Geometry-based Constraint Generation for Large-scale Radiation Therapy Treatment Planning

by

Kourosh Khedriliraviasl

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

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisor(s): Houra Mahmoudzadeh
Assistant Professor, Department of Management Sciences,
Faculty of Engineering, University of Waterloo

Other Member(s): Qi-Ming He
Professor, Department of Management Sciences,
Faculty of Engineering, University of Waterloo

Jim Bookbinder
Professor, Department of Management Sciences,
Faculty of Engineering, University of Waterloo

Author's Declaration

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

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

Abstract

Intensity-Modulated Radiation Therapy (IMRT) is a high precision radiotherapy with many beams that have different intensities to accurately irradiate a tumor/cancerous cells considering minimization of the dose to surrounding normal tissues. Different optimization methods are developed to achieve optimal beam intensities considering different criteria in objective function and number of constraints. IMRT optimization models are large-scale in nature because large number of constraints and/or variables associated with tumor/-cancerous cells can be defined which is very hard to solve and takes a huge amount of computational time. A novel constraint generation solution method for IMRT treatment planning is presented in this research. A Fluence Map Optimization (FMO) framework is used to formulate the optimization problem for a breast cancer treatment planning case study. In addition, a data processing method is developed to form clusters of similar voxels with the highest dose associated from beamlets. Because in this optimization problem there are large number of constraints related to each voxel, these data clusters will help in deciding which voxels to pick as set of new constraints for the sub-problems. This is for the purpose of reducing the number of constraints. The optimal beam intensities do not cause a violation of any other constraints in the main model. The results show that the objective function value from constraint generation model is same as the objective function value that is resulted from the full-size model implementation considering all the constraints. Also, the optimal beam intensities from these two models are comparable and they create similar fluence map pattern. This shows that the new model is able to maintain the same quality as large-scale model. The advantages of this new method are, reducing the number of constraints that should be considered for the optimization significantly, and maintaining solution quality which can help create faster optimization algorithms.

Acknowledgements

I would like to thank my supervisor and the committee for investing time to read my thesis and participate in my thesis seminar. I would like to thank students and friends who supported me with their comments.

Table of Contents

List of Figures

List of Tables

Abbreviations

3D-CRT three-dimensional conformal radiation therapy 1, 2

3D-RTP three-dimensional radiation treatment planning 1, 2

CCRT computer controlled radiation treatment 1

CTV Critical Target Volume 15–17, 24, 25, 28, 33–35, 48

FMO Fluence Map Optimization 5–10, 12, 13, 24, 35, 47

IMRT Intensity Modulated Radiation Therapy 2–4, 6, 10, 12, 13, 15, 47, 48

MLC Multi-Leaf Collimator 3, 5

OAR organs at risk 7, 8, 15, 28, 48

RT Radiation Therapy 1, 3

Nomenclature

D_{ij} A dose Matrix from set of all weights of the beamlet 5

$D^H_{i,j}$ data in dose-influence matrix for Heart 24

$D^T_{i,j}$ data in dose-influence matrix for CTV 24

H set of voxels in Heart 24

T set of voxels in CTV 24

$V^{(k)}$ subset of voxels at iteration k 25

$V^{(k+1)}$ new subset of voxels at iteration $k+1$ 26

δ violation amount in Gy 29

$\mathcal{V}^b$ set of all voxels in structure b 8

$\mathcal{V}_s$ number of voxels in structure s 7

$\overline{D}^b$ upper bound on dose to each voxel in structure b 8

$\overline{T}_s$ above threshold of penalized dose in structure s 7

$\overline{w}_s$ weight on overdose for structure s 7

θ_i lower bound for total CTV dose 24

$\underline{D}^b$ lower bound on dose to each voxel in structure b 8

$\underline{T}_s$ below threshold of penalized dose in structure s 7

$\underline{w}_s$ weight on under-dose for structure s 7

d_{js} dose in structure s 7

$viol_i^{(k)}$ violation for each voxel i at iteration k 26

$w^{*(k)}$ optimal beam intensity at iteration k 26

w_j intensity of beamlet j 3

Chapter 1

Introduction

1.1 Intensity Modulated Radiation Therapy (IMRT)

In 2019, 220,400 Canadians are estimated to be diagnosed with cancer with 82,100 mortality rate, and 63% survival rate is estimated within first 5 years of treatment. Also, Breast cancer is the second most commonly diagnosed cancer in 2019 with 1 in 8 females. Radiation therapy is one of the cancer treatment ways that can destroy cancer cells over a repetitive treatment period without causing too much harm to healthy cells in other organs (CCSAC, 2019). Use of Radiation Therapy (RT) treatment planning is growing with the utilization of cutting edge three-dimensional radiation treatment planning (3D-RTP) and computer controlled radiation treatment (CCRT) frameworks (Gupta et al., 2012). As a consequence, three-dimensional conformal radiation therapy (3D-CRT) is intended to adjust the spatial distribution of the prescribed dose to the 3D target volume (cancerous cells) while minimizing the dose to the ordinary surrounding structures (Prado et al., 2007). Ordinarily, the 3D-CRT uses fixed radiation beams which are formed utilizing the target

volume projection (Hall and Wuu, 2003). A new 3D-CRT technique, Intensity Modulated Radiation Therapy (IMRT) (as shown in Figure 1.1), is based on the application of optimized non-uniform radiating beam intensities to the patient (Bortfeld, 2006). Automated 3D-RTP optimization is often used in the IMRT treatment plans to determine the distribution of intensities over the target volume using a computer optimization technique. An integral component of IMRT is the detection of optimal beam fluence. The main planning problem for IMRT is in determining the modulated fluence profiles which are physically deliverable and lead to a dose distribution that best matches the intended dose (Boyer et al., 2001). IMRT is a form of radiation therapy that enables the intensity of beam radiation to be modulated. It becomes necessary to calculate treatment plans that best achieve the treatment objectives. Also, it is of high importance to use optimization models, methods and theories in the design of intensity modulated radiation therapy (Ehrgott et al., 2010).

Figure 1.1: Intensity Modulated Radiation Therapy (IMRT) System (UHN, 2015)

1.2 Mathematical Programming Models for IMRT

There are a number of mathematical problems to manage IMRT optimally. Treatment is carried out by rotating the accelerator about the patient and by coordinating the motions of the leaf in a device known as Multi-Leaf Collimator (MLC) so as to achieve an appropriate dose dispensing at each angle. The identification of the numbers and values of the gantry angles is a higher-level optimization problem, often known as the optimization problem of the beam-angles. It is assumed that gantry angles are given in the proposed model. Fluence maps, are characterized by intensity of beamlet j, w_j. The identification process of these beamlet intensities is known as *fluence map optimization* (Zhang and Merritt, 2006).

Mathematical models are classified into following parts; (a) feasibility problems (b) linear programming models, (c) nonlinear programming models, (d) mixed integer programming models, and (e) multiple objective programming models. When formulating the optimization model, objective function and constraints should be specified and the optimization search algorithm needs to be defined. The clinical objectives in the form of an objective function (also known as the score function or cost function) for the IMRT planning are defined mathematically. In order to determine beam parameters (beam weight) most closely to reach the desirable solution, computer optimization techniques are developed. The objective function value is the supposed index of the treatment plan's goodness. This value is often described by the term score or cost. The objective of optimization is therefore to minimize (or maximize) the score according to the selection of the objective function. In some previous attempts, objective functions are used to optimize the RT plans based on the distribution of dose (Bortfeld et al., 1990) (Bortfeld and Schlegel, 1993).For instance, the minimum dose of the target volume could be maximised under a maximum dose con-

straint for certain organ at risk. Dose based and/or dose-volume-based criteria are used in most IMRT optimization systems. A common method for the creation of dose-based and dose-volume objective functions is to minimize the dose variance for target or dose limits of risk organ values in relation to the prescribed dose.The variance is the sum of the squares of the differences between the dose or dose limit calculated and prescribed. Therefore, the typical dose-based or dose-based objective function is to sum up the variance terms that each anatomical structure is represented, multiplied by corresponding penalty factors. This is known as quadratic objective function. Fundamentally, linear programming models objectives are in the forms of (a) minimize the maximum deviation from upper bounds on dose to critical organs and normal tissue, (b) maximize the minimum dose to tumor, and (c) minimize the maximum deviation from prescribed dose to target (Ehrgott et al., 2010).

The constraints can be formed into (a) beam intensity non-negativity constraints, (b) dose to the organs at risk and normal tissues upper-bound, (c) dose to the target lower and upper bounds, (d) ratio between the maximum beamlet intensity and the average beamlet intensity upper bound, and (e) mean dose to organs at risk upper-bound. Mixture of objectives and constraints can create many types of models. In general, when the objective function is based on target, critical organs and normal tissue will form the constraints. Additionally, when critical organs and normal tissue form the objective function, so the target would form the constraints. The non-negativity constraints should always exist to cover as a physical constraint. The next section is focused on IMRT fluence map optimization.

1.3 Fluence Map Optimization (FMO) for IMRT

The distribution of the radiation dose that was measured in Gray (Gy) in the patient should be evaluated correctly for optimization reasons. The volume of each structure is discretized by voxels according to taking into account of every beamlet contribution. A dose matrix D_{ij} is usually built from the set of all weights of the beamlet, via indexation of D_{ij} rows for every voxel and D_{ij} columns for every beamlet. In other words, the row numbers in D_{ij} matrix will be associated with voxel numbers (V) and the column numbers will be related to beamlet numbers (N) where all beam directions are taken into consideration . Accordingly, with this matrix formulation, the total dose that the voxel i receives can be determined by $\sum_{j=1}^{N} D_{ij}w_j$, where in this formulation w_j represents the intensity beamlet j, also $i \in \{\ 1, ..., V\ \}$ where V is number of voxels and $j \in \{\ 1, ..., N\ \}$ where N is number of beamlets. Additionally, the number of all voxels that are taken into consideration is in the range of tens of thousands, so the dimension of the dose-influence matrix is usually very high. The large-scale problems are associated from the big size of this dose-influence matrix D which is considered as the number one reason that solving the Fluence Map Optimization (FMO) problems is difficult. The majority of the FMO literature models belong in a class of constrained optimization models in order to optimize an objective function while satisfying the dose requirements. A range of criteria can be regarded as part of the objective function, which leads to various objective functions (Rocha et al., 2015).

In Figure 1.2, every element in D_{ij} can be related to a beamlet radiated from the MLC and a voxel in region of interest that were discretized.

Let D be the matrix with non-zero elements in rows and/or columns. Accordingly, every voxel receives a non-zero radiation quantity and each of the beamlets has some effects on

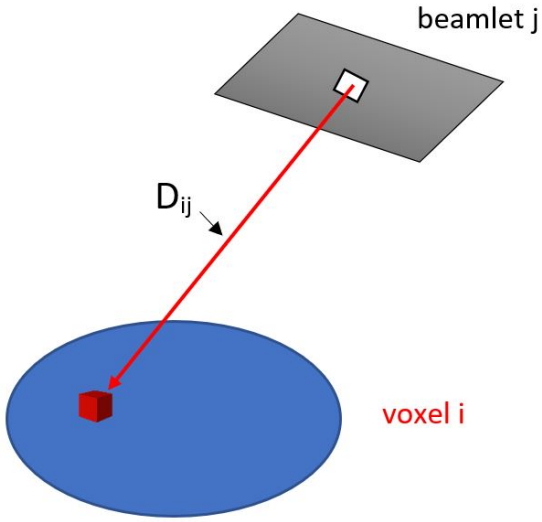


Figure 1.2: Influence Matrix Element

one dose of a voxel.

The problem of fluence map optimization is an instance of an IMRT planning that an acceptable dose distribution can be determined and a beamlet intensity vector can be identified which best realizes the given distribution. Multiple methods are presented in the literature in order to formulate the mentioned problem into optimization format that uses various objective functions consisting biological and physical ones. Different biological reactions are represented as statistical theories to build biological models; tumor control probability is an example of it. Recently, "weighted least squares" is being used as the main modelling technique which most commercial IMRT systems use. The objective function can represent every anatomical structure, so in this case the problem can be converted to a multi-objective optimization problem.

The objective function for the FMO model is sometimes formulated as follows. A convex penalty function to penalize the over- and underdose of each voxel is presented;

$$F(\mathbf{d}) = \sum_{s\in\mathcal{S}} \sum_{j\in\mathcal{V}_s} \frac{1}{|\mathcal{V}_s|} \left(\overline{w}_s \big(d_{js} - \overline{T}_s\big)_+^p + \underline{w}_s \big(\underline{T}_s - d_{js}\big)_+^p \right)$$

In this equation, $\mathcal{V}_s$ is the number of voxels in structure s, d_{js} is the dose in structure s associated from beamlet j, $(.)_+$ indicates $\max\{.,0\}$ and $\overline{T}_s$ and $\underline{T}_s$ are above and below thresholds of penalized dose in structure s and weights on over- and under- dose for structure s are represented by $\overline{w}_s$ and $\underline{w}_s$. In order to prevent large structures to dominate small structures, the penalty for each voxel is divided by the size of the structure. Quadratic versions of this penalty function (p=2) are common mostly (Aleman and Sharpe, 2011), (Romeijn et al., 2003). To improve the computational performance, linear versions (p=1) are taken into consideration (Lim et al., 2008), (Cao and Lim, 2011). Experiments show there is no need to include hard constraints for voxel doses due to the power of the quadratic convex penalty function, but tuning the threshold and weight parameters is not an easy task.

Additionally, in order to consider uncertainties in organ positions, robust formulations of FMO are presented. The objective is to minimize total dose to organs at risk (OAR) (Bortfeld et al., 2008), (Chan et al., 2006),

$$F(\mathbf{d}) = \sum_{H} \sum_{i\in\mathcal{V}} d_{ih}$$

or other version of it as weighted total dose to OARs (Mahmoudzadeh et al., 2015):

$$F(\mathbf{d}) = \sum_{H} \sum_{i\in\mathcal{V}} \frac{1}{|\mathcal{V}|} \underline{w}_h d_{ih}$$

OARs are the healthy organs surrounding the tumour that are at risk of being damaged

due to excessive radiations. As for the constraints, full−volume constraints represent that the dose in each structure voxel should be within predesignated upper and lower bounds (Romeijn et al., 2003) and is formulated as follows;

$$d_{ih} \geq \underline{D}^b \qquad \forall i \in \mathcal{V}^b$$
$$d_{ih} \leq \overline{D}^b \qquad \forall i \in \mathcal{V}^b$$

In this equation, lower and upper bounds on dose to each voxel in structure b are represented by $\underline{D}^b$ and $\overline{D}^b$, and $\mathcal{V}^b$ is the set of all voxels in structure b. Target dose homogeneity is guaranteed with these constraints. The constraints for full-volume upper bound can be identified in order for organs to be still functional after the treatment ends, so the maximum dose is considered for these OARs.

In the literature, FMO problems are presented in various formulations which consist of linear programming (LP)-based multi-criteria optimization (Craft et al., 2007), (Hamacher and Küfer, 1999) and mixed-integer linear programming (MILP) (Bednarz et al., 2002) , (Ferris et al., 2006b), (Shepard et al., 1999). Most of the FMO models are in the form of optimizing physical dose criteria and other methods focus on biological aspects to be analyzed (Alber and Nüsslin, 1999), (Das, 2009), (Zinchenko et al., 2008). Also, some advanced techniques for FMO modelling can be determined that have the advantage of being computationally tractable considering physical criteria. In this research the general formulation of FMO model is used but the proposed constraint generation algorithm is applicable to other model formats.

1.4 Solution Methods for IMRT

Optimal treatment design requires solving a very large-scale optimization model. In FMO, there are huge number of constraints and/or variables to consider for optimization and there is a need to develop an optimization solution method that can be accurate and at the same time it is potentially fast. In the following sections, different solution methods including sampling methods, heuristic and meta-heuristics and exact methods are discussed, and pros and cons of each are explained.

1.4.1 Sampling Methods

As it is discussed earlier, radiation treatment planning employs a range of optimization problems and methods. These problems usually include large quantities of information from patient anatomy simulations and delivery device characteristics. An approach of three phases has been examined, based on samples of the subordinate data, which in patient examples determine optimal beam angles, wedge directions and delivery intensities. Phase I utilizes several coarse information samples and linear programming to adapt the sample and establish a promising angle collection to use. Phase II resolves the sample problems adapted as mixed integer programs with promising angles only. Phase III further refines the sample and sets out the majority of discrete choice factors to decrease calculation times. General principles applicable to major classes of treatment planning problems have been given particular emphasis. Specific instances have shown huge speed increases without affecting the quality of the solution (Ferris et al., 2006a).

1.4.2 Heuristic and Meta-heuristics

Projected gradient techniques for linearly constrained FMO models with quadratic objectives are most often used for their speed, lack of reliance on the calculations from Hessian and easy execution (Lahanas et al., 2003). Projected gradient methods are sufficiently common to have a review of their IMRT optimization speed (Zhang et al., 2004).But, projected gradient techniques do not ensure optimal treatments. While mathematical optimality is of no relevance from the clinical point of view, failure to ensure solution quality implies that treatment plans may be incompatible and non standard, and may rely on the duration of the algorithm, not to mention algorithm-specific implementation by a specific dosimetrist. Different line-search strategies were discovered to produce dramatic variations in IMRT results within the projected gradient routine (with the quickest, but tailored, non-uniform retrieval technique producing best objective function values) (Aleman et al., 2013). In addition to traditional near-zero starting, it can be found that a warm start improves the final value of the objective function and the total calculation time significantly in comparison with the magnitude of the most extensive treatment sites.

Customized interior point methods are as quick as projected gradient approaches for linearly constrained FMO models also it has the advantage of guaranteed optimality gaps. It has also been shown that reformulating the quadratic FMOs to conical problems, or to new constraint generation techniques, with commercial software, lead to ϵ- optimal alternatives. Although these methods are empirically better than projected gradient methods and may be comparable or better in time for computation, in the real world application, practical difficulties may arise with commercial software or in obtaining approval by clinicians of complicated, possibly non-intuitive solutions to optimization (Lan et al., 2012).

If the change in the weight of the beamlet yields a better result, the suggested change is

approved and if it is not, the change is dismissed which depends on the method of search. Since each beam is improved from a variety of distinct directions and every beam has an impact on many points, only relatively small changes in the weight of the beam are allowed at once. For all beamlets, this method is repeated. A tiny enhancement of the treatment plan probably results at the end of each full cycle (an iteration). A new dose distribution and new score for the target function is then calculated by using the new pattern of beamlet intensities, which is then used as a basis to further enhance the next cycle. This iterative method continues until no further enhancement occurs (Deasy, 1997).

Gradient methods are considered to be computationally fastest, but a gradient method assumes that a single extremum (depending upon the type of the objective function, a minimum or maximum) is present. This is true of a quadratic objective (dose-based) function when only the weight of the beam is optimized. In other cases, it can be required to establish if there are multiple extremes, and if such multiple extremes affect the quality of the solutions found. Several extremes have been discovered in optimized beam paths and in the use to optimize standardized beam weights by dose-response objective functions. It can also be assumed that there are various minimum levels in order to optimize IMRT plans with dose-response objective functions. Multiple minimum levels have also been shown when using dose-volume-based objectives (Deasy, 1997). If there are several extremes, gradient search techniques converge to the closest extreme. The treatment plan for this extreme point may be far from the best possible alternative and completely unsatisfactory.

The most frequently used meta-heuristic method is simulated annealing, or its variation, fast simulated annealing (Webb, 1991). Other types have also been suggested, such as genetic algorithms (Ezzell, 1996). The simulated annealing method makes it possible to escape the optimization process from local extrema traps and thus discover the global extreme, but this is only true if a large amount of configurations is evaluated. There is prac-

tically no assurance that the absolute best solution will be discovered among those studied. In addition, meta-heuristics are fast but there is no guarantee or bound on optimality.

Simulated annealing and genetic algorithms are very slow to achieve the optimum, typically requiring tens of thousands of iterations. The gradient-based algorithms are fast compared to SA. Although they do not provide global optimality, they appear to provide adequate treatment plans. Therefore, when beam intensity optimization problem is considered, a gradient-based technique seems natural.

Iterative filtered back projection and fast iterative techniques are suggested, such as the image reconstruction from projections. Iterative algorithms begin with an initial solution that is approximated and generate a sequence of solutions that correspond to optimal solution. Being simple and straightforward are the advantages of these algorithms. Unlike iterative techniques, constrained least square (CLS) algorithms are used by the direct matrix manipulation to achieve the solution of a linear system.For conformal radiotherapy, CLS was applied and extended to IMRT. In addition, random search and maximum entropy and methods of maximum likelihood were also applied to solve the problem of optimizing the beam intensity.

An approach to FMO that is called *aperture modulation*, does not address individual beamlets but apertures explicitly. (Siebers et al., 2002) have proposed incorporating the option of aperture in a heuristic FMO local search technique. (Bednarz et al., 2002) developed a mixed-integer linear programming model, incorporating a predetermined pool of apertures to choose from. Also, an improved algorithm has been suggested in order to determine a treatment plan using only a few apertures (Shepard et al., 2002). Such methods have been implemented in instances of head-and-neck and prostate cancer, and empirical studies indicate that only a restricted number of apertures are able to achieve high quality treatment plans. All of these approaches, however, are heuristic in nature

and cannot ensure an optimal solution. Formal integration of FMO and decomposition problems is determined to be key in achieving the best apertures.

1.4.3 Exact Methods

Development of an IMRT treatment plan that determines an optimal set of apertures and their associated intensities is further investigated in (Romeijn et al., 2005). In the first phase, each beam of radiation shall be seen as a collection of beamlets, each with a specific intensity. In the second phase, the optimal intensity profile is approximated and decomposed into a set of apertures with corresponding intensities. A large-scale convex program is formulated for this problem, and because of its large dimensions, a column generation approach is discovered. In each iteration, one or more apertures can be added to the problem by the related optimization problem. Several versions of this sub problem are identified, each corresponding to a specific set of constraints that apertures need to meet in one or more kinds of commercial IMRT equipment presently accessible. Polynomial-time algorithms are considered to solve each of these sub problems in order to reach optimality. This approach proves to be clinically efficient.

Therefore, for specific problems, some algorithms are developed such as Direct Aperture Optimization (DAO) and column-generation, but there are no exact methods for solving FMO efficiently and accurately without random sampling.

1.5 Proposed Study

In this research, a geometry based data-driven optimization approach is presented for arranging clusters of dominant voxel-beamlet pairs to be used in a constraint generation

algorithm in order to improve the computational efficiency and memory shortages while running the optimization problem. A case study of breast cancer radiation therapy treatment planning optimization is studied. This constraint generation approach is compared to the original fluence map optimization problem which is clinically acceptable, and the result shows similar quality in terms of objective function values and optimal beamlet intensities. The results of different sensitivity analysis scenario considerations are also presented in a later chapter.

Chapter 2

Methodology

In this chapter, a data processing technique to create voxel-beamlet clusters along with some test scenarios are explained. Moreover, the constraint generation algorithm formulation and implementation are clarified.

2.1 Data Processing

A breast cancer radiation therapy treatment planning case-study is considered. As can be seen in Figure 2.1, the blue 3D scattered data represents the voxels of the Critical Target Volume (CTV) which in the IMRT system will receive radiation from two beams, beam 1 and beam 2, which are emitted from opposite sides and have the same 2D shape. Each of these beams consist of 760 beamlets. In this case-study there are 33,592 voxels in CTV, and the Heart is considered as an OAR with 42,342 voxels. Note that the size of a voxel is fixed (e.g., $1{\times}1 \times 1\text{mm}^3$), and hence, the number of voxels in a structure depends on the shape and volume of the structure. In this figure, each voxel requires a constraint in the

optimization model and that is why it is considered a large-scale optimization problem. This visual representation helps in identifying voxel locations and studying the individual effect of beamlets on the voxels and looking for similar patterns and/or shapes that different beamlets can produce. Also, as it is explained in next parts, these visualizations are helpful to develop insights for creating different clusters of voxels, which create highest dose profiles that can be considered in order to develop a constraint generation algorithm based on those clustered dose profiles, enabling solution of the model efficiently.

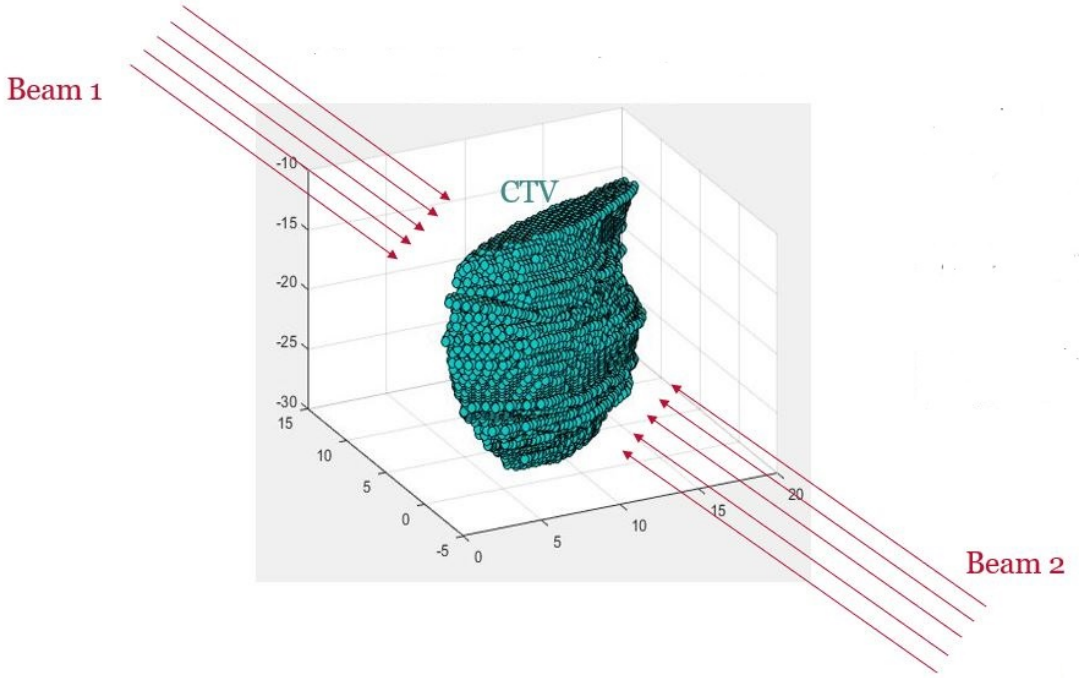


Figure 2.1: Visualization of voxels in CTV

In Figure 2.2, the effect of unit intensity of a sample beamlet (beamlet 744) on all voxels in the CTV is shown. In Figure 2.2, bright colors represent the high intensity and dark colors represent low intensity. This shows that according to the Dose Influence matrix for

this case-study, each beamlet has effects on certain number of voxels in different locations of the CTV. Two scenarios are studied in order to see whether any common voxel dose patterns can be found associated from neighboring beamlets in the x-y coordinate. The first scenario considered two beamlets along x-axis and the second scenario considered two beamlets along y-axis. Because the higher dose (i.e. bright area) is more important to be taken into consideration, another set of 3D plots are created in the next parts which only shows the voxels with the certain range of highest dose. This makes it easier to discover any potential common patterns and compare different scenarios.

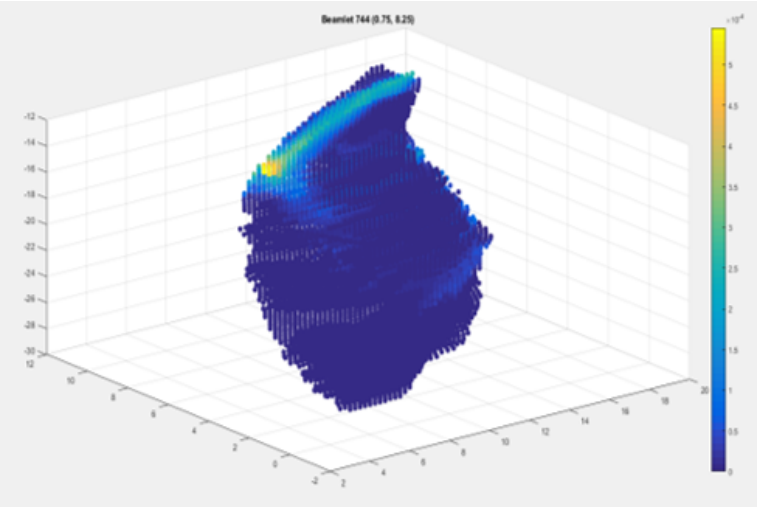


Figure 2.2: Visualization of affected voxels associated from beamlet 744 in CTV

2.1.1 Comparing Neighbouring Beamlets (First Scenario)

In Figure 2.3 the beamlet coordinates is plotted. This is for the purpose of visual comparison of dose for neighbouring beamlets which in this case is for beamlet 742 and beamlet 743.

Figure 2.3: beamlet coordinate (scenario 1)

As can be seen in Figure 2.4 visual comparison of dose for neighbouring beamlets is plotted . This is for all ranges of D_{ij} values.

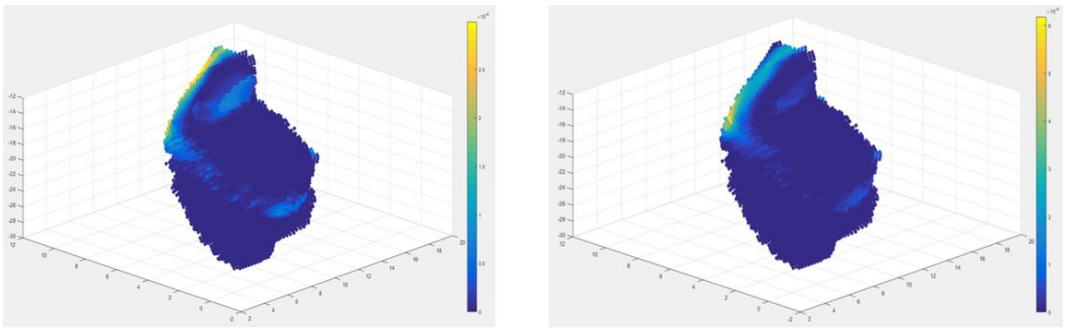


Figure 2.4: beamlets 742 and 743 comparison from beam 1

Figure 2.5 shows the visual comparison of the same neighbouring beamlets for a specific range of D_{ij} values greater than 2×10^{-4}.

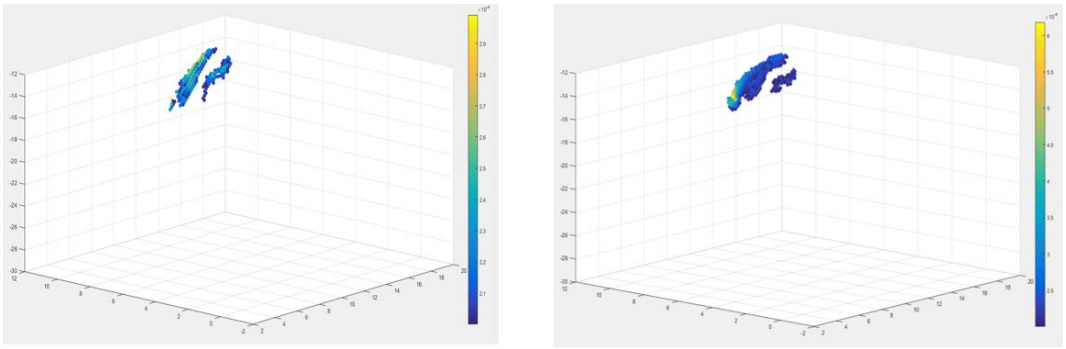


Figure 2.5: beamlets 742 and 743 comparison from beam 1 with $D_{ij} > 2 \times 10^{-4}$

As can be seen, neighbouring beamlets seem to affect different set of voxels.

2.1.2 Comparing Neighbouring Beamlets (Second Scenario)

As it is shown in Figure 2.6, another example is now considered for the visual comparison of neighbouring beamlets along y-axis. This differs from the previous example, which considered neighbouring beamlets in a row.

Beamlet 729 and beamlet 710 affected voxels visualizations as presented in Figure 2.7. These are presented for all the D_{ij} values.

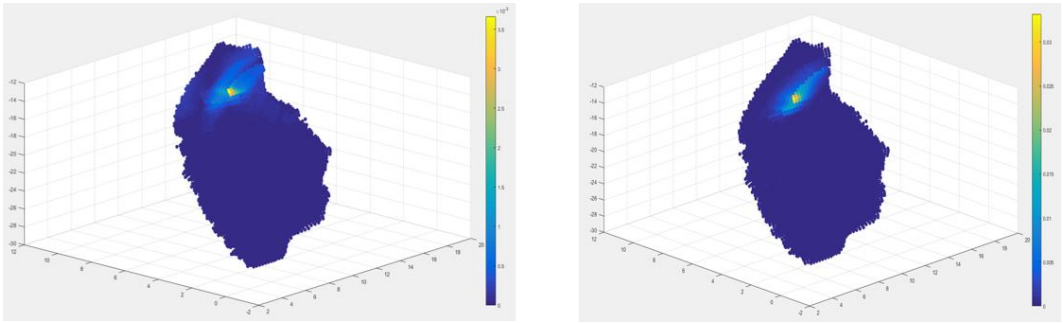


Figure 2.7: beamlets 729 and 710 comparison from beam 1 with all D_{ij}

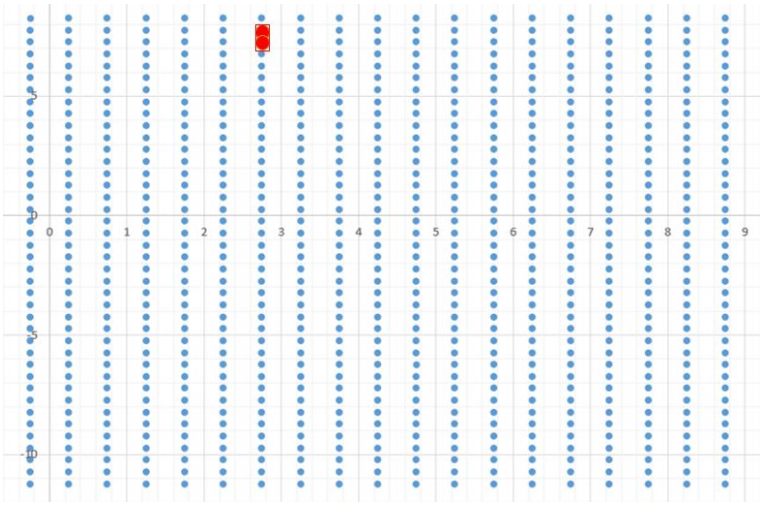


Figure 2.6: beamlet coordinate (scenario 2)

The Figure 2.8 compares the same beamlets, when D_{ij} has the same range as considered in the previous example.

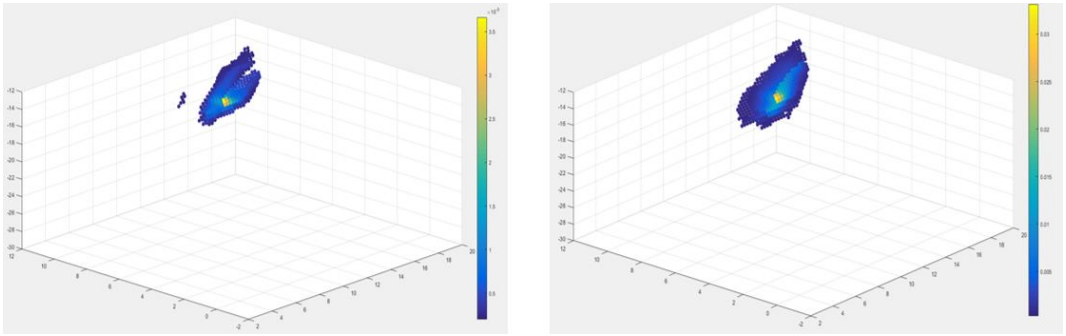


Figure 2.8: beamlets 729 and 710 comparison from beam 1 with $D_{ij} > 2 \times 10^{-4}$

As for the outcome of these observations, some repeated patterns can be seen but the shape varies in different cases. This is not helpful in terms of clustering the voxels, since the shapes are too different and voxels belong to several clusters.

2.1.3 Comparing Neighbouring Voxels

To improve upon the results so far, another set of examples has been tested which shows which beamlets does each voxel see. The visual comparisons are shown in Figure 2.9.

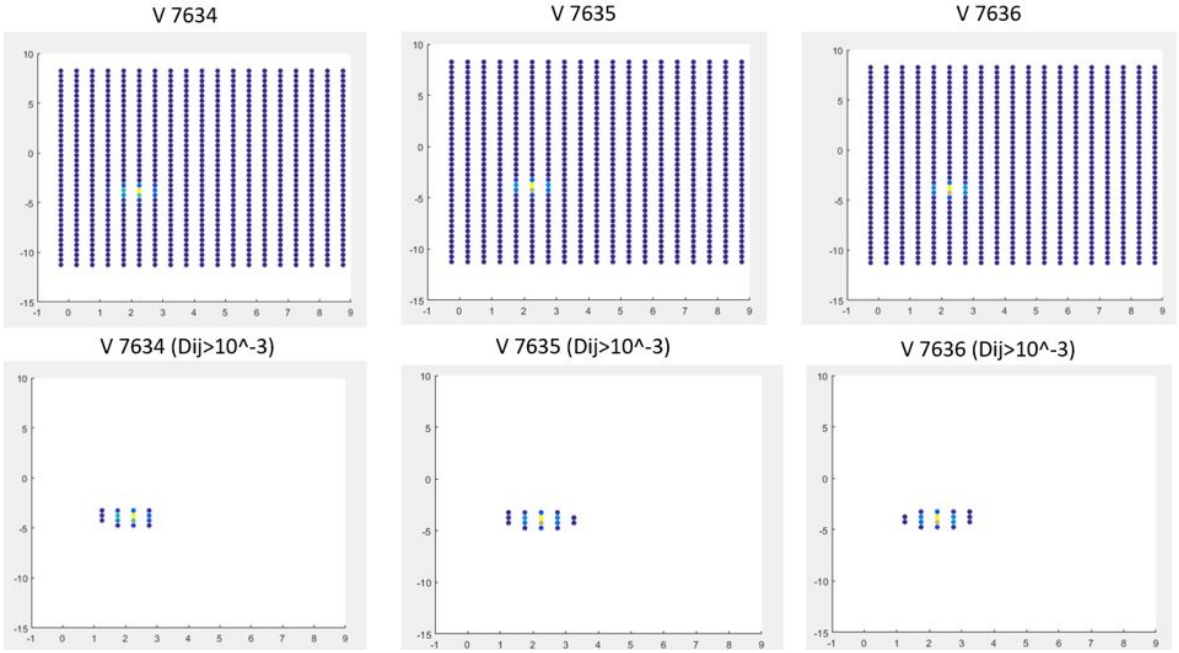
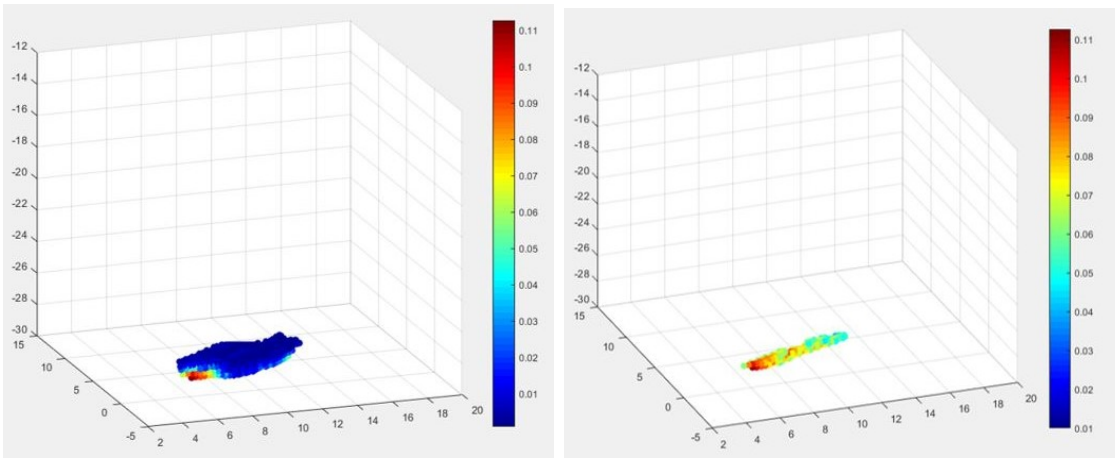


Figure 2.9: which beamlets does each voxel see

There are some differences but there are 1) repeated patterns, and 2) the main beamlet with the highest intensity was similar for a set of neighbouring voxels.

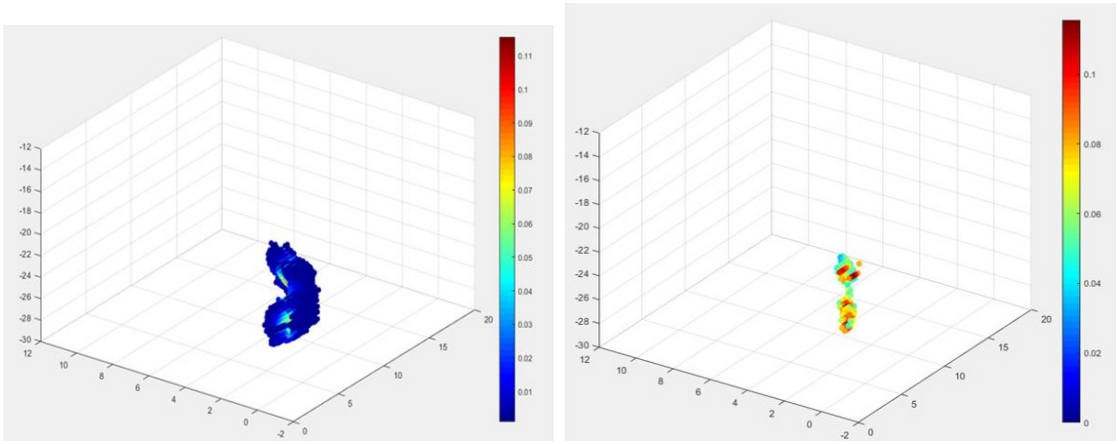
2.1.4 Voxels that Have Same Dominant Beamlet

To further investigate, voxels that have the same dominant beamlets are plotted in Figures 2.10, 2.11, 2.12, 2.13.



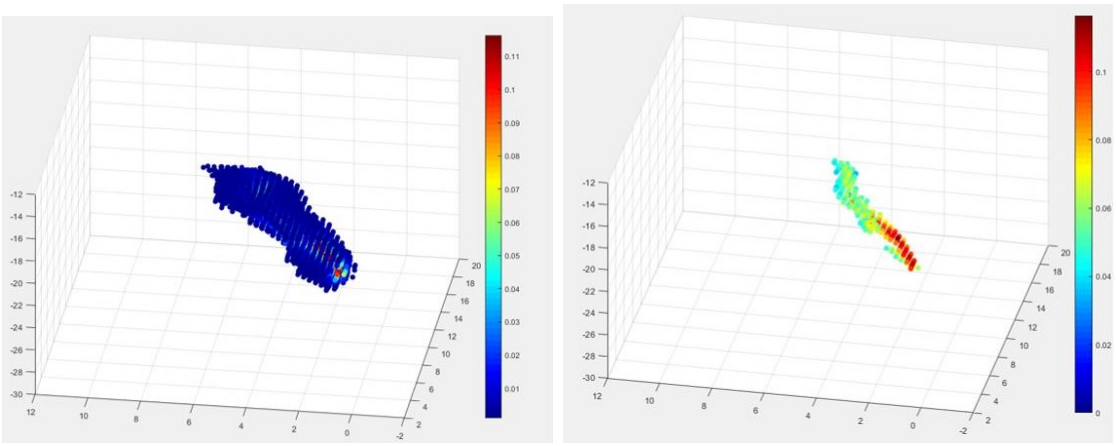
(a) (b)

Figure 2.10: (a) what beamlet 60 sees and (b) voxels that have beamlet 60 as the main dominant beamlet



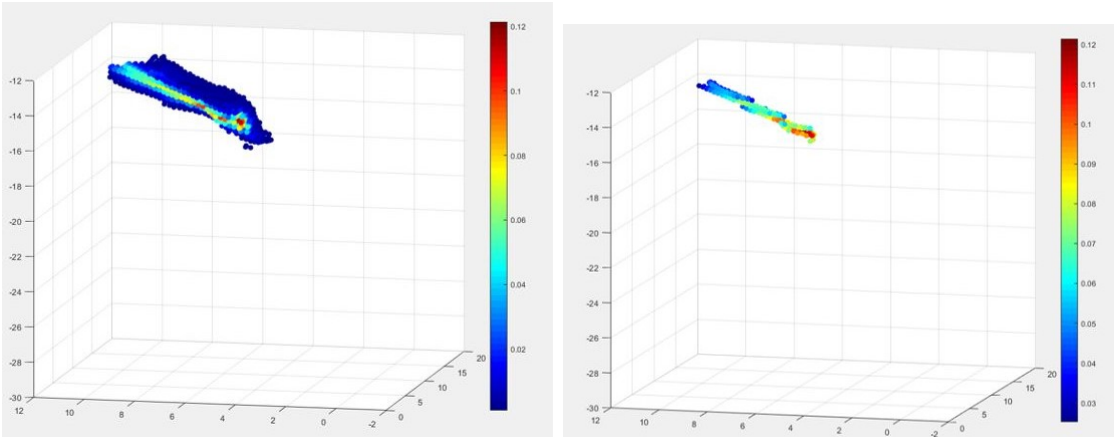
(a) (b)

Figure 2.11: (a) what beamlet 175 sees and (b) voxels that have beamlet 175 as the main dominant beamlet



(a) (b)

Figure 2.12: (a) what beamlet 406 sees and (b) voxels that have beamlet 406 as the main dominant beamlet



(a) (b)

Figure 2.13: (a) what beamlet 649 sees and (b) voxels that have beamlet 649 as the main dominant beamlet

For example, in Figure 2.13a, if beamlet 649 is on, then all voxels in this cluster will

get a high dose. Voxels in Figure 2.13b receive the highest possible dose from beamlet 649. That means these voxels are within the direct range of this beamlet, and hence, have beamlet 649 as their "dominant" beamlet. The advantages are, each voxel is in one cluster only and there are similar patterns across beamlets. All the 32,592 voxels are clustered based on their dominant associated beamlet and they are sorted based on their beamlet index. There are 302 clusters of voxels in total with different number of voxels, ranges from 1 to 300. These 302 voxel clusters associated with 302 beamlets and because they are representing the dominant beamlets, the rest of the beamlets are not considered in the model.

So, the dominant beamlet-voxel pairs are clustered differently and the data is processed in order to move on to the next stage. That is the FMO problem, considering the use of a novel constraint-generation algorithm to solve the linear program.

2.2 Optimization Algorithm

The original linear FMO model is as follows;

$$\begin{aligned} \text{Minimize} \quad & \sum_{i \in H} \sum_{j} D^{H}_{i,j} w_j \\ \text{Subject to} \quad & \sum_{j} D^{T}_{i,j} w_j \geq \theta_i, \forall i \in T \end{aligned} \tag{2.1}$$

where $D^{H}_{i,j}$ in objective function represents the data in dose-influence matrix for set of Heart (H) voxels, w_j represents the beam intensity which is the decision variable, and $D^{T}_{i,j}$ in constraint formulation is the data in dose-influence matrix for set of T voxels in CTV. Also, θ_i is a constant number, in this case 42.4 Gy, which represents a lower bound for

amount of dose in CTV. This problem concerns minimizing the total dose to the Heart, while meeting the minimum required dose for every voxel on the CTV.

For the constraint generation algorithm initialization, the large number of constraints in model (2.1) is relaxed and the beamlet-voxel clustered data that resulted in the previous section is used to construct a set of initial constraints. The initial optimization model solution and set of optimal beamlet intensities are generated. It is then checked whether there are any violations for the rest of the constraints. After that it is possible to iterate through with this process until there are no more voxels to be considered from the clusters and there are no more violation of constraints.

2.2.1 Master Problem

In the constraint generation algorithm, the master problem is formulated as follows;

$$\begin{aligned} \text{Minimize} \quad & \sum_{i \in H} \sum_{j} D^{H}_{i,j} w_j \\ \text{Subject to} \quad & \sum_{j} D^{T}_{i,j} w_j \geq \theta_i, \forall i \in V^{(k)} \end{aligned} \tag{2.2}$$

where $V^{(k)}$ is a subset of voxels at iteration k. From each of the 302 clusters, one voxels is randomly chosen to be included as the initial set of constraints, so 302 constraints are included in iteration 1 and this set is updated at each iteration. The model is solved to obtain $w^{*(k)}$, which is the set of optimal beam intensities considering these initial constraints. In the master problem, the objective function is to minimize total radiation dose to Heart and the constraints are meeting the minimum CTV dose requirement for a subset of voxels.

2.2.2 Sub Problem

In the sub problem formulation, $w^{*(k)}$ is used to find $viol_i^{(k)}$, which is the voxels that are violated by the initial optimal solution. To determine the violated voxels, first $viol_i^{(k)} = \theta_i - \sum_j D_{i,j}^T w_j$ is calculated for each voxel, second, if the result of this calculation is less than or equal to zero, the constraint of that voxel is not violated by the current optimal beam intensity solution, but if it is greater than zero, there is a violation. The maximum amount of violation per cluster is determined and the corresponding constraints are added to the master problem. The iteration process starts by updating set of voxels at each iteration by adding the voxels with maximum violation per cluster to $V^{(k)}$ which creates new set of voxels, $V^{(k+1)}$.

Furthermore, the solution of the Master Problem is updated starting from the previous solution, $w^{*(k)}$, and it is updated using dual simplex. This procedure is continued until there are no further violations and there are no more voxels to add.

Figure 2.14 shows the general overview of the proposed constraint generation algorithm. Also, the pseudo-code of the mentioned algorithm can be found in the appendices. Moreover, the results and the sensitivity analysis are also explained in the next chapter.

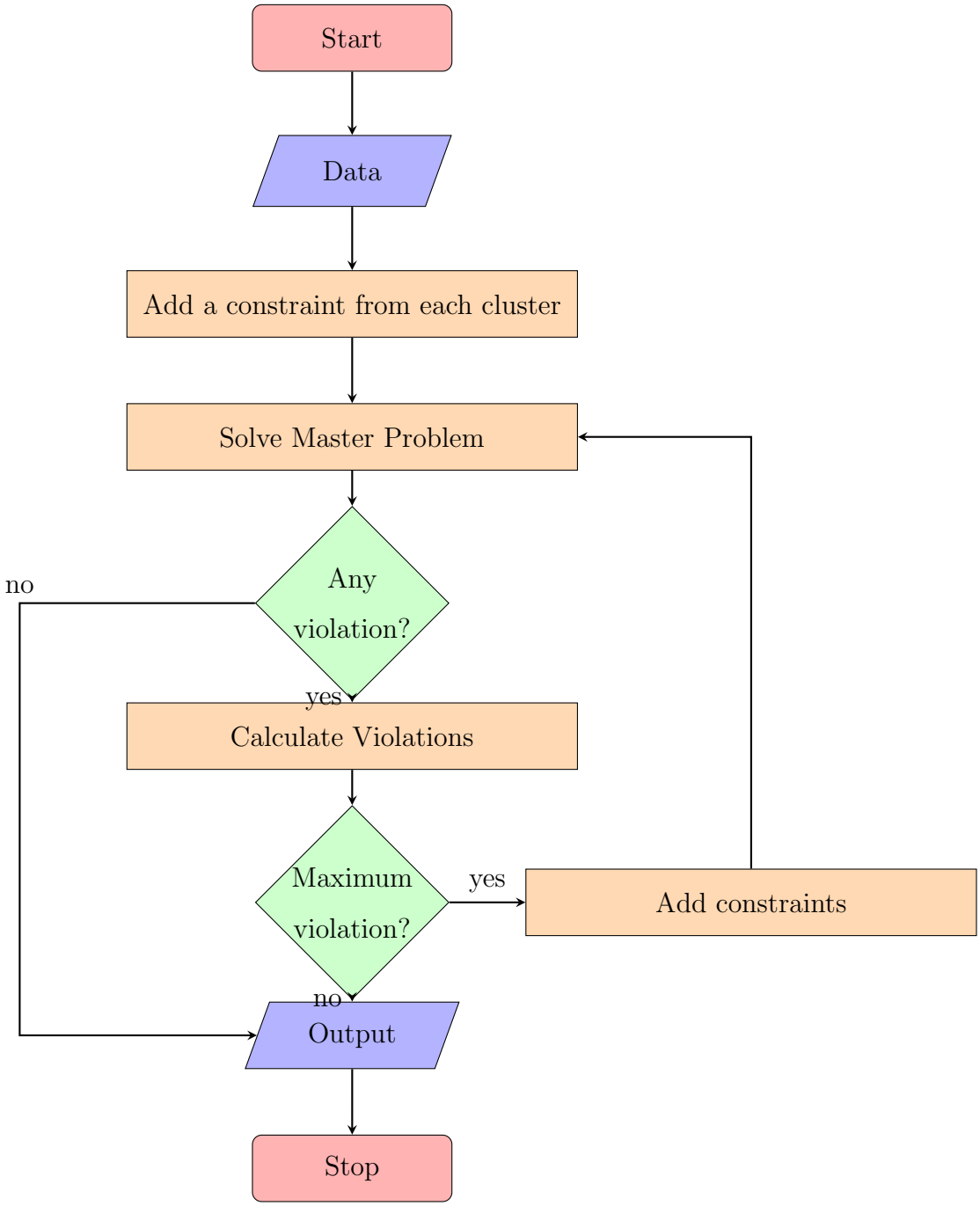


Figure 2.14: Constraint Generation Algorithm

Chapter 3

Results

The optimization problem is coded in Eclipse Java using CPLEX libraries in order to formulate the optimization algorithm explained in the previouse chapter, solve it and get the results. In this section, two patients data sets are considered in order to demonstrate the results of the proposed methodology. The results confirm that the constraint generation algorithm works regardless of the shape and/or size of the CTV and OAR. The initial results are summarized in Table 3.1 where number of voxels in CTV and Heart, number of beamlets, number of clusters and optimal objective value, for patient #1 and patient #2, are described. In the second stage, the optimization model with the constraint generation algorithm is solved with both patient data sets, and it produced the same objective function values, for both cases. This means the proposed constraint generation algorithm provides the same quality as large-scale model and potentially a faster process.

Table 3.1: Summary of initial results for Patient #1 and Patient #2

	#voxels CTV	#voxels Heart	# of beamlets	# of clusters	optimal obj. val.
Patient 1	35,592	42,342	760	302	46.87
Patient 2	56,923	51,709	968	445	51.57

3.1 Analysing Initial Results

Tables 3.2 and 3.3 show results for patient #1 and patient #2, respectively. The different columns show iteration number, number of constraints added in each iteration, total number of added constraints in each iteration and objective function value in each iteration. In this table, δ (i.e. violation amount in Gy) represents the violation criteria for this model. As for the initial considerations, the δ is set to 0.1 and the model met the algorithm stopping condition after 13 iterations for both patients. As can be seen, the number of added constraints at each iteration is decreasing until there are no more constraints that can be added and there are not more violations.

Also, the objective function value starts from 37.20 and reaches an optimal point of 46.87 which cannot be improved further. So, adding further constraints would not improve the solution quality. In the last iteration, the model solved with 1353 constraints, which is 4 percent of the total number of constraints that is used for the large-scale model.

3.2 Sensitivity Analysis

For patient #1, in Tables 3.4 and 3.5, it is shown that with different amounts of δ (i.e. violation amount in Gy), how the model behaves in terms of number of added constraints,

Table 3.2: Results when $\delta = 0.1$ for the Patient #1

Iteration #	# of Constraints Added	Total # of Constraints	Objective Function Value
1	302	302	37.20
2	289	591	45.00
3	257	848	46.26
4	215	1063	46.68
5	138	1201	46.81
6	69	1270	46.85
7	37	1307	46.86
8	12	1319	46.86
9	8	1327	46.86
10	21	1348	46.87
11	2	1350	46.87
12	3	1353	46.87
13	0	1353	46.87

number of iterations and the value of objective function. It is shown that with a higher δ, the total number of constraints is reduced, and with lower δ the total number of constraints is increased. Similar experiments are considered for the second patient and the results are shown in Tables 3.6 and 3.7, which confirm a similar trend.

As for a summary of the analysis for δ's and the outcomes for both patients, Tables 3.8 and 3.8 are created which show different δ's that are considered, total number of iterations, total number of constraints that are taken into consideration, percentage of total number of constraints that are considered in the constraint generation model for each δ, and the

Table 3.3: Results when $\delta = 0.1$ for the Patient #2

Iteration #	# of Constraints Added	Total # of Constraints	Objective Function Value
1	445	445	42.28
2	413	858	48.45
3	387	1245	50.56
4	320	1565	51.20
5	238	1803	51.45
6	143	1946	51.54
7	64	2010	51.56
8	24	2034	51.56
9	12	2046	51.56
10	8	2054	51.56
11	3	2057	51.56
12	2	2059	51.56
13	0	2059	51.56

objective function value.

As can be seen in Table 3.8, the solution quality (i.e. objective function value) is partially affected when δ is equal to 1. But, when δ is equal to 1, there are lower number of iterations for constraint generation and lower number of constraints are needed to satisfy the stopping condition of the optimization compared to other δ values. On the other hand, using δ equal to 0.1 (in this case which is for Patient #1) is suggested for this model because the objective function value would not be affected and maintained the same quality.

In case of Patient #2, the results suggest a different recommended value for δ as it is

Table 3.4: Results when $\delta = 1$ for the Patient #1

Iteration #	# of Constraints Added	Total # of Constraints	Objective Function Value
1	302	302	37.20
2	280	582	45.00
3	213	795	46.16
4	146	941	46.62
5	44	985	46.71
6	16	1001	46.74
7	11	1012	46.76
8	1	1013	46.76
9	0	1013	46.76

shown in Table 3.9. In this case, when δ is equal to 0.1, the solution quality is decreased at the same time of having less number of required constraints. But, δ value of 1 is recommended in this case because it maintains the same solution quality and it also gives lower number of constraints compared to when δ is close to zero. The number of iterations varies between these two cases and would not be a proper criteria to suggest a δ value.

3.3 Beamlet Intensities Fluence Map Results

For verification purposes, the beamlet intensity profiles are generated for both full model implementation and the constraint generation algorithm in Figures 3.1 and 3.2. As can be seen, they both produced similar beamlet intensity profiles for patient #1. This can also certify that the constraint generation model produces the same solution in terms of

Table 3.5: Results when $\delta = 10^{-6}$ for Patient #1

Iteration #	# of Constraints Added	Total # of Constraints	Objective Function Value
1	302	302	37.20
2	289	591	45.00
3	261	852	46.26
4	224	1076	46.68
5	167	1243	46.82
6	106	1349	46.86
7	45	1394	46.87
8	20	1414	46.87
9	9	1423	46.87
10	0	1423	46.87

beamlet profiles as the full model implementation. In these figures, the black pixels have the lowest beamlet intensity and the white pixles have the highest beamlet intensity.

Additionally, an unexpected issue happened in these illustrations. In these 2D profiles, in the margins, high beamlet intensities also exist. This is because the amount of the intensity is not constrained, so outside of the region of interest, beamlets can have high intensities. As shown in Figure 3.1, there are three areas including Heart on the left side, CTV (i.e. the breast) in the center the main area and the outside of body area, which is the air. The high intensities in the margins would not have any effect on CTV area; they are outside the patient's body, which affects the area in the air but not the Heart or CTV areas.

Table 3.6: Results when $\delta = 1$ for the Patient #2

Iteration #	# of Constraints Added	Total # of Constraints	Objective Function Value
1	445	445	42.28
2	408	853	48.45
3	353	1206	50.55
4	232	1438	51.15
5	124	1562	51.36
6	45	1607	51.42
7	12	1619	51.43
8	8	1627	51.44
9	2	1629	51.44
10	1	1630	51.44
11	4	1634	51.44
12	0	1634	51.44

3.4 CTV Dose Statistics

In Figures 3.3 - 3.6, the dose statistics for CTV in both full model and constraint generation model when δ has the value of 1, 0.1 and 0.000001, are provided for the Patient #1. In Figure 3.3 there is no CTV dose violation and 42.4 Gy is the minimum required prescribed dose. Figures 3.4, 3.5 and 3.6, however, contain an allowance for dose violation in the constraint generation model, which are δ values in Gy, hence few voxels violate the minimum required prescribed CTV dose, and, Table 3.10 summarizes the dose statistics, minimum and maximum dose, mean dose and the 90th, 75th, 50th, 25th and 5th percentiles. It is detemined that when δ is close to zero (i.e. 0.000001) the histogram pattern is very similar

Table 3.7: Results when $\delta = 10^{-6}$ for Patient #2

Iteration #	# of Constraints Added	Total # of Constraints	Objective Function Value
1	445	445	42.28
2	414	859	48.52
3	385	1244	50.56
4	320	1564	51.25
5	260	1824	51.47
6	176	2000	51.54
7	89	2089	51.56
8	50	2139	51.56
9	20	2159	51.61
10	61	2220	51.57
11	3	2223	51.57
12	2	2225	51.57
13	0	2225	51.57

to full model results. In these histogram, also, few of the voxels in CTV receives very high dose and this is because of the effect of not having an upper-bound for the original FMO model and also having very high beam dose in the margins of the fluence map.

Similar analysis is done for Patient #2 and Figures 3.7 - 3.10 show the results. Also, when these histograms are compared, it is determined that the results of the full model and the constraint generation model when δ is 0.000001, are identical and produce same patterns. These results certify that the proposed constraint generation algorithm works perfectly without concerning patient's different organ size, shape and number of voxels.

Table 3.8: Summary of δ's for Patient #1

δ	# Iter	Total Const.	% Const	Obj. Val.
10^{-6}	10	1423	4.23	46.87
0.1	13	1353	4.02	46.87
1	9	1013	3.01	46.76

Table 3.9: Summary of δ's for Patient #2

δ	# Iter	Total Const.	% Const	Obj. Val.
10^{-6}	13	2225	3.9	51.57
0.1	12	1634	2.8	51.44
1	13	2059	3.6	51.56

Moreover, Table 3.11 provides a summary of statistics for the full model and constraint generation when the Patient #2 data is being used.

Table 3.10: Summary of CTV dose statistics for both models for Patient #1 ($\delta = 1$)

	Min	Max	Mean	90th	75th	50th	25th	5th
full model	42.4	108.59	46.31	49.89	47.42	45.56	44.20	42.85
constraint generation	41.49	109.10	46.20	49.93	47.37	45.42	44.03	42.69

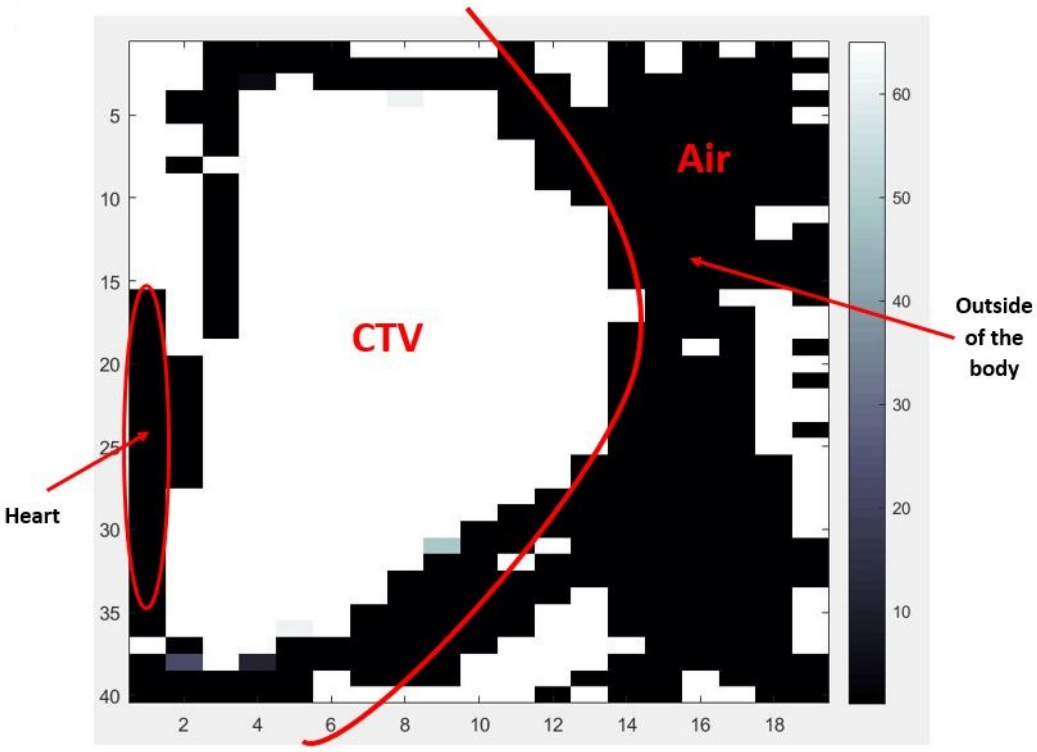


Figure 3.1: optimal beamlet intensities for full model implementation

Table 3.11: Summary of CTV dose statistics for both models for Patient #2 ($\delta = 1$)

	Min	Max	Mean	90th	75th	50th	25th	5th
full model	42.4	174.33	51.15	60.16	54.63	49.82	46.42	43.13
constraint generation	41.47	154.08	51.02	60.12	54.53	49.68	46.26	42.96

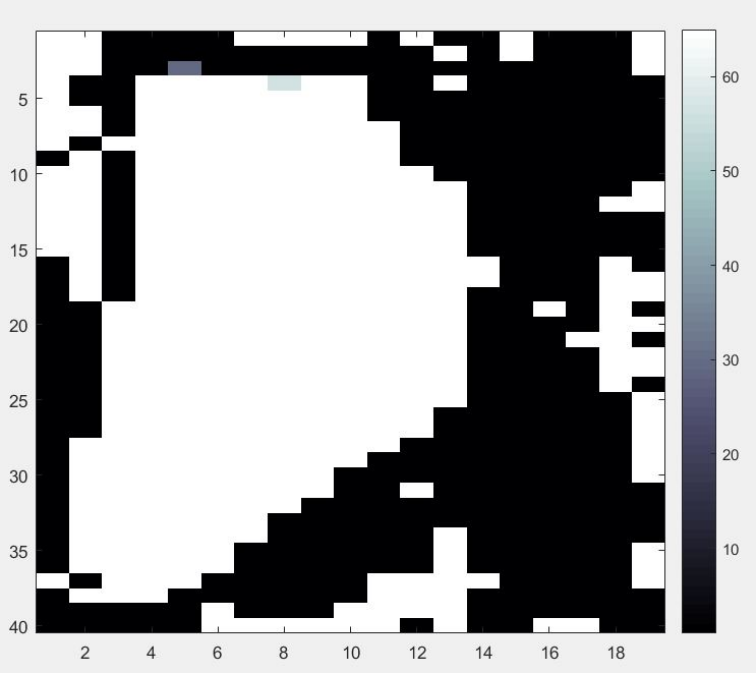


Figure 3.2: optimal beamlet intensities for constraint generation algorithm

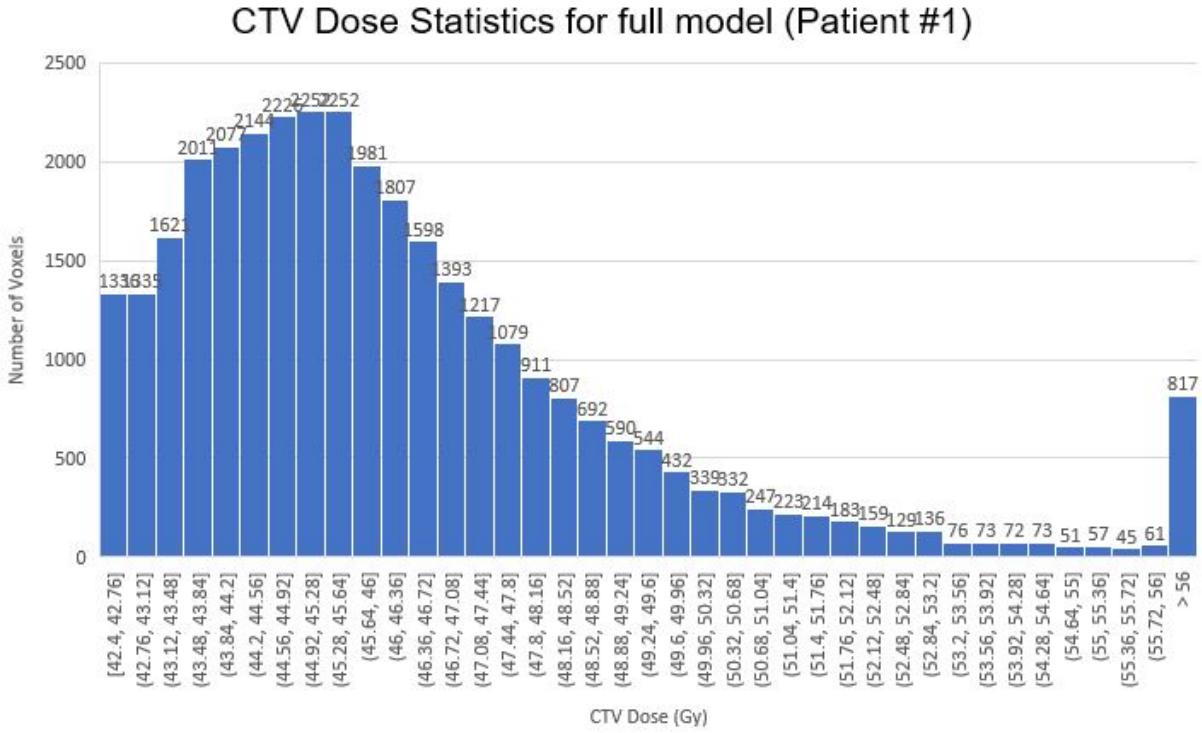


Figure 3.3: CTV dose histogram for full model Patient #1

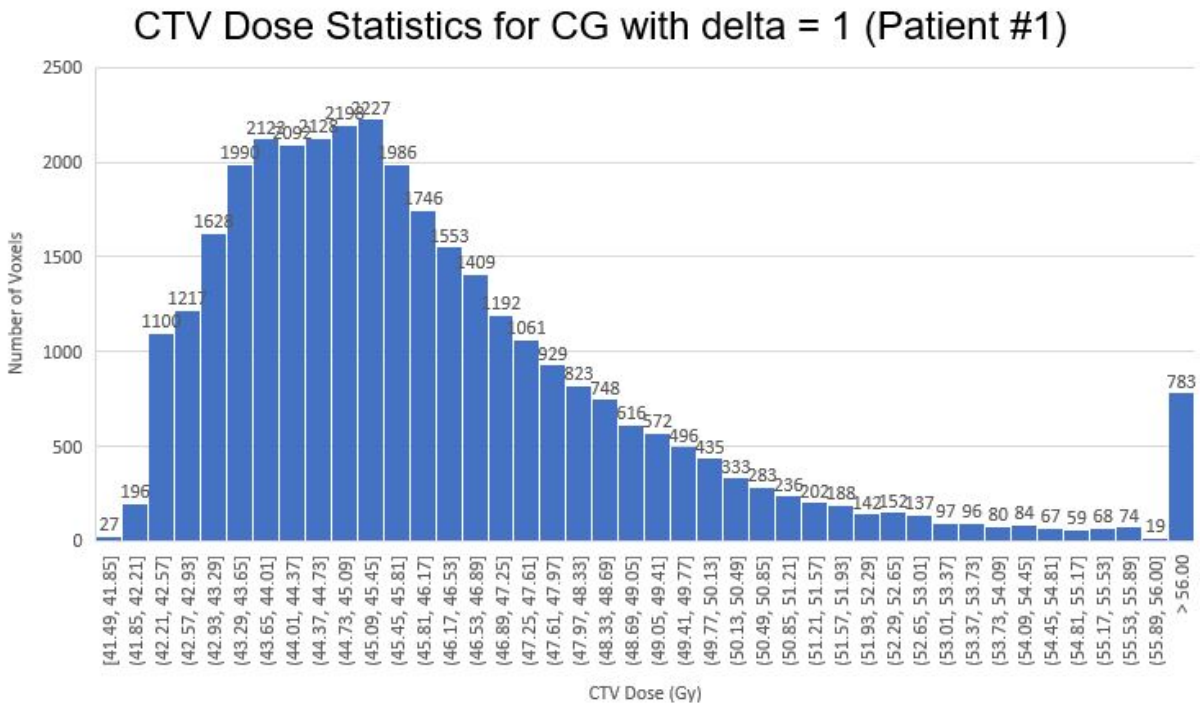


Figure 3.4: CTV dose histogram for constraint generation model when $\delta = 1$ for Patient #1

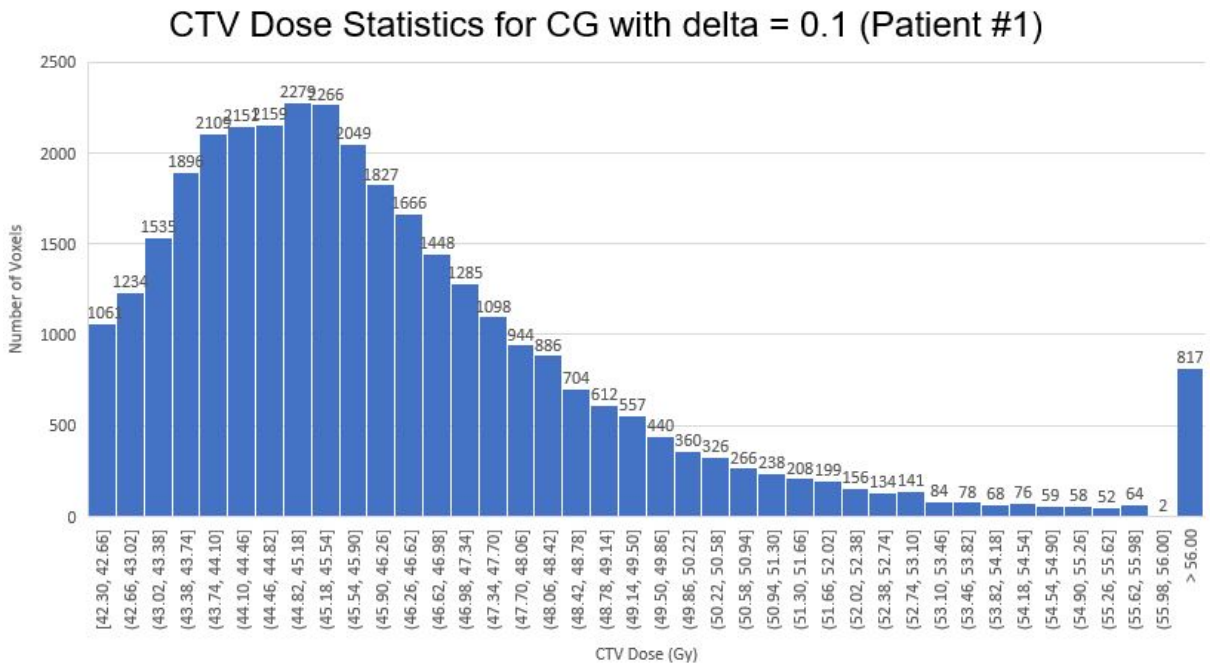


Figure 3.5: CTV dose histogram for constraint generation model when $\delta = 0.1$ for Patient #1

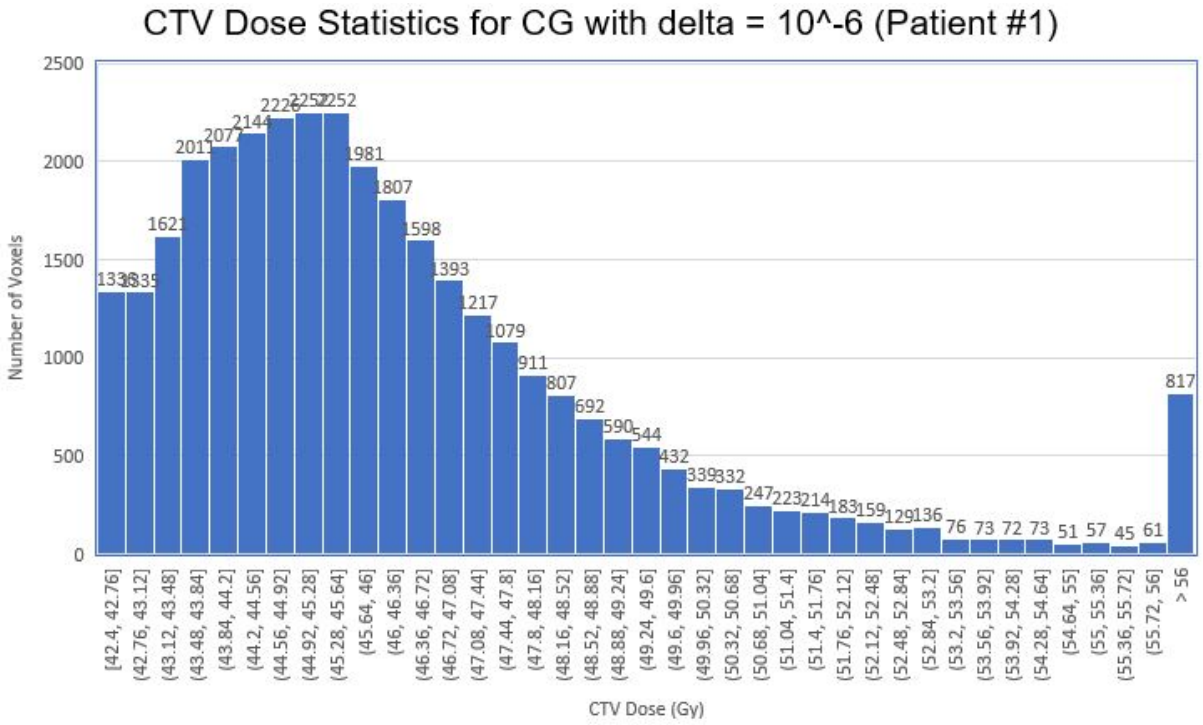


Figure 3.6: CTV dose histogram for constraint generation model when $\delta = 0.000001$ for Patient #1

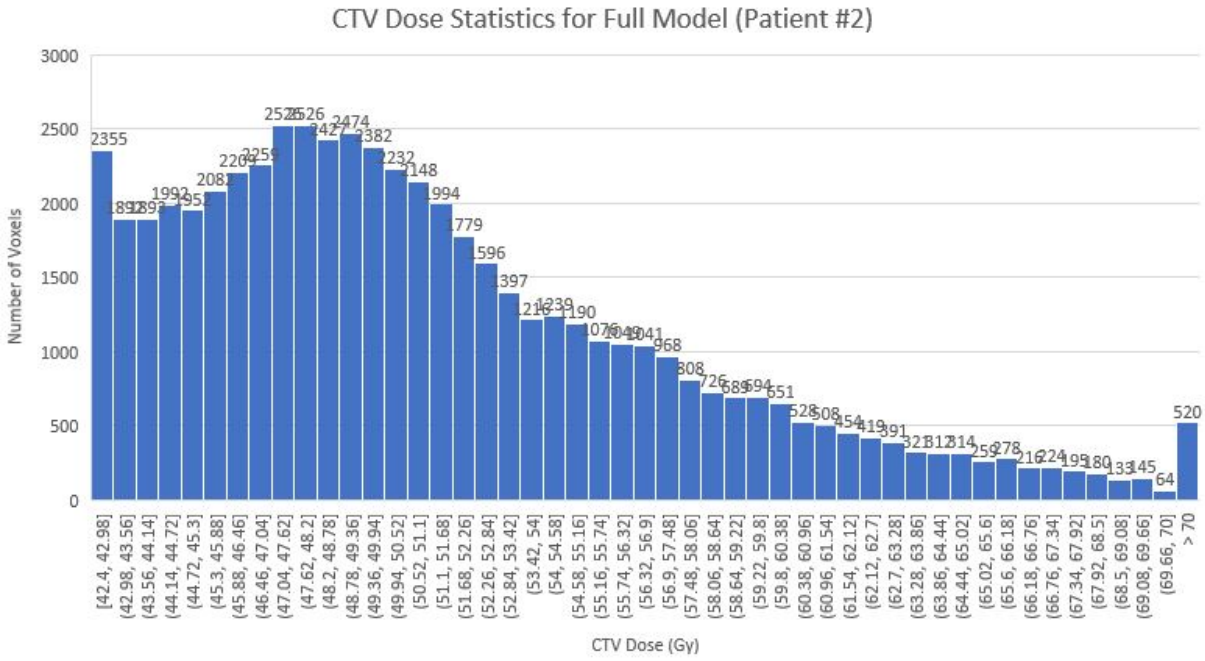


Figure 3.7: CTV dose histogram for full model Patient #2

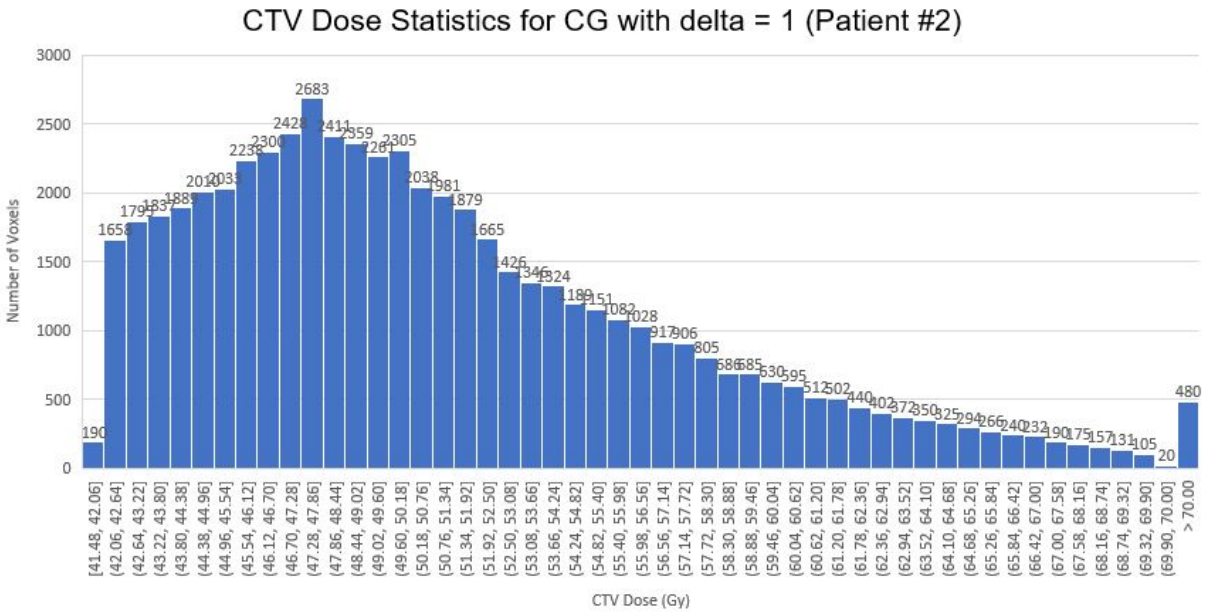


Figure 3.8: CTV dose histogram for constraint generation model when $\delta = 1$ for Patient #2

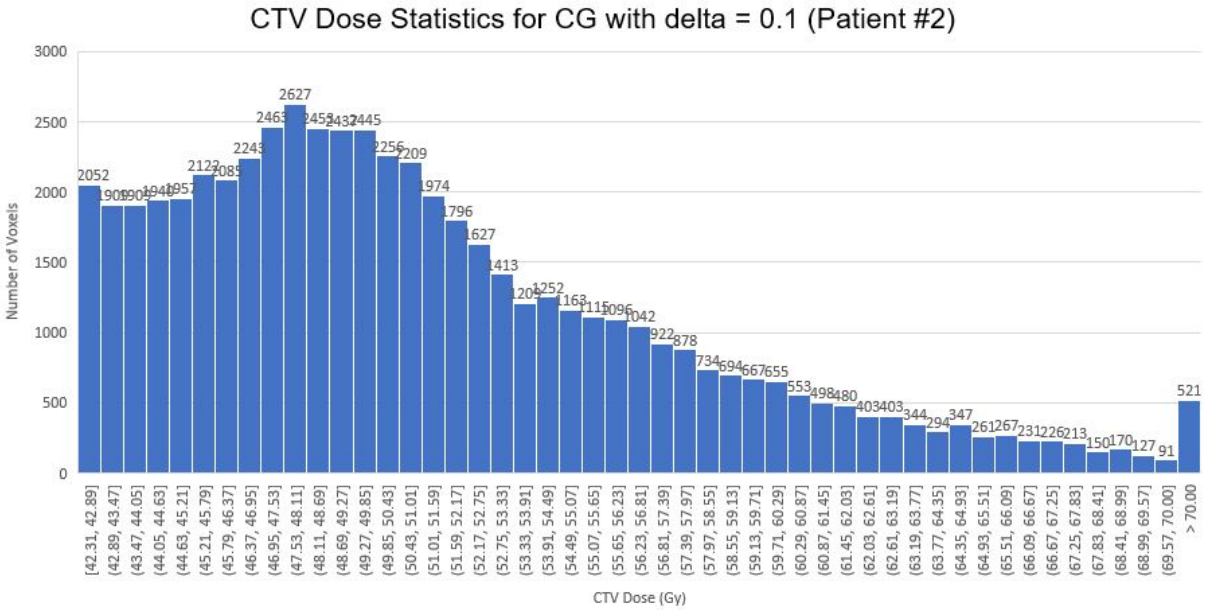


Figure 3.9: CTV dose histogram for constraint generation model when $\delta = 0.1$ for Patient #2

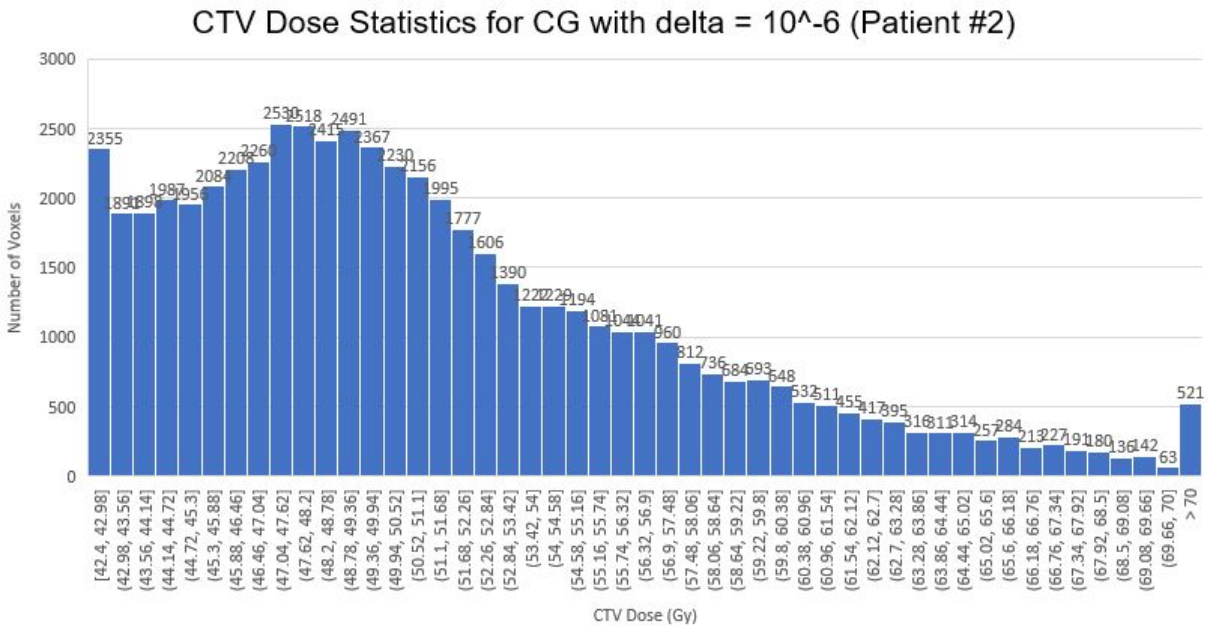


Figure 3.10: CTV dose histogram for constraint generation model when $\delta = 0.000001$ for Patient #2

Chapter 4

Conclusions

A constraint generation solution method for IMRT treatment planning is proposed, and the steps of the algorithm are elaborated. The formulation of FMO framework is explained for two breast cancer treatment planning case studies. Furthermore, the data processing method for creating the clusters of similar voxels with the highest dose associated from beamlets is investigated.

Due to the large number of constraints, these data clusters helped in deciding which voxels to pick as set of new constraints to formulate the sub-problems. It is shown that number of constraints is significantly reduced and optimal beam intensity profile does not violate any of the other constraints in the main model. At the same time, the same quality was maintained in terms of the objective function value. Also, the optimal beam intensities created similar fluence map pattern as clinical model. So, the advantages of this new method which are reducing the number of constraints and maintaining the similar results to full model is proved and faster optimization algorithms can be produced with this method. Additionally, the proposed methodology is tested on two different patient

cases and same behaviour in terms of significant reduction in number of constrains and maintaining the same solution quality, is observed. In other words, this methodology works with different Patient's data.

For the future research, the time component for each algorithm can be investigated further in order to have an exact estimate of time in different algorithms. Also, different sets of criteria for CTV and OAR can be added for this problem, for which reformulation of the mathematical optimization model in both objective function and constraints can be studied. Accordingly, a base model can be created for solving a more advanced optimization model called robust optimization. In this reformulation for IMRT optimization model, the number of constraints can significantly increase in order to incorporate uncertainty in the model.

References

Alber, M. and Nüsslin, F. (1999). An objective function for radiation treatment optimization based on local biological measures. *Physics in Medicine & Biology*, 44(2):479.

Aleman, D., Mišić, V., and Sharpe, M. (2013). Computational enhancements to fluence map optimization for total marrow irradiation using imrt. *Computers Operations Research*, 40(9):2167 – 2177. Operations research for health care delivery.

Aleman, D. M. and Sharpe, M. B. (2011). Optimization methods for total marrow irradiation using intensity modulated radiation therapy. *INFOR: Information Systems and Operational Research*, 49(4):234–240.

Bednarz, G., Michalski, D., Houser, C., Huq, M. S., Xiao, Y., Anne, P. R., and Galvin, J. M. (2002). The use of mixed-integer programming for inverse treatment planning with pre-defined field segments. *Physics in Medicine & Biology*, 47(13):2235.

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

Bortfeld, T., Bürkelbach, J., Boesecke, R., and Schlegel, W. (1990). Methods of image reconstruction from projections applied to conformation radiotherapy. *Physics in Medicine and Biology*, 35(10):1423–1434.

Bortfeld, T., Chan, T. C. Y., Trofimov, A., and Tsitsiklis, J. N. (2008). Robust management of motion uncertainty in intensity-modulated radiation therapy. *Operations Research*, 56(6):1461–1473.

Bortfeld, T. and Schlegel, W. (1993). Optimization of beam orientations in radiation therapy: some theoretical considerations. *Physics in Medicine and Biology*, 38(2):291–304.

Boyer, A., Butler, B., and DiPetrillo, T. (2001). Intensity-modulated radiotherapy: current status and issues of interest. *International Journal of Radiation Oncology*Biology*Physics*, 51(4):880 – 914.

Cao, W. and Lim, G. J. (2011). *Optimization Models for Cancer Treatment Planning*. American Cancer Society.

CCSAC (2019). Canadian cancer statistics 2019. *Canadian Cancer Society*.

Chan, T. C. Y., Bortfeld, T., and Tsitsiklis, J. N. (2006). A robust approach to IMRT optimization. *Physics in Medicine and Biology*, 51(10):2567–2583.

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

Das, S. (2009). A role for biological optimization within the current treatment planning paradigm. *Medical physics*, 36(10):4672–4682.

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

Ehrgott, M., W., H. H., and Lizhen, S. (2010). Mathematical optimization in intensity modulated radiation therapy1. *Annals of Operations Research*, pages 309–365.

Ezzell, G. A. (1996). Genetic and geometric optimization of three-dimensional radiation therapy treatment planning. *Medical Physics*, 23(3):293–305.

Ferris, M. C., Einarsson, R., Jiang, Z., and Shepard, D. (2006a). Sampling issues for optimization in radiotherapy. *Annals of Operations Research*, 148(1):95–115.

Ferris, M. C., Meyer, R. R., and D'Souza, W. (2006b). Radiation treatment planning: Mixed integer programming formulations and approaches. In *Handbook on modelling for discrete optimization*, pages 317–340. Springer.

Gupta, T., Agarwal, J., Jain, S., Phurailatpam, R., Kannan, S., Ghosh-Laskar, S., Murthy, V., Budrukkar, A., Dinshaw, K., Prabhash, K., et al. (2012). Three-dimensional conformal radiotherapy (3d-crt) versus intensity modulated radiation therapy (imrt) in squamous cell carcinoma of the head and neck: a randomized controlled trial. *Radiotherapy and Oncology*, 104(3):343–348.

Hall, E. J. and Wuu, C.-S. (2003). Radiation-induced second cancers: the impact of 3d-crt and imrt. *International Journal of Radiation Oncology* Biology* Physics*, 56(1):83–88.

Hamacher, H. W. and Küfer, K.-H. (1999). Inverse radiation therapy planning: A multiple objective optimisation approach. In *Monitoring, Evaluating, Planning Health Services*, pages 177–189. World Scientific.

Lahanas, M., Schreibmann, E., and Baltas, D. (2003). Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Physics in Medicine and Biology*, 48(17):2843–2871.

Lan, Y., Li, C., Ren, H., Zhang, Y., and Min, Z. (2012). Fluence map optimization (fmo) with dose–volume constraints in imrt using the geometric distance sorting method. *Physics in Medicine & Biology*, 57(20):6407.

Lim, G. J., Choi, J., and Mohan, R. (2008). Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning. *OR Spectrum*, 30(2):289–309.

Mahmoudzadeh, H., Lee, J., Chan, T. C. Y., and Purdie, T. G. (2015). Robust optimization methods for cardiac sparing in tangential breast imrt. *Medical Physics*, 42(5):2212–2222.

Prado, K., Starkschall, G., and Mohan, R. (2007). Three-dimensional conformal radiation therapy. *Treatment Planning in Radiation Oncology, Khan, FM, Editor. Lippincott Williams & Wilkins: Philadelphia, PA, USA*, pages 116–141.

Rocha, H., Dias, J. M., Ferreira, and Do Carmo Lopes, M. (2015). A two-stage programming approach to fluence map optimization for intensity-modulated radiation therapy treatment planning. *6th European Conference of the International Federation for Medical and Biological Engineering*, pages 687–690.

Romeijn, H. E., Ahuja, R. K., Dempsey, J. F., and Kumar, A. (2005). A column generation approach to radiation therapy treatment planning using aperture modulation. *SIAM Journal on Optimization*, 15(3):838–862.

Romeijn, H. E., Ahuja, R. K., Dempsey, J. F., Kumar, A., and Li, J. G. (2003). A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Physics in Medicine and Biology*, 48(21):3521–3542.

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

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

Shepard, D. M., Ferris, M. C., Olivera, G. H., and Mackie, T. R. (1999). Optimizing the delivery of radiation therapy to cancer patients. *Siam Review*, 41(4):721–744.

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

UHN (2015). Imrt.

Webb, S. (1991). Optimization of conformal radiotherapy dose distributions by simulated annealing: II. inclusion of scatter in the 2d technique. *Physics in Medicine and Biology*, 36(9):1227–1237.

Zhang, X., Liu, H., Wang, X., Dong, L., Wu, Q., and Mohan, R. (2004). Speed and convergence properties of gradient algorithms for optimization of imrt. *Medical Physics*, 31(5):1141–1152.

Zhang, Y. and Merritt, M. (2006). Fluence map optimization in imrt cancer treatment planning and a geometric approach. In *Multiscale Optimization Methods and Applications*, pages 205–227. Springer.

Zinchenko, Y., Craig, T., Keller, H., Terlaky, T., and Sharpe, M. (2008). Controlling the dose distribution with geud-type constraints within the convex radiotherapy optimization framework. *Physics in Medicine & Biology*, 53(12):3231.

APPENDICES

Input : D_{ij} for Heart, D_{ij} for CTV and Voxel Clusters Data

Output: Optimal Beamlet Intensities

```
1  Create new IloCplex Model;
2  Initialize new IloNumVar[] w variable for beam;
3  foreach b=0; b < number of beamlets; b++ do
4  |  identify w[b] in the context of cplex.numVar ranges from 0 to Infinity;
5  end
6  Initialize IloLinearNumExpr[] doseCtv for array of voxel indexes in CTV;
7  foreach v = 0; v < number of voxels in CTV; v++ do
8  |  foreach b=0; b < number of beamlets; b++ do
9  |  |  create doseCtv terms for each voxel which is summation of matrix
   |  |   multiplication of intensity vector w[b] and CTV matrix;
10 |  end
11 end
12 foreach b=0; b < number of beamlets; b++ do
13 |  foreach v = 0; v < number of voxels in CTV; v++ do
14 |  |  if voxel index in cluster data == voxel index in CTV data then
15 |  |  |  add a constraint to the model;
16 |  |  |  break and go to next iteration;
17 |  |  end
18 |  end
19 end
```

Algorithm 1: Constraint Generation Algorithm

```
Initialize Objective Function for the optimization model;
foreach b=0; b < number of beamlets; b++ do
    foreach v = 0; v < number of voxels in Heart; v++ do
        add average Heart dose;
    end
    foreach v=0; v <number of voxels in CTV; v++ do
        add average dose CTV;
    end
end
Create Minimization function for objective;
solve the model;
foreach j=0; j<number of voxels in CTV; j++ do
    foreach k=0; k<760; k++ do
        calculate sum[j] += w1[k] * D2[k][j];
    end
    calculate violation delta[j] = theta - sum[j] ;
    if the violation is maximum in that cluster then
        add a new constraint;
    end
end
solve the model;
```