fluence map into n_s discrete segments. The function, for this study, utilizes the anisotropic version of total variation to account for the MLC leaf direction. The convexly relaxed segmentation labeling array, u, has been successfully applied to the Mumford-Shah formulation for piecewise constant segmentation[20,25]. The rows of the labeling array, u_{bp},

are subject to the probability simplex, which is described by two constraints on u_{bp}—the elements of u_{bp} are non-negative and their sum is equal to 1. This probability simplex constraint is applied in order to push u_{bps} to equal 1 for the segment where $\left|x_{bp} - c_{bs}\right|^2$ is the smallest, but the total variation regularization term encourages each segment in u to be piecewise continuous. The index for segment, s, starts at 0, to account for a "segment" with zero fluence or completely closed MLC. By integrating MLC segmentation with FMO in a single objective function, the beam apertures become optimizable variables directly contributing to the dosimetry.

Term set 3 is not convex, rendering the whole objective formulation non-convex. Fortunately, the formulation is multiconvex[26,27]—by updating one variable while holding the other two constant, we have a convex module. Similarly to a previous study[113], the algorithm is broken down into 3 modules, which are evaluated in an alternating block fashion.

3.2.1.1 Module 1

Module 1 updates x while holding u and c constant. The formulation for module 1 is written as

$$\begin{aligned} \underset{x}{argmin} \quad & \frac{1}{2}\left\| W\left(\sum_{b=1}^{n_b}(A_b x_b) - d\right)\right\|_2^2 + \sum_{b=1}^{n_b}\left(\lambda_1\left\|D_b^{\parallel}x_b\right\|_1 + \lambda_2\|D_b^{\perp}x_b\|_1\right) \\ & + \frac{\gamma}{2}\sum_{b=1}^{n_b}\sum_{p=1}^{n_p}\sum_{s=0}^{n_s} u_{bps}\left|x_{bp} - c_{bs}\right|^2 \\ subject\ to \quad & x \geq 0. \end{aligned}$$

Equation 3.2

The optimization formulation can be solved with a proximal-class, first-order primal-dual algorithm known as the Chambolle-Pock algorithm[20,135]. This study uses the overrelaxed version of the algorithm[136] for faster convergence rate, as well as a preconditioning process[135] to select the step size.

3.2.1.2 Module 2

Module 2 updates c, while holding x and u constant. Until after the initial update of u, c is first initialized to be

$$c_{bs} = \frac{s}{n_s} max(x_b) \qquad for\ all\ b, s,$$

Equation 3.3

and the subsequent iterations will update c using a closed form solution[19]

$$c_{bs} = \frac{u_{bs}^T x_b}{u_{bs}^T \vec{\mathbf{1}}} \qquad for\ all\ b, s.$$

Equation 3.4

In the case where $u_{b,s=k}^T \vec{1} = 0$, its corresponding $c_{b,s=k}$ takes on an average value from its neighbor values. In other words $c_{b,s=k} = \frac{1}{2}\left(c_{b,s=k-1} + c_{b,s=k+1}\right)$. If this turned off segment

happens to be the last segment, meaning that $u_{b,s=n_s}^T \vec{1} = 0$, then $c_{b,s=n_s} = 2 * c_{b,s=n_s-1}$. By default, $c_{b,s=0}$ will always be set at zero to define an "off" segment. This strategy is implemented to assign an otherwise undefined $c_{b,s=k}$ a unique non-zero value, which may be effective in finding useful segments.

3.2.1.3 Module 3

Module 3 updates u, while holding x and c, constant. The terms in the objective setting that have u are simply the terms that belong to the Mumford-Shah function:

$$\underset{u}{argmin} \quad \sum_{b=1}^{n_b}\left(\frac{1}{2}\sum_{p=1}^{n_p}\sum_{s=0}^{n_s} u_{bps}\left|x_{bp}-c_{bs}\right|^2 + \sum_{s=1}^{n_s}\left(\omega_1\left\|D_b^{\parallel}u_{bs}\right\|_1 + \omega_2\|D_b^{\perp}u_{bs}\|_1\right)\right)$$

$$subject\ to \qquad u \geq 0$$

$$\vec{1}^T u_{bp} = 1,$$

Equation 3.5

where γ weighting is ignored during module 3 since it does not influence the outcome of the optimization in Equation 3.5. The formulation shown in Equation 3.5 can also be efficiently evaluated using the preconditioned overrelaxed Chambolle-Pock algorithm.

3.2.1.4 Alternating module schedule and other heuristics to handle non-convexity

In order to allow for a change in the fluence to immediately affect the average value of its aperture, module 1 and module 2 are alternated back-to-back, updating c as soon as x is updated. After the active modules 1 and 2 update x and c, module 3 will then update u. This routine would repeat until the aperture shapes converged to a constant shape. After analysis with a test case, the number of subiterations for each module was decided to be 100 for modules 1 and 2, and 1500 for module 3.

Once the aperture shapes have converged, a polishing step is added where the intensity values are directly optimized while the apertures are locked. The polishing optimization takes the form of

$$\begin{aligned} &\underset{c}{argmin} && \frac{1}{2}\left\| W\left(\sum_{b=1}^{n_b}(A_b u_b c_b) - d\right)\right\|_2^2 \\ &subject\ to && c \geq 0, \end{aligned}$$

Equation 3.6

where the aperture intensities c are the only optimization variables. The polishing step allows for fine tuning of the aperture intensities that has been shown to improve the dosimetry. Ultimately, the aperture shapes and intensities are obtained from u and c.

To ensure an acceptable local minimum that emphasizes the dose fidelity, a graduated weighting technique was implemented, where γ starts at 0 and is monotonically increased over the course

of the optimization, until it reaches a max value γ_{max}. The idea behind this technique is to start with the convex part of the objective function that includes the dose fidelity term. During the optimization, the non-convexity is slowly introduced into the problem, ensuring that the optimized fluence segments still result in high dose fidelity. The γ was chosen to monotonically increase via the equation $\gamma = \gamma_{max} \cdot \min\left(1, \left(\frac{k}{\hat{k}}\right)^8\right)$, where k is the current subiteration number for the updates involving modules 1 and 2, and $\hat{k}$ is a set subiteration number when $\gamma = \gamma_{max}$ for the first time. For all patients in this study, $\hat{k}$ was set to 500, which would allow for the alternating module schedule to cycle through 5 times before $\gamma = \gamma_{max}$. Note that γ is not updated while module 3 is active, since γ does not directly influence the outcome of module 3.

3.2.2 Simulated Annealing DAO

Simulated annealing is a stochastic technique developed for non-convex optimization[7], and has been successfully used for DAO[8-14]. We developed simulated annealing based DAO for comparison. Starting from an initial set of apertures such as the conformal aperture from beams' eye view, the method randomly selects either to change an aperture intensity or a leaf position for modification. A random number from a Gaussian distribution is sampled to determine the size and direction of the change. The standard deviation, SD, of the Gaussian is defined as

$$SD_{leaf} = 1 + (\vartheta - 1)\frac{1}{(n_{succ} + 1)^{1/T_0^{step}}}$$

$$SD_{intensity} = (\vartheta - 1)\frac{1}{(n_{succ} + 1)^{1/T_0^{step}}},$$

Equation 3.7

where ϑ is the initial value, n_{succ} is the number of successes, and T_0^{step} defines the cooling rate[7]. A success is defined as a change that results in a decrease of the cost function. Since the leaf position is quantized by beamlets, SD_{leaf} is defined such that the smallest possible value is 1 beamlet. On the other hand, $SD_{intensity}$ is allowed to diminish to 0.

Once either the aperture intensity or the leaf position is randomly modified, a cost function is evaluated. For this study, the cost function used is the dose fidelity term of Equation 3.1:

$$minimize \qquad \frac{1}{2}\left\| W\left(\sum_{b=1}^{n_b}(A_b x_b) - d\right)\right\|_2^2,$$

Equation 3.8

where x_b a function of the leaf positions and aperture intensities—the variables that simulated annealing will be affecting directly. The theoretical leaf positions are constrained such that the two leaves on the same row will not overlap, and the aperture intensities are constrained to be non-negative. While the cost function in Equation 3.8 is convex, the conversion of the leaf positions and aperture intensity information to x_b is non-convex. Any change that results in decreasing the cost value is accepted. If the change increases the cost value, the change can still be accepted with a probability defined as

$$P_{accept} = \phi \frac{1}{(n_{succ}+1)^{1/T_0^{prob}}},$$

Equation 3.9

where ϕ is the initial probability of acceptance, and T_0^{prob} defines the cooling rate. The motivation for allowing changes that increase the cost function is to allow for the jumping out of local minima.

3.2.3 Evaluation

Three planning cases—one glioblastoma multiforme (GBM) patient, one lung (LNG) patient, and one head and neck (H&N) patient with 3 PTVs at different prescription levels—were evaluated in this study to test the feasibility of the optimization formulation. Using a collapsed cone convolution/superposition code with a 6 MV polyenergetic kernel—with a dose array resolution of $(0.25\ \text{cm})^3$ and a beamlet size of $(0.5\ \text{cm})^2$—the beamlet dose was calculated for 1162 beams, evenly spaced across the 4π steradian. After beam angles that would cause gantry to couch/patient collisions were removed, a column generation and pricing approach[17,31,32,68,92-94] was utilized to automatically select 20 beam angles for each patient. The beamlet dose of the selected beams are stored in the dose array, A, for optimization. A shell structure around the PTV and skin structure were added to every patient to reduce dose spillage outside the PTV. The thickness of the shell structure was calculated based on the equation: $thickness_{shell} = \sqrt[3]{\frac{3}{4\pi}TV}$, where TV is the target volume.

Patient	**Number of allowed segments per beam (n_s)**	**Prescription dose (Gy)**	**PTV volume (cc)**

Glioblastoma Multiforme	10	30	57.77
Lung	10	50	47.84
Head & Neck	20	54	197.54
		59.4	432.56
		69.96	254.98

Table 3.2: Number of allowed segments per beam, prescription dose, and PTV volume for each patient.

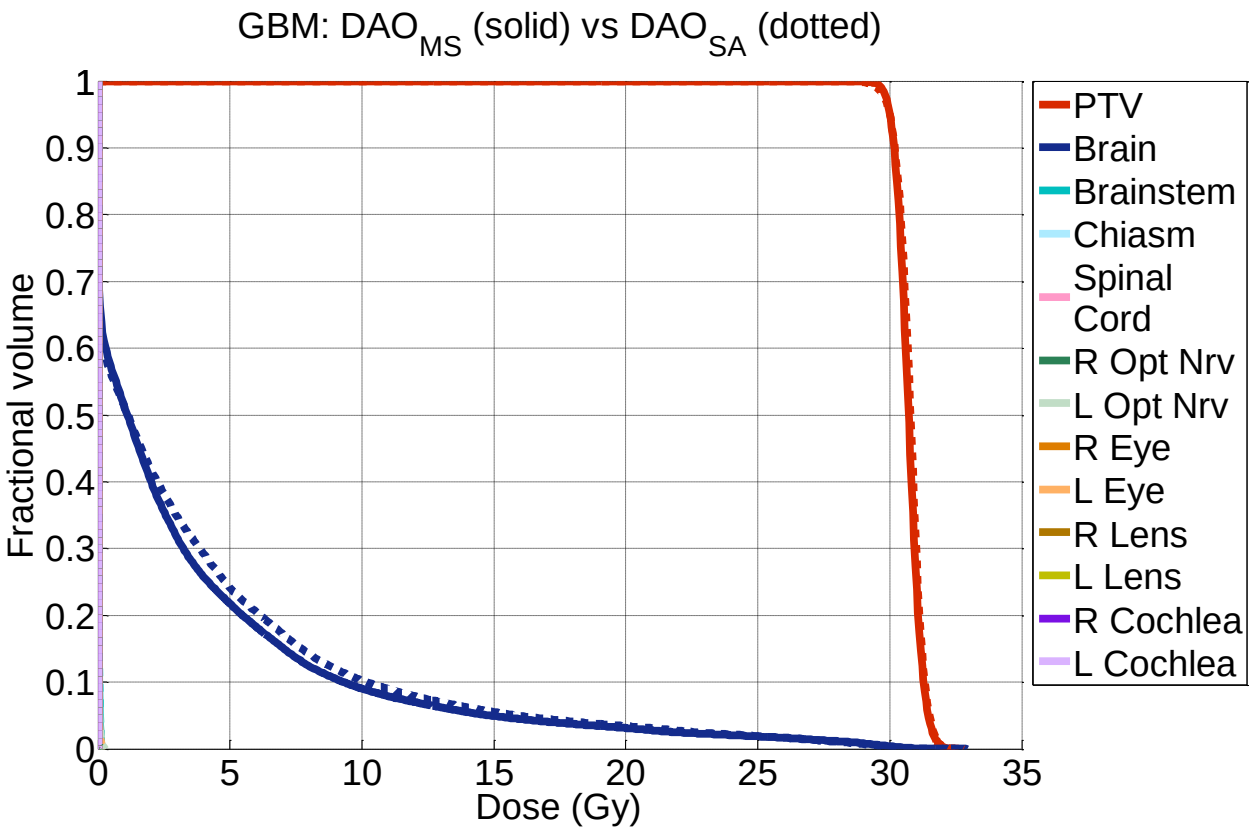
Patient plans were created using Mumford-Shah based DAO (DAO_{MS}) and simulated annealing DAO (DAO_{SA}). The same beam angles and number of allowed segments per beam are used for both DAO_{MS} and DAO_{SA} for the same patient. To ensure an unbiased plan comparison, the weighting matrices W of DAO_{MS} and DAO_{SA} plans are iteratively updated to outperform each other in terms of OAR sparing and PTV coverage and homogeneity until further improvement in one aspect is impossible without sacrificing the other. At the end, with the same fidelity term, the structure weightings used by both the DAO_{MS} plans and the DAO_{SA} plans were found to be similar.

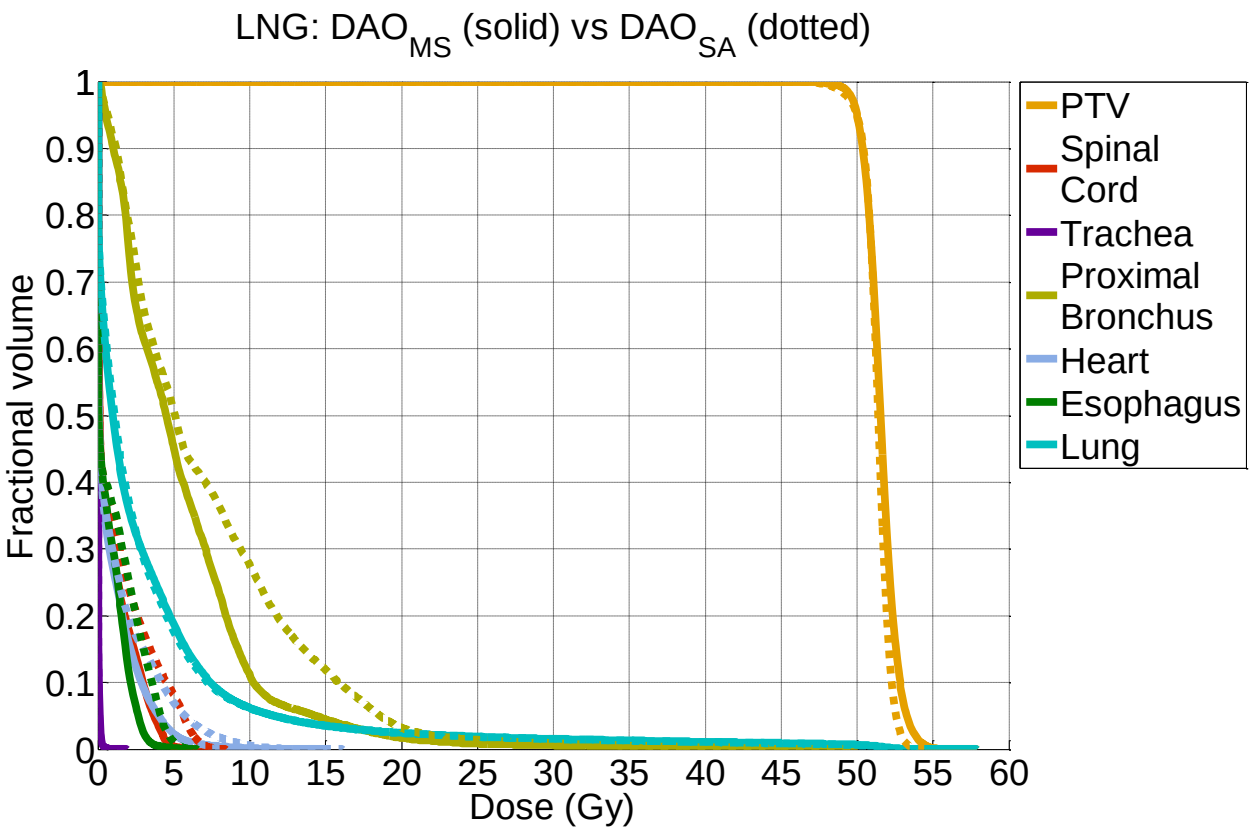
For evaluation, all treatment plans were normalized such that 100% of the prescription dose is delivered to 95% of the PTV. For the H&N case, the 69.96 Gy PTV was used for normalization. PTV homogeneity $\left(\frac{D95}{D5}\right)$, D98, D99, and D_{max} were evaluated to assess the PTV coverage, hot spots, and homogeneity. OAR D_{max} and D_{mean} were assessed to determine dose sparing to the critical structures. D_{max} is defined as the dose at 2% of the structure volume, D2, which is

recommended by the ICRU-83 report[58]. To quantify the amount of high dose spillage to the normal tissue, R50, defined as the 50% isodose volume divided by the target volume, was determined. Lastly, the van't Reit conformation number (VRCN)[137] was also assessed. VRCN is defined as $\frac{TV_{RI}}{TV} \times \frac{TV_{RI}}{V_{RI}}$, where TV_{RI} is the target volume covered by the reference isodose, TV is the target volume, and V_{RI} is the volume of the reference isodose. VRCN is global conformity index that takes into account both the irradiation of the target volume and healthy tissues, and ranges from 0 to 1, with 1 being the perfect case. For the H&N case, the R50 and VRCN values were calculated using the sum of all 3 PTVs as the target volume and 54 Gy as the reference dose.

The computer used to solve the optimizations has 32 GB RAM, an NVIDIA GeForce GTX 690, and an Intel Core i7-3960X CPU, with 6 physical cores overclocked to 4.00 GHz.

3.3 Results


GBM: DAO$_{MS}$ (solid) vs DAO$_{SA}$ (dotted)
Fractional volume
Dose (Gy)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30
35
PTV
Brain
Brainstem
Chiasm
Spinal Cord
R Opt Nrv
L Opt Nrv
R Eye
L Eye
R Lens
L Lens
R Cochlea
L Cochlea


LNG: DAO_MS (solid) vs DAO_SA (dotted)
Fractional volume
Dose (Gy)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30
35
40
45
50
55
60
PTV
Spinal Cord
Trachea
Proximal Bronchus
Heart
Esophagus
Lung

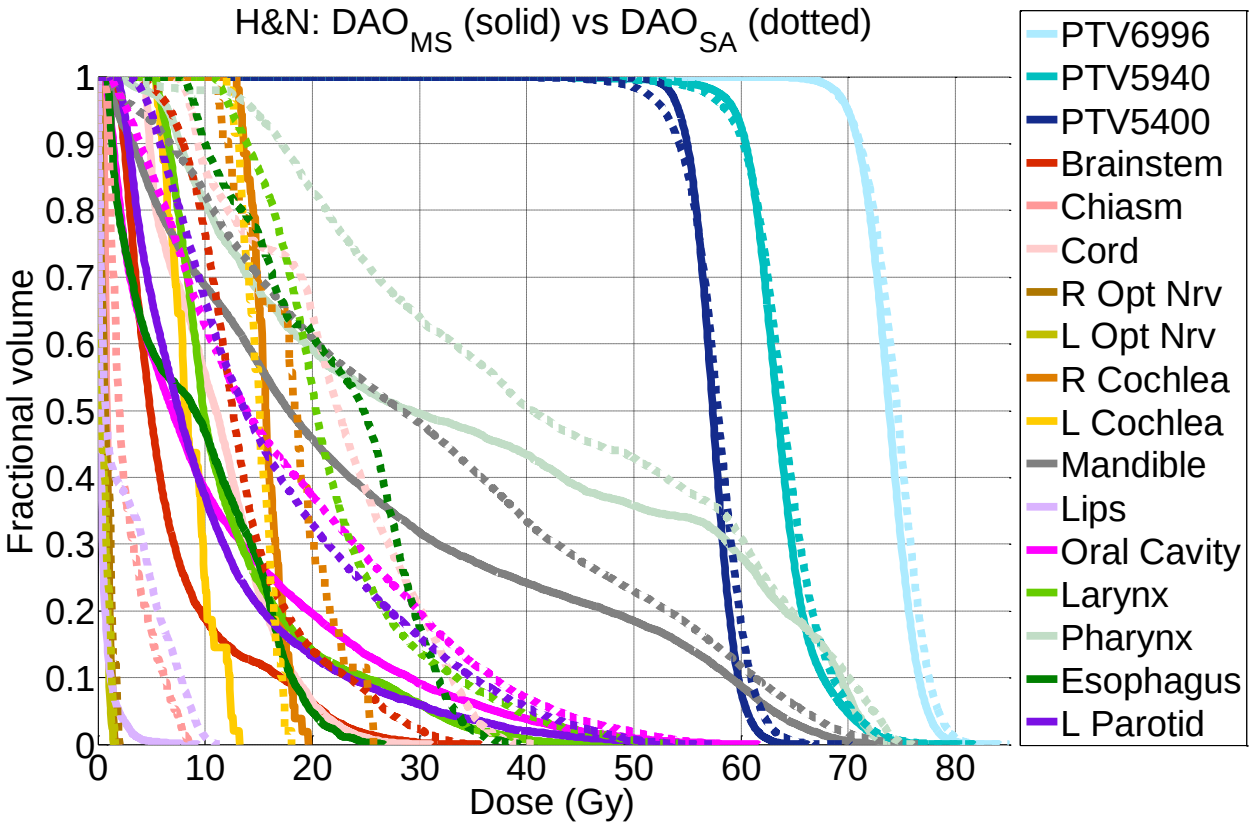


Figure 3.1: DVH comparisons of the GBM, LNG, and H&N cases.

Patient Case	**PTV Statistics**					**R50**		**VRCN**	
	Homogeneity		D98	D99	D_{max}				
	DAOMS	DAOSA	DDAOMS − DDAOSA (Gy)			DAOMS	DAOSA	DAOMS	DAOSA
GBM	0.954	0.952	+0.089	+0.174	–0.047	2.553	2.712	0.816	0.847
LNG	0.938	0.951	+0.294	+0.494	+0.935	2.840	2.680	0.831	0.892

H&N	54	0.895	0.853	+2.626	+3.771	–1.595	2.248	3.074	0.748	0.607
	59.4	0.850	0.833	+1.237	+1.448	–0.483				
	69.96	0.904	0.892	+0.153	–0.101	–1.054				
Average		0.908	0.896	+0.880	+1.157	–0.449	2.547	2.822	0.798	0.782

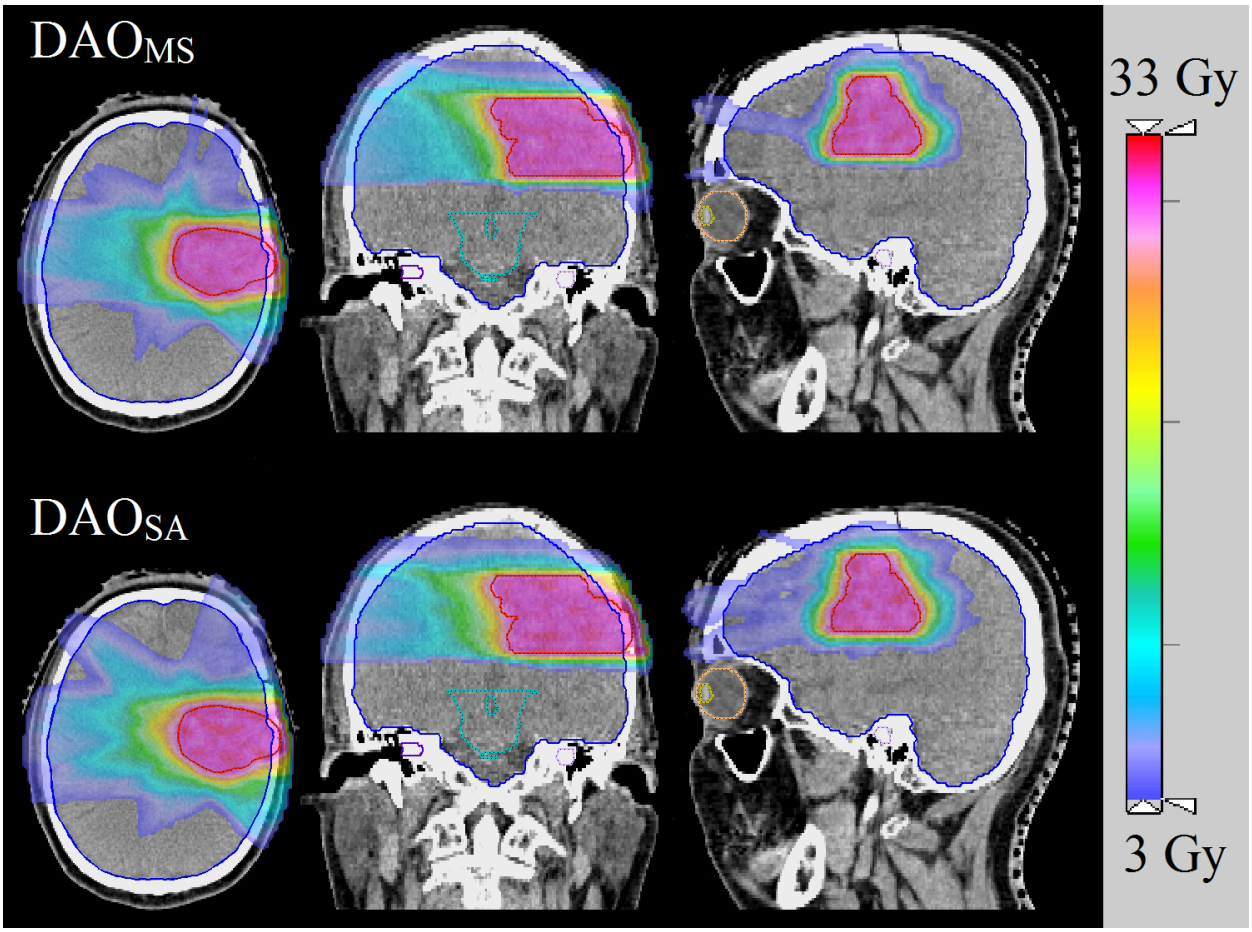
Table 3.3: Comparison of the PTV homogeneity, D98, D99, and D_{max}, as well as R50 and VRCN.

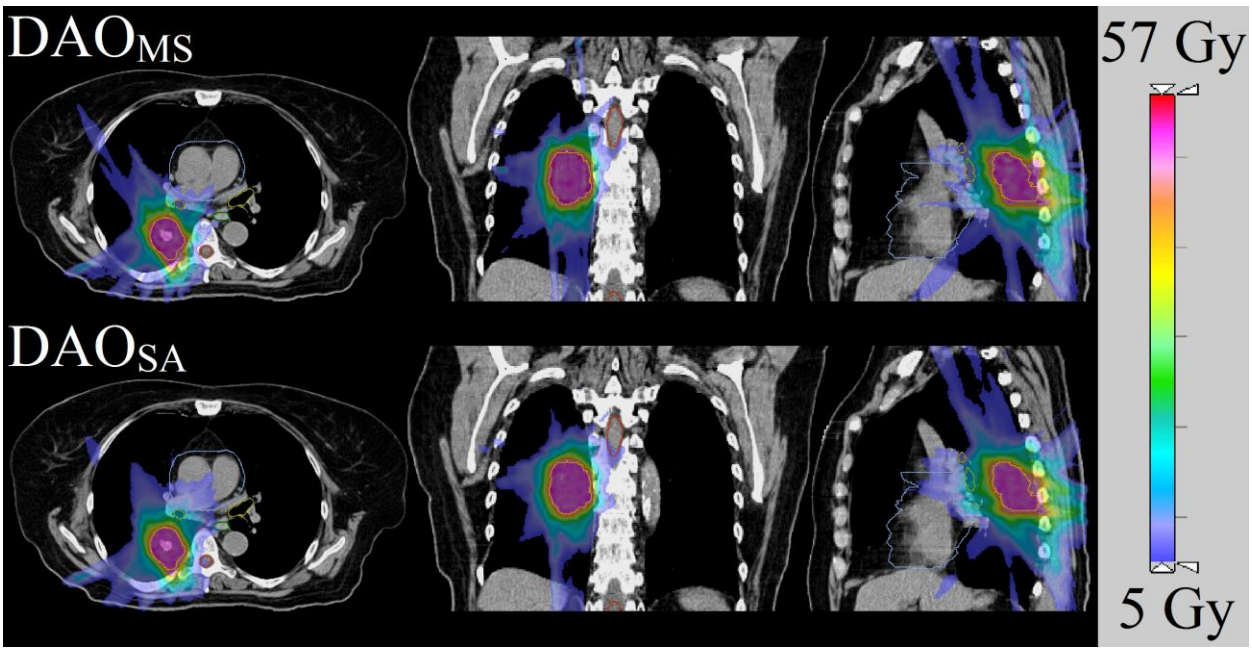
Dose Difference $\left(\mathbf{D_{DAO_{MS}}} - \mathbf{D_{DAO_{SA}}}\ (\mathbf{Gy})\right)$	**D_{max}**			**D_{mean}**		
	Largest value	Smallest value	Average value	Largest value	Smallest value	Average value
GBM	+0.021 Chiasm	–0.031 L Eye	–0.004	+0.003 Chiasm	–0.003 L Eye	–0.0004
LNG	+0.036 Trachea	–3.000 ProxBronch	–1.835	+0.004 Trachea	–1.593 ProxBronch	–0.541
H&N	–0.162 L Opt Nrv	–13.067 Spinal Cord	–5.891	–0.046 R Opt Nrv	–13.073 Esophagus	–5.714

Table 3.4: Largest, smallest, and average values found for $\left(\mathbf{D_{DAO_{MS}}} - \mathbf{D_{DAO_{SA}}}\ (\mathbf{Gy})\right)$ dose differences for $\mathbf{D_{max}}$ and $\mathbf{D_{mean}}$. Negative values represent dose sparing for $\mathbf{DAO_{MS}}$, while positive values represent dose sparing for $\mathbf{DAO_{SA}}$. OARs that received 0 Gy in both cases are excluded in the evaluation.

Figure 3.1 shows the DVHs of all patients. The GBM DAO_{MS} and DAO_{SA} plans are almost identical in OAR doses with difference less than 0.05% of the prescription dose. For the LNG case, the DVH shows that the DAO_{MS} method was able to better spare the proximal bronchus dose, but otherwise, the two methods performed similarly. On average, for the LNG case, the DAO_{MS} method was able to spare proximal bronchus D_{max} by 3.67% and D_{mean} by 1.08% of the prescription dose. For the more complicated H&N plan, the DAO_{MS} plan evidently improved OAR sparing compared to the DAO_{SA} plan, with as much as a 13 Gy reduction to the spinal cord. The improvement is consistent for all OARs in the H&N DAO_{MS} plan. On average, D_{max} and D_{mean} were reduced by 10.91% and 10.58% of the prescription dose. The OAR sparing is summarized in Table 3.4.

Table 3.3 shows PTV dose statistics, as well the dose spillage (R50) and the conformity number (VRCN). On average, the PTV was better covered using DAO_{MS}. D98 and D99 was increased by 1.66% and 2.21% of the prescription dose. The H&N plan had the largest improvement in dose coverage, with an increase as large as 3.77 Gy in D99 for the 54 Gy PTV. The average homogeneity of the two planning methods is similar with DAO_{MS} being slightly better by 0.012. The average R50 and VRCN was more clearly improved in the H&N case using DAO_{MS}. The R50 for the GBM and LNG cases were similar for the two plan types, differing by 0.16. For the H&N plan, R50 was lowered by 0.826 and VRCN was increased by 1.355 using DAO_{MS}, indicating substantially improved dose compactness and conformality.


DAOMS
DAOSA
33 Gy
3 Gy


DAOMS
DAOSA
57 Gy
5 Gy

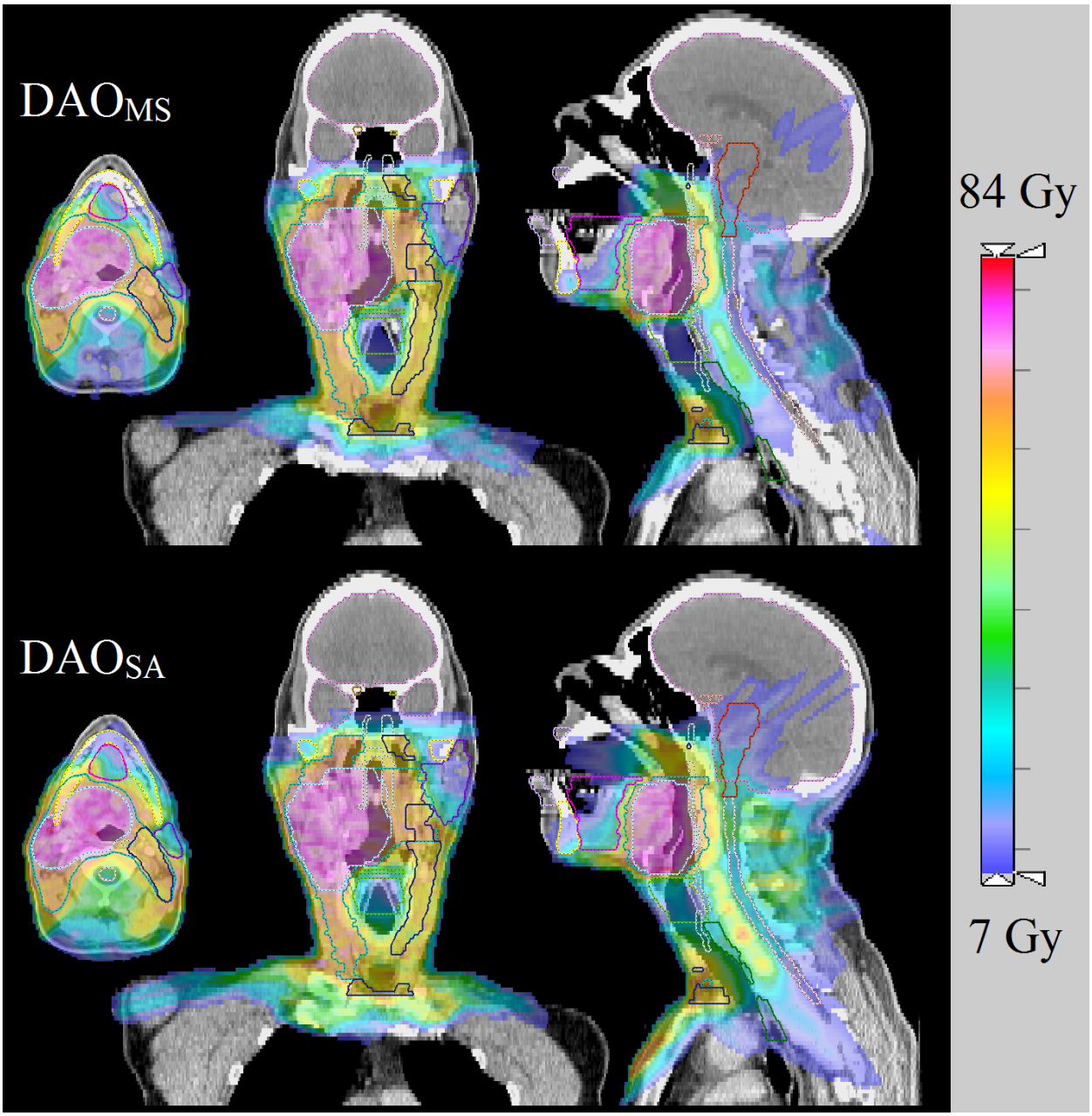


Figure 3.2: Dose color washes of the GBM, LNG, and H&N patients. The low dose cutoff for viewing was set to be 10% of the prescription dose.

Figure 3.1 shows the isodose comparison. The low dose cutoff for viewing was set at 10% of the prescription dose. Qualitatively, the dose distributions produced by the DAO_{MS} and the DAO_{SA}

methods for GBM and LNG plans are similar. On the other hand, the dose distribution differences of the H&N plans are more compelling. The dose is visually more compact in the DAO_{MS} plan, with lower dose to the brainstem, spinal cord, and posterior neck region in the sagittal slice, as well as the larynx and left parotid in the coronal slice.

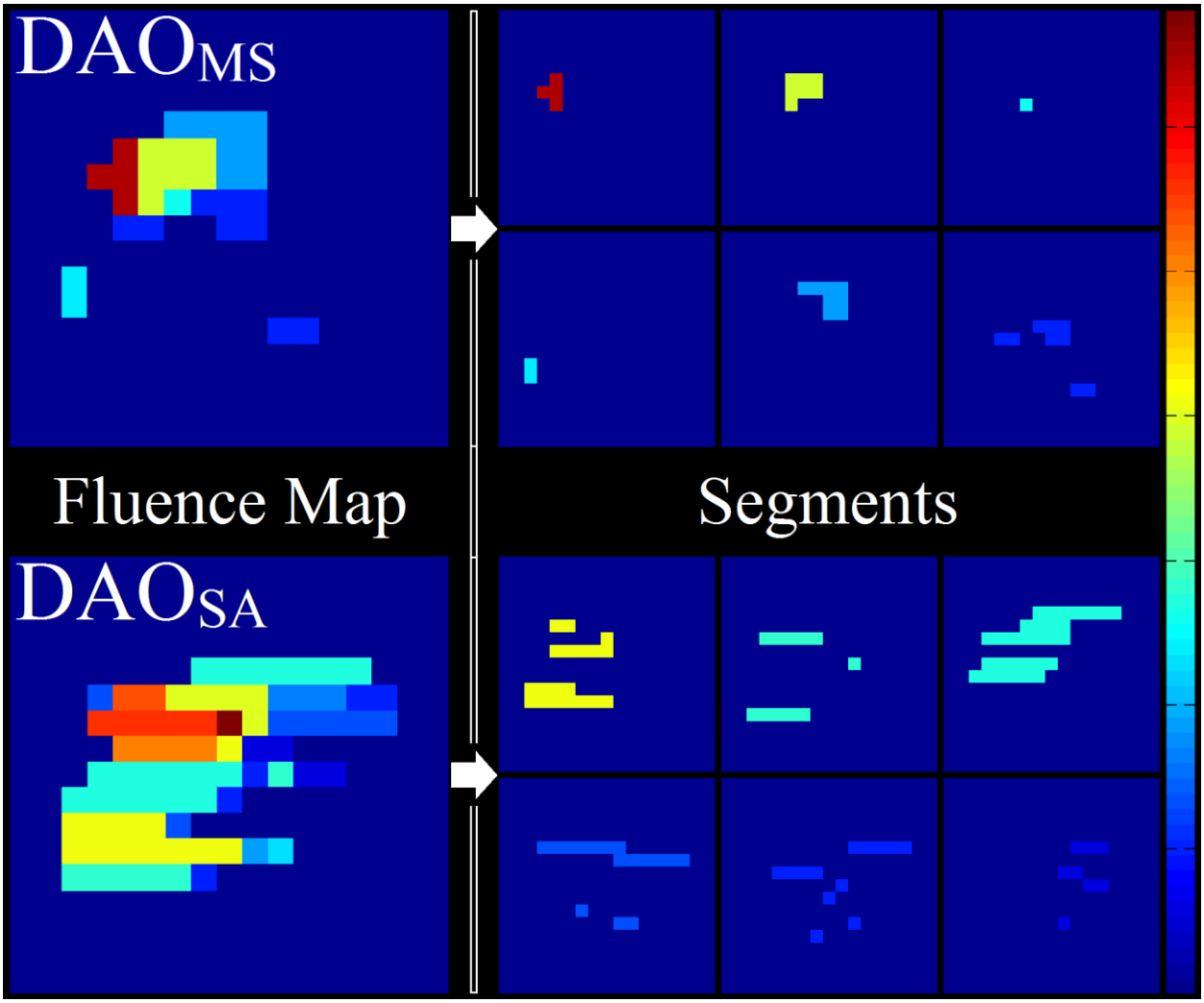


Figure 3.3: Schematic of an optimized fluence from each plan type, and the breakdown of the fluences to their segments.

	Allowed number of segments per beam (n_s)	**Average number of solved segments per beam**		**Mean number of beamlets in a segment**	
		$\mathrm{DAO_{MS}}$	$\mathrm{DAO_{SA}}$	$\mathrm{DAO_{MS}}$	$\mathrm{DAO_{SA}}$
GBM	10	6.10	7.80	6.72	17.51
LNG	10	9.40	9.15	4.93	14.93
H&N	20	16.00	10.20	8.84	50.50

Table 3.5: Aperture statistics for the $\mathrm{DAO_{MS}}$ and the $\mathrm{DAO_{SA}}$ methods for the patient cases.

Figure 3.3 shows an example of MLC segments and resultant fluence map for the GBM patient. The pairwise disjoint nature of the piecewise constant Mumford-Shah formulation is apparent in the segmentation results, as the $\mathrm{DAO_{MS}}$ segments do not overlap. The $\mathrm{DAO_{SA}}$ MLC segments, on the other hand, can overlap, resulting in more complex fluence. On average, the $\mathrm{DAO_{MS}}$ segments contain fewer beamlets than that of $\mathrm{DAO_{SA}}$. Interestingly, Table 3.5 shows that none of the planning methods utilized the maximal allowed segments for delivery.

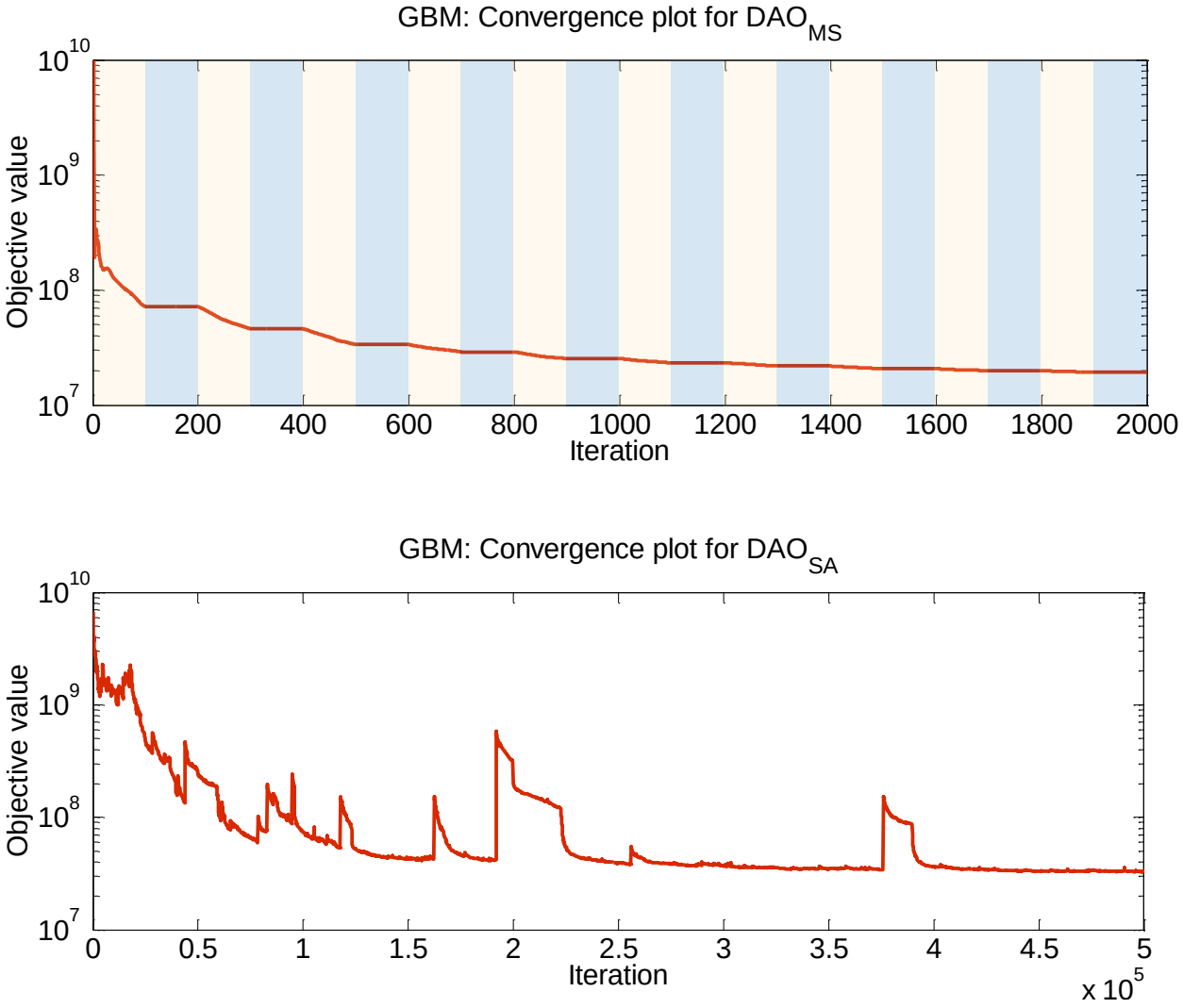


Figure 3.4: Example convergence plots for the GBM case. For DAO$_{MS}$, the yellow regions represent where module 1 and module 2 are updating x and c at each subiteration. The blue region represents where module 3 is updating u, and each data point in the blue region represents 15 subiterations. Objective value calculation for DAO$_{MS}$ uses the γ_{max} value for all iterations.

	Solve time (s)	
	DAO_{MS}	DAO_{SA}

GBM	599.8	5719.6
LNG	647.1	25747.7
H&N	2170.2	35682.8

Table 3.6: Solve time for one optimization run for each case and plan type. DAO_{MC} solve times include the polishing step.

DAO_{MS} is substantially faster than DAO_{SA}, with a 9.5 to 40 fold decrease in total solve time. Figure 3.4 shows example convergence plots for the GBM case. The DAO_{MS} approach is able to consistently lower the cost function, since each module is a convex subproblem. The stochastic property of simulated annealing causes the cost function to randomly increase at times, while attempting to escape from local minima. Figure 3.4 and Table 3.6 highlight the strengths of the multiconvex approach to quickly and deterministically reach a minimum that contains a high quality dosimetric plan.

3.4 Discussion

In this study, we developed a novel direct aperture optimization technique that integrates fluence map optimization—employing L2-norm fidelity and L1-norm regularization terms—with the multiphase piecewise constant Mumford-Shah method. We utilized a first-order primal-dual proximal-class algorithm, known as the Chambolle-Pock algorithm, to solve the multiconvex direct aperture optimization problem. This novel method simultaneously solves for the fluence segments while minimizing the dose fidelity term offers a number of advantages.

The optimization formulation offers an elegant theoretical description of the physical problem. The fidelity term penalizes deviations from the prescription dose, the total variation term promotes piecewise smoothness on the fluence map, and the piecewise constant Mumford-Shah formulation defines segments that constitute the fluence map. As a non-convex optimization, this method cannot guarantee to find the global minima of the problem. However, with the graduated weighting technique on the non-convex term, the optimization is able to find a local minimum that is biased to satisfying to dose fidelity term. By slowly introducing the non-convexity, optimizer can begin to form segments for the fluence map that indirectly penalize deviations between the projected dose and the prescription dose.

In addition to its theoretical appeal, the deterministic multiconvex approach offers practical advantages in computational speed and dosimetric performance, compared to the existing stochastic simulated annealing method for DAO. The DAO_{SA} method struggled to include more apertures in the optimization but the explicit addition is heavily penalized by computational time. Given the vast number of possible apertures as a combination of beamlets, the previous methods are inherently inefficient. Our method, on the other hand, segments the MLC apertures *on the fly*, thus avoiding being limited to a small subset of possible apertures. As hypothesized, the advantage is shown more evident for the multilevel head and neck plan requiring complex modulation.

The Chambolle-Pock algorithm was selected to solve the optimization problem for several reasons. As a proximal-class algorithm, it can solve many types of non-differentiable optimization formulations exactly, such as formulations involving the L1-norm. The algorithm is highly efficient on memory usage and computation cost because it does not require a to solve system of linear equations involving the fluence to dose transformation matrix at every iteration,

contrasting to other methods such as alternating direction method of multipliers (ADMM)[138]. The computation cost of Chambolle-Pock relies on the simple multiplication of a large matrix and its transpose at each iteration.

While the total variation regularization promotes apertures to be MLC deliverable, the DAO_{MS} method does not explicitly guarantee that the resulting aperture is MLC deliverable in 1 segment. For example, in Figure 3.3, the last segment of the DAO_{MS} plan is not deliverable for horizontal MLCs. The simplest solution is to break down the segment into 2 delivery segments, which should minimally add to the treatment time since only a small fraction of solved segments are undeliverable in this manner.

A drawback of the current approach is that the Mumford-Shah formulation is designed to describe non-overlapping segments. The ability to have overlapping segments would allow more complex fluences with fewer segments. Furthermore, enforcing pairwise disjoint regions results in overall smaller segments, shown in Table 3.5, since each segment is now competing for space with the other segments. Although the drawback did not prevent DAO_{MS} from outperforming the stochastic approach, particularly in complex cases, there is clearly space for improvement. As a topic of further investigation, the smaller segment size problem may be alleviated by devising an improved segmentation algorithm to divide the DAO_{MS} fluence into fewer and larger overlapping segments, without modifying the current fluence.

Another interesting observation one can make from the result is that both methods did not use the maximal allowed number of segments. For the DAO_{MS} plans, this is due to the total variation term that reduces the fluence map complexity and subsequently fewer segments. The DAO_{SA} method, on the other hand, is not regularized to penalize more segments. Given that DAO_{SA}

plans are suboptimal compared to the DAO_{MC} plans, this reflects the weakness of using simulated annealing method to search a larger aperture solution space.

The run time of the DAO_{SA} method used in this study is noticeably longer than the simulated annealing implemented commercial software. The longer computational time can be attributed to the following reasons. First, the plans used 20 beams, which increased the fluence-to-dose transformation matrix size. Second, the fluence-to-dose transformation matrices in this study were not downsampled. Last, the plans allowed many more segments than what would be allowed in a commercial planning software.

While considerably faster the the DAO_{SA} method, the DAO_{MC} method is still too slow and unusable for on-demand optimization. The single largest cost per iteration in the Chambolle-Pock algorithm is the multiplication of the fluence-to-dose transformation matrix on the primal and dual variable. Cutting down the matrix size, via methods such as downsampling or removing some penumbra information, may yield faster optimization times that may be adequate for on-demand optimizaiton.

3.5 Conclusion

A novel deterministic direct aperture optimization formulation combining fluence map optimization and multiphase piecewise constant Mumford-Shah segmentaion into a unified framework was proposed and evaluated. The new approach enables generating MLC segments on the fly without being limited to a small subset of possible apertures as previous methods did. The non-convex optimization formulation was split into multiple convex modules, and solved alternatingly using a first-order primal-dual proximal-class algorithm. The new deterministic

method solved the DAO problem is considerably faster than the simulated annealing method and is dosimetrically superior, particularly for the complex head and neck case.

4 APPENDIX

THE CHAMBOLLE-POCK ALGORITHM

4.1 Problem Formulation

The Chambolle-Pock algorithm, a first-order primal dual algorithm[20,21], has been successfully implemented for all of the problems presented in this disseration[28,68,98,113]. The algorithm solves a primal problem of the canonical form

$$minimize\ F(Kx) + G(x),$$

Equation 4.1

where F and G are lower semicontinuous functions and K is a matrix. To solve an optimization problem with Chambolle-Pock, the optimization must be rewritten to fit this canonical form. All optimizations in this dissertation that use the Chambolle-Pock algorithm can easily be expressed in this canonical form.

4.2 Overrelaxed Algorithm

The overrelaxed version of the Chambolle-Pock algorithm[136] solves the primal problem via the iteration

$$\bar{x}^{n+1} = prox_{\tau G}(x^n - \tau K^T z^n)$$

$$\bar{z}^{n+1} = prox_{\sigma F^*}\left(z^n + \sigma K\left(\bar{x}^{n+1} + \theta(\bar{x}^{n+1} - x^n)\right)\right)$$

$$x^{n+1} = \rho\bar{x}^{n+1} + (1-\rho)x^n$$

$$z^{n+1} = \rho\bar{z}^{n+1} + (1-\rho)z^n,$$

Equation 4.2

where z variable of the dual problem: $minimize\ G^*(-K^T z) + F^*(z)$. The Chambolle-Pock algorithm is solving both the primal and dual problem simultaneously. The function, F^*, is the convex conjugate of F, defined as $F^*(z) = \sup_y \left(z^T y - F(y)\right)$. The overrelaxation parameter, $\rho \in [0,2]$, can be adjusted to control the convergence. The algorithm degenerates to the original algorithm when $p = 1$. The parameter, θ, is typically set to 1. As a proximal-class algorithm[139], the Chambolle-Pock algorithm relies on the "proximal mapping", or "prox operator", defined as

$$prox_{th}(x) = \underset{v}{argmin}\left(h(v) + \frac{1}{2t}\|v - x\|_2^2\right),$$

Equation 4.3

where h is a lower semicontinous function and t is a parameter that serves as a step size. Intuitively, $prox_{th}(x)$ will try to find an argument that minimizes h, without straying too far from x. The efficacy of proximal algorithms relies on that the function h has a simple evaluation with the prox operator. A useful property of the prox operator is the separable sum rule,

$$prox_{\sigma F^*}\left(\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}\right) = \begin{bmatrix} prox_{\sigma f_1^*}(z_1) \\ \vdots \\ prox_{\sigma f_n^*}(z_n) \end{bmatrix},$$

Equation 4.4

where $F\left(\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right) = f_1(y_1) + \cdots + f_n(y_n)$, and this allows us to break down $F(Kx)$ into a sum of convex functions. Another valuable property is the Moreau decomposition,

$$prox_{\sigma F^*}(z) = z - \sigma \cdot prox_{\sigma^{-1}F}(\sigma^{-1}z),$$

Equation 4.5

which allows us to calculated the prox operator of the convex conjugate of a function without having to evaluate the convex conjugate itself.

4.3 Step Size Selection

In the algorithm, we have the parameters τ and σ, which serve as our step sizes during the optimization. These can be selected based on a diagonal preconditioning approach[135]:

$$\tau_j = \frac{1}{\sum_{i=1}^{i=n_{rows}} K(i,j)} \qquad \sigma_i = \frac{1}{\sum_{j=1}^{j=n_{cols}} K(i,j)}$$

$$\tau = \begin{bmatrix} \tau_{j=1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau_{j=n_{cols}} \end{bmatrix} \qquad \sigma = \begin{bmatrix} \sigma_{i=1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{i=n_{rows}} \end{bmatrix},$$

Equation 4.6

where τ and σ are diagonal matrices with the same number of diagonal elements as the length as x and z, respectively. The diagonal preconditioning guarantees convergence under the assumption that $F(Kx)$ and $G(x)$ have sum separable operations. When using the preconditioned τ and σ, the mathematical operations involving these step sizes are performed elementwise. In the case where these functions are not sum separable, but are block separable, it is possible to assign the block of variables the smallest step size—found during the preconditioning process—within its block. Otherwise, τ and σ can be scalars and follow the relation, $\tau\sigma\|K\|^2 \leq 1$, to guarantee convergence. The operator norm, $\|K\|$, can be estimated via power iteration[140].

5 REFERENCES

1. Khan FM, Gibbons JP. *Khan's the physics of radiation therapy.* Lippincott Williams & Wilkins; 2014.
2. Brahme A. Optimization of stationary and moving beam radiation therapy techniques. *Radiotherapy and Oncology.* 1988;12(2):129-140.
3. Bortfeld T, Bürkelbach J, Boesecke R, Schlegel W. Methods of image reconstruction from projections applied to conformation radiotherapy. *Physics in Medicine and Biology.* 1990;35(10):1423.
4. Webb S. Optimisation of conformal radiotherapy dose distribution by simulated annealing. *Physics in Medicine and Biology.* 1989;34(10):1349.
5. Convery D, Rosenbloom M. The generation of intensity-modulated fields for conformal radiotherapy by dynamic collimation. *Physics in Medicine and Biology.* 1992;37(6):1359.
6. Xia P, Verhey LJ. Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments. *Medical Physics.* 1998;25(8):1424-1434.
7. Kirkpatrick S. Optimization by simulated annealing: Quantitative studies. *Journal of statistical physics.* 1984;34(5-6):975-986.
8. Earl M, Shepard D, Naqvi S, Li X, Yu C. Inverse planning for intensity-modulated arc therapy using direct aperture optimization. *Physics in medicine and biology.* 2003;48(8):1075.
9. Shepard DM, Earl MA, Li XA, Naqvi S, Yu C. Direct aperture optimization: A turnkey solution for step-and-shoot IMRT. *Medical Physics.* 2002;29(6):1007-1018.

10. Earl M, Afghan M, Yu C, Jiang Z, Shepard D. Jaws-only IMRT using direct aperture optimization. *Medical physics.* 2007;34(1):307-314.

11. Bergman AM, Bush K, Milette M-P, Popescu IA, Otto K, Duzenli C. Direct aperture optimization for IMRT using Monte Carlo generated beamlets. *Medical physics.* 2006;33(10):3666-3679.

12. Sui H, CA PIC. Direct Aperture Optimization.

13. Zhang G, Jiang Z, Shepard D, Zhang B, Yu C. Direct aperture optimization of breast IMRT and the dosimetric impact of respiration motion. *Physics in medicine and biology.* 2006;51(20):N357.

14. Jiang Z, Earl MA, Zhang GW, Yu CX, Shepard DM. An examination of the number of required apertures for step-and-shoot IMRT. *Phys Med Biol.* 2005;50(23):5653-5663.

15. Men C, Romeijn HE, Taşkın ZC, Dempsey JF. An exact approach to direct aperture optimization in IMRT treatment planning. *Physics in Medicine and Biology.* 2007;52(24):7333.

16. Salari E, Unkelbach J. A column-generation-based method for multi-criteria direct aperture optimization. *Physics in medicine and biology.* 2013;58(3):621.

17. Romeijn H, Ahuja R, Dempsey J, Kumar A. A Column Generation Approach to Radiation Therapy Treatment Planning Using Aperture Modulation. *SIAM Journal on Optimization.* 2005;15(3):838-862.

18. Mumford D, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics.* 1989;42(5):577-685.

19. Vese LA, Chan TF. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International journal of computer vision.* 2002;50(3):271-293.

20. Chambolle A, Pock T. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision.* 2011;40(1):120-145.

21. Pock T, Cremers D, Bischof H, Chambolle A. An algorithm for minimizing the Mumford-Shah functional. Paper presented at: Computer Vision, 2009 IEEE 12th International Conference on2009.

22. Chan TF, Vese LA. Active contours without edges. *IEEE Transactions on image processing.* 2001;10(2):266-277.

23. Esedoglu S, Tsai Y-HR. Threshold dynamics for the piecewise constant Mumford–Shah functional. *Journal of Computational Physics.* 2006;211(1):367-384.

24. Chan TF, Vese LA. A level set algorithm for minimizing the Mumford-Shah functional in image processing. Paper presented at: Variational and Level Set Methods in Computer Vision, 2001. Proceedings. IEEE Workshop on2001.

25. Chan TF, Esedoglu S, Nikolova M. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM journal on applied mathematics.* 2006;66(5):1632-1648.

26. Shen X, Diamond S, Udell M, Gu Y, Boyd S. Disciplined Multi-Convex Programming. *arXiv preprint arXiv:160903285.* 2016.

27. Xu Y, Yin W. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences.* 2013;6(3):1758-1789.

28. Nguyen D, O'Connor D, Yu VY, et al. Dose domain regularization of MLC leaf patterns for highly complex IMRT plans. *Medical Physics.* 2015;42(4):1858-1870.

29. Voet PW, Breedveld S, Dirkx ML, Levendag PC, Heijmen BJ. Integrated multicriterial optimization of beam angles and intensity profiles for coplanar and noncoplanar head and neck IMRT and implications for VMAT [published online ahead of print 2012/08/17]. *Med Phys.* 2012;39(8):4858-4865.

30. Kim H, Li R, Lee R, et al. Dose optimization with first-order total-variation minimization for dense angularly sampled and sparse intensity modulated radiation therapy (DASSIM-RT) [published online ahead of print 2012/07/27]. *Med Phys.* 2012;39(7):4316-4327.

31. Dong P, Lee P, Ruan D, et al. 4π Non-Coplanar Liver SBRT: A Novel Delivery Technique. *International Journal of Radiation Oncology*Biology*Physics.* 2013;85(5):1360-1366.

32. Dong P, Lee P, Ruan D, et al. 4π Noncoplanar Stereotactic Body Radiation Therapy for Centrally Located or Larger Lung Tumors. *International Journal of Radiation Oncology*Biology*Physics.* 2013;86(3):407-413.

33. Dong P, Nguyen D, Ruan D, et al. Feasibility of prostate robotic radiotherapy on conventional C-arm linacs. *Practical Radiation Oncology.* 2013;In press.

34. Rossi L, Breedveld S, Heijmen BJ, Voet PW, Lanconelli N, Aluwini S. On the beam direction search space in computerized non-coplanar beam angle optimization for IMRT-prostate SBRT [published online ahead of print 2012/08/07]. *Phys Med Biol.* 2012;57(17):5441-5458.

35. Kairn T, Crowe SB, Kenny J, Knight RT, Trapp JV. Predicting the likelihood of QA failure using treatment plan accuracy metrics. *Xvii International Conference on the Use of Computers in Radiation Therapy (Iccr 2013).* 2014;489.

36. Mohan R, Arnfield M, Tong S, Wu Q, Siebers J. The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy. *Medical Physics.* 2000;27(6):1226-1237.

37. Keller-Reichenbecher MA, Bortfeld T, Levegrun S, Stein J, Preiser K, Schlegel W. Intensity modulation with the "step and shoot" technique using a commercial MLC: A planning study. *Int J Radiat Oncol.* 1999;45(5):1315-1324.

38. Evans PM, Hansen VN, Swindell W. The optimum intensities for multiple static multileaf collimator field compensation. *Medical physics.* 1997;24(7):1147-1156.

39. Wu Y, Yan D, Sharpe MB, Miller B, Wong JW. Implementing multiple static field delivery for intensity modulated beams [published online ahead of print 2002/01/05]. *Med Phys.* 2001;28(11):2188-2197.

40. Luan S, Wang C, Chen DZ, et al. A new MLC segmentation algorithm/software for step-and-shoot IMRT delivery [published online ahead of print 2004/05/06]. *Med Phys.* 2004;31(4):695-707.

41. Spirou SV, Fournier-Bidoz N, Yang J, Chui CS, Ling CC. Smoothing intensity-modulated beam profiles to improve the efficiency of delivery [published online ahead of print 2001/11/07]. *Med Phys.* 2001;28(10):2105-2112.

42. Matuszak MM, Larsen EW, Fraass BA. Reduction of IMRT beam complexity through the use of beam modulation penalties in the objective function. *Medical Physics.* 2007;34(2):507-520.

43. Jin RC, Min ZF, Song EM, Liu H, Ye YY. A novel fluence map optimization model incorporating leaf sequencing constraints. *Physics in Medicine and Biology.* 2010;55(4):1243-1264.

44. Siebers JV, Lauterbach M, Keall PJ, Mohan R. Incorporating multi-leaf collimator leaf sequencing into iterative IMRT optimization. *Medical Physics.* 2002;29(6):952-959.

45. Broderick M, Leech M, Coffey M. Direct aperture optimization as a means of reducing the complexity of Intensity Modulated Radiation Therapy plans [published online ahead of print 2009/02/18]. *Radiat Oncol.* 2009;4:8.

46. Men C, Romeijn HE, Taskin ZC, Dempsey JF. An exact approach to direct aperture optimization in IMRT treatment planning [published online ahead of print 2007/12/11]. *Phys Med Biol.* 2007;52(24):7333-7352.

47. Webb S. Direct aperture optimization for a variable aperture collimator for intensity-modulated radiation therapy [published online ahead of print 2004/04/09]. *Phys Med Biol.* 2004;49(5):N47-55.

48. Shepard DM, Earl MA, Li XA, Naqvi S, Yu C. Direct aperture optimization: a turnkey solution for step-and-shoot IMRT [published online ahead of print 2002/07/04]. *Med Phys.* 2002;29(6):1007-1018.

49. Suss P, Kufer KH, Thieke C. Improved stratification algorithms for step-and-shoot MLC delivery in intensity-modulated radiation therapy. *Physics in Medicine and Biology.* 2007;52(19):6039-6051.

50. Alber M, Nusslin F. Intensity modulated photon beams subject to a minimal surface smoothing constraint [published online ahead of print 2000/06/08]. *Phys Med Biol.* 2000;45(5):N49-52.

51. Matuszak MM, Larsen EW, Jee KW, McShan DL, Fraass BA. Adaptive diffusion smoothing: a diffusion-based method to reduce IMRT field complexity [published online ahead of print 2008/05/22]. *Med Phys.* 2008;35(4):1532-1546.

52. Boyd S, Vandenberghe L. *Convex optimization.* Cambridge university press; 2009.

53. Wright SJ. *Primal-dual interior-point methods.* Vol 54: Siam; 1997.

54. Ben-Tal A, Nemirovski A. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* Vol 2: Siam; 2001.

55. Boyd S. Alternating direction method of multipliers. Paper presented at: Talk at NIPS Workshop on Optimization and Machine Learning2011.

56. Nguyen D, Rwigema JC, Yu VY, et al. Feasibility of extreme dose escalation for glioblastoma multiforme using 4 inverted question mark radiotherapy [published online ahead of print 2014/11/08]. *Radiat Oncol.* 2014;9(1):239.

57. Rwigema J, Nguyen D, Heron D, et al. 4pi Noncoplanar Stereotactic Body Radiation Therapy for Head and Neck Cancers - Potential to Improve Tumor Control and Late Toxicity. *Int J Radiat Oncol Biol Phys.* 2014;In Press.

58. Grégoire V, Mackie TR. State of the art on dose prescription, reporting and recording in Intensity-Modulated Radiation Therapy (ICRU report No. 83). *Cancer/Radiothérapie.* 2011;15(6–7):555-559.

59. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Medical Physics.* 1998;25(5):656-661.

60. Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Medical Physics.* 2003;30(9):2455-2464.

61. Sharpe MB, Miller BM, Yan D, Wong JW. Monitor unit settings for intensity modulated beams delivered using a step-and-shoot approach. *Medical Physics.* 2000;27(12):2719-2725.

62. Kung JH, Chen GTY. Intensity modulated radiotherapy dose delivery error from radiation field offset inaccuracy. *Medical Physics.* 2000;27(7):1617-1622.

63. Eklund K, Ahnesjo A. Modeling silicon diode dose response factors for small photon fields [published online ahead of print 2010/11/26]. *Phys Med Biol.* 2010;55(24):7411-7423.

64. Cranmer-Sargison G, Liu PZ, Weston S, Suchowerska N, Thwaites DI. Small field dosimetric characterization of a new 160-leaf MLC [published online ahead of print 2013/10/01]. *Phys Med Biol.* 2013;58(20):7343-7354.

65. Luan S, Wang C, Chen DZ, et al. An improved MLC segmentation algorithm and software for step-and-shoot IMRT delivery without tongue-and-groove error [published online ahead of print 2006/06/07]. *Med Phys.* 2006;33(5):1199-1212.

66. Luan S, Saia J, Young M. Approximation algorithms for minimizing segments in radiation therapy. *Inform Process Lett.* 2007;101(6):239-244.

67. van Asselen B, Schwarz M, van Vliet-Vroegindeweij C, Lebesque JV, Mijnheer BJ, Damen EM. Intensity-modulated radiotherapy of breast cancer using direct aperture optimization [published online ahead of print 2006/05/23]. *Radiother Oncol.* 2006;79(2):162-169.

68. Nguyen D, Thomas D, Cao M, O'Connor D, Lamb J, Sheng K. Computerized triplet beam orientation optimization for MRI-guided Co-60 radiotherapy. *Medical Physics.* 2016;43(10):5667-5675.

69. Yang Y, Cao M, Sheng K, et al. Longitudinal Diffusion MRI for Treatment Response Assessment: Preliminary Experience using an MRI-Guided tri-Cobalt 60 Radiotherapy System *Medical Physics.* 2016;43(3).

70. Mutic S, Dempsey JF. The ViewRay System: Magnetic Resonance–Guided and Controlled Radiotherapy. *Seminars in Radiation Oncology.* 2014;24(3):196-199.

71. Dong P, Lee P, Ruan D, et al. 4pi noncoplanar stereotactic body radiation therapy for centrally located or larger lung tumors [published online ahead of print 2013/03/26]. *Int J Radiat Oncol Biol Phys.* 2013;86(3):407-413.

72. Dong P, Lee P, Ruan D, et al. 4pi non-coplanar liver SBRT: a novel delivery technique. *International journal of radiation oncology, biology, physics.* 2013;85(5):1360-1366.

73. Dong P, Nguyen D, Ruan D, et al. Feasibility of prostate robotic radiation therapy on conventional C-arm linacs [published online ahead of print 2014/07/12]. *Pract Radiat Oncol.* 2014;4(4):254-260.

74. Nguyen D, Rwigema JC, Yu VY, et al. Feasibility of extreme dose escalation for glioblastoma multiforme using 4pi radiotherapy. *Radiation oncology.* 2014;9:239.

75. Rockafellar RT. *Convex analysis.* Princeton university press; 2015.

76. Romeijn HE, Ahuja RK, Dempsey JF, Kumar A. A column generation approach to radiation therapy treatment planning using aperture modulation. *Siam J Optimiz.* 2005;15(3):838-862.

77. Nguyen D, O'Connor D, Yu VY, et al. Dose domain regularization of MLC leaf patterns for highly complex IMRT plans [published online ahead of print 2015/04/04]. *Med Phys.* 2015;42(4):1858-1870.

78. Keller-Reichenbecher M-A, Bortfeld T, Levegrün S, Stein J, Preiser K, Schlegel W. Intensity modulation with the “step and shoot” technique using a commercial MLC: A planning study. *International Journal of Radiation Oncology* Biology* Physics.* 1999;45(5):1315-1324.

79. Wu Y, Yan D, Sharpe M, Miller B, Wong J. Implementing multiple static field delivery for intensity modulated beams. *Medical physics.* 2001;28(11):2188-2197.

80. Luan S, Wang C, Chen DZ, et al. A new MLC segmentation algorithm/software for step-and-shoot IMRT delivery. *Medical physics.* 2004;31(4):695-707.

81. Josep Sempau and Scott JWaAFB. DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations. *Physics in Medicine and Biology.* 2000;45(8):2263.

82. Acharya S, Fischer-Valuck BW, Kashani R, et al. Online Magnetic Resonance Image Guided Adaptive Radiation Therapy: First Clinical Applications [published online ahead of print 2015/12/19]. *Int J Radiat Oncol Biol Phys.* 2016;94(2):394-403.

83. Hu Y, Rankine L, Green OL, et al. Characterization of the onboard imaging unit for the first clinical magnetic resonance image guided radiation therapy system [published online ahead of print 2015/10/03]. *Med Phys.* 2015;42(10):5828-5837.

84. Noel CE, Parikh PJ, Spencer CR, et al. Comparison of onboard low-field magnetic resonance imaging versus onboard computed tomography for anatomy visualization in radiotherapy [published online ahead of print 2015/07/25]. *Acta Oncol.* 2015;54(9):1474-1482.

85. Wooten HO, Green O, Yang M, et al. Quality of Intensity Modulated Radiation Therapy Treatment Plans Using a (6)(0)Co Magnetic Resonance Image Guidance Radiation Therapy System [published online ahead of print 2015/06/25]. *Int J Radiat Oncol Biol Phys.* 2015;92(4):771-778.

86. Sun B, Yang D, Esthappan J, et al. Three-dimensional dose accumulation in pseudo-split-field IMRT and brachytherapy for locally advanced cervical cancer [published online ahead of print 2015/05/11]. *Brachytherapy.* 2015;14(4):481-489.

87. Wooten HO, Rodriguez V, Green O, et al. Benchmark IMRT evaluation of a Co-60 MRI-guided radiation therapy system [published online ahead of print 2015/03/10]. *Radiother Oncol.* 2015;114(3):402-405.

88. Li HH, Rodriguez VL, Green OL, et al. Patient-specific quality assurance for the delivery of (60)Co intensity modulated radiation therapy subject to a 0.35-T lateral magnetic field [published online ahead of print 2014/12/03]. *Int J Radiat Oncol Biol Phys.* 2015;91(1):65-72.

89. Saenz DL, Yan Y, Christensen N, et al. Characterization of a 0.35T MR system for phantom image quality stability and in vivo assessment of motion quantification [published online ahead of print 2015/12/25]. *J Appl Clin Med Phys.* 2015;16(6):5353.

90. Wojcieszynski AP, Rosenberg SA, Brower JV, et al. Gadoxetate for direct tumor therapy and tracking with real-time MRI-guided stereotactic body radiation therapy of the liver [published online ahead of print 2015/12/03]. *Radiother Oncol.* 2015.

91. Saenz DL, Paliwal BR, Bayouth JE. A dose homogeneity and conformity evaluation between ViewRay and pinnacle-based linear accelerator IMRT treatment plans [published online ahead of print 2014/05/30]. *J Med Phys.* 2014;39(2):64-70.

92. Dong P, Nguyen D, Ruan D, et al. Feasibility of prostate robotic radiation therapy on conventional C-arm linacs. *Practical radiation oncology.* 2014;4(4):254-260.

93. Dong P, Yu V, Nguyen D, et al. Feasibility of using intermediate x-ray energies for highly conformal extracranial radiotherapy. *Medical physics.* 2014;41(4):041709.

94. Nguyen D, Rwigema J-CM, Yu VY, et al. Feasibility of extreme dose escalation for glioblastoma multiforme using 4π radiotherapy. *Radiation Oncology.* 2014;9(1):1-9.

95. Rwigema J-CM, Nguyen D, Heron DE, et al. 4π Noncoplanar Stereotactic Body Radiation Therapy for Head-and-Neck Cancer: Potential to Improve Tumor Control and Late Toxicity. *International Journal of Radiation Oncology*Biology*Physics.* 2015;91(2):401-409.

96. Leksell L. The stereotaxic method and radiosurgery of the brain [published online ahead of print 1951/12/13]. *Acta Chir Scand.* 1951;102(4):316-319.

97. Yu CX, Shao X, Zhang J, et al. GammaPod-a new device dedicated for stereotactic radiotherapy of breast cancer [published online ahead of print 2013/05/03]. *Med Phys.* 2013;40(5):051703.

98. Nguyen D, Ruan D, O'Connor D, et al. A novel software and conceptual design of the hardware platform for intensity modulated radiation therapy. *Medical Physics.* 2016;43(2):917-929.

99. Chang SX, Cullip TJ, Deschesne KM, Miller EP, Rosenman JG. Compensators: an alternative IMRT delivery technique [published online ahead of print 2005/03/09]. *J Appl Clin Med Phys.* 2004;5(3):15-36.

100. Tajima Y, Nakayama H, Itonaga T, et al. Dosimetric Evaluation of Compensator Intensity Modulation-based Stereotactic Body Radiotherapy for Stage I Non-Small Cell Lung Cancer [published online ahead of print 2015/05/23]. *Br J Radiol.* 2015:20150122.

101. Waghorn BJ, Staton RJ, Rineer JM, Meeks SL, Langen K. A comparison of the dosimetric effects of intrafraction motion on step-and-shoot, compensator, and helical tomotherapy-based IMRT [published online ahead of print 2013/05/09]. *J Appl Clin Med Phys.* 2013;14(3):4210.

102. Robinson J, Opp D, Zhang G, et al. Evaluating dosimetric accuracy of flattening filter free compensator-based IMRT: measurements with diode arrays [published online ahead of print 2012/01/10]. *Med Phys.* 2012;39(1):342-352.

103. Nangia S, Chufal KS, Arivazhagan V, Srinivas P, Tyagi A, Ghosh D. Compensator-based intensity-modulated radiotherapy in head and neck cancer: our experience in achieving dosimetric parameters and their clinical correlation [published online ahead of print 2006/08/17]. *Clin Oncol (R Coll Radiol).* 2006;18(6):485-492.

104. Mu G, Xia P. A feasibility study of using conventional jaws to deliver complex IMRT plans for head and neck cancer [published online ahead of print 2009/09/03]. *Phys Med Biol.* 2009;54(18):5613-5623.

105. Earl MA, Afghan MK, Yu CX, Jiang Z, Shepard DM. Jaws-only IMRT using direct aperture optimization [published online ahead of print 2007/02/07]. *Med Phys.* 2007;34(1):307-314.

106. Webb S. Intensity-modulated radiation therapy using only jaws and a mask [published online ahead of print 2002/02/12]. *Phys Med Biol.* 2002;47(2):257-275.

107. Dai JR, Hu YM. Intensity-modulation radiotherapy using independent collimators: An algorithm study. *Medical Physics.* 1999;26(12):2562-2570.

108. Webb S, Poludniowski G. Intensity-modulated radiation therapy (IMRT) by a dynamic-jaws-only (DJO) technique in rotate-translate mode [published online ahead of print 2010/10/13]. *Phys Med Biol.* 2010;55(21):N495-506.

109. Kim Y, Verhey LJ, Xia P. A feasibility study of using conventional jaws to deliver IMRT plans in the treatment of prostate cancer. *Physics in Medicine and Biology.* 2007;52(8):2147-2156.

110. Romeijn HE, Ahuja RK, Dempsey JF, Kumar A, Li JG. A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Physics in Medicine and Biology.* 2003;48(21):3521-3542.

111. Kirk J. *Open traveling salesman problem–genetic algorithm*. http://www.mathworks.com/matlabcentral/fileexchange/21196-open-traveling-salesman-problem-genetic-algorithm.

112. Dong P, Yu V, Nguyen D, et al. Feasibility of using intermediate x-ray energies for highly conformal extracranial radiotherapy [published online ahead of print 2014/04/04]. *Med Phys.* 2014;41(4):041709.

113. Nguyen D, Lyu Q, Ruan D, O'Connor D, Low DA, Sheng K. A comprehensive formulation for volumetric modulated arc therapy planning. *Medical Physics.* 2016;43(7):4263-4272.

114. Verbakel WFAR, Cuijpers JP, Hoffmans D, Bieker M, Slotman BJ, Senan S. Volumetric Intensity-Modulated Arc Therapy Vs. Conventional Imrt in Head-and-Neck Cancer: A Comparative Planning and Dosimetric Study. *Int J Radiat Oncol.* 2009;74(1):252-259.

115. Kawashima M, Ozawa S, Haga A, et al. Comparison of total MU and segment areas in VMAT and step-and-shoot IMRT plans [published online ahead of print 2012/07/04]. *Radiol Phys Technol.* 2013;6(1):14-20.

116. Rao M, Yang W, Chen F, et al. Comparison of Elekta VMAT with helical tomotherapy and fixed field IMRT: plan quality, delivery efficiency and accuracy [published online ahead of print 2010/04/14]. *Med Phys.* 2010;37(3):1350-1359.

117. Bedford JL. Treatment planning for volumetric modulated arc therapy [published online ahead of print 2009/12/10]. *Med Phys.* 2009;36(11):5128-5138.

118. Yu CX. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Physics in Medicine and Biology.* 1995;40(9):1435.

119. Shepard DM, Cao D, Afghan MKN, Earl MA. An arc-sequencing algorithm for intensity modulated arc therapy. *Medical Physics.* 2007;34(2):464-470.

120. Jin J-Y, Wen N, Ren L, Glide-Hurst C, Chetty IJ. Advances in treatment techniques: arc-based and other intensity modulated therapies. *The Cancer Journal.* 2011;17(3):166-176.

121. Crooks SM, Wu X, Takita C, Watzich M, Xing L. Aperture modulated arc therapy [published online ahead of print 2003/06/19]. *Phys Med Biol.* 2003;48(10):1333-1344.

122. Earl MA, Shepard DM, Naqvi S, Li XA, Yu CX. Inverse planning for intensity-modulated arc therapy using direct aperture optimization [published online ahead of print 2003/05/14]. *Phys Med Biol.* 2003;48(8):1075-1089.

123. Cao D, Afghan MK, Ye J, Chen F, Shepard DM. A generalized inverse planning tool for volumetric-modulated arc therapy [published online ahead of print 2009/10/21]. *Phys Med Biol.* 2009;54(21):6725-6738.

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

125. Bortfeld T, Webb S. Single-Arc IMRT? [published online ahead of print 2008/12/17]. *Phys Med Biol.* 2009;54(1):N9-20.

126. Craft D, McQuaid D, Wala J, Chen W, Salari E, Bortfeld T. Multicriteria VMAT optimization [published online ahead of print 2012/02/11]. *Med Phys.* 2012;39(2):686-696.

127. Xia P, Hwang AB, Verhey LJ. A leaf sequencing algorithm to enlarge treatment field length in IMRT [published online ahead of print 2002/07/04]. *Med Phys.* 2002;29(6):991-998.

128. Que W. Comparison of algorithms for multileaf collimator field segmentation [published online ahead of print 1999/12/10]. *Med Phys.* 1999;26(11):2390-2396.

129. Peng F, Jia X, Gu X, Epelman MA, Romeijn HE, Jiang SB. A new column-generation-based algorithm for VMAT treatment plan optimization [published online ahead of print 2012/06/23]. *Phys Med Biol.* 2012;57(14):4569-4588.

130. Cheng L-T, Dong B, Men C, Jia X, Jiang S. Binary level-set shape optimization model and algorithm for volumetric modulated arc therapy in radiotherapy treatment. *SIAM Journal on Scientific Computing.* 2013;35(6):B1321-B1340.

131. Osher S, Fedkiw R. *Level set methods and dynamic implicit surfaces.* Vol 153: Springer Science & Business Media; 2006.

132. Zhu L, Lee L, Ma Y, Ye Y, Mazzeo R, Xing L. Using total-variation regularization for intensity modulated radiation therapy inverse planning with field-specific numbers of segments. *Physics in medicine and biology.* 2008;53(23):6653.

133. Gorski J, Pfeuffer F, Klamroth K. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research.* 2007;66(3):373-407.

134. Neylon J, Sheng K, Yu V, et al. A nonvoxel-based dose convolution/superposition algorithm optimized for scalable GPU architectures [published online ahead of print 2014/10/06]. *Med Phys.* 2014;41(10):101711.

135. Pock T, Chambolle A. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. Paper presented at: Computer Vision (ICCV), 2011 IEEE International Conference on2011.

136. Condat L. A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms. *Journal of Optimization Theory and Applications.* 2013;158(2):460-479.

137. Feuvret L, Noël G, Mazeron J-J, Bey P. Conformity index: a review. *International Journal of Radiation Oncology* Biology* Physics.* 2006;64(2):333-342.

138. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning.* 2011;3(1):1-122.

139. Parikh N, Boyd S. Proximal algorithms. *Foundations and Trends in Optimization.* 2013;1(3):123-231.

140. Golub GH, Van Loan CF. Matrix Computations Johns Hopkins University Press. In: Baltimore, London; 1989.