Elin Hynning

An Integer Programming Approach to Conversion from Static to Continuous Delivery of Intensity Modulated Radiation Therapy

Master Thesis

November, 2011

Division of Optimization and Systems Theory, Department of Mathematics, Kungliga Tekniska Högskolan, Stockholm RaySearch Laboratories, Stockholm E-mail: ehynning@kth.se
Abstract

The aim of this thesis is to generate a starting guess for volumetric modulated arc therapy (VMAT) plan optimization by using data from an intensity modulated radiation therapy (IMRT) plan. We develop a mixed-binary linear programming model, based on choosing segments among a set of predefined segments. The objective is to deliver intensity modulation as similar as possible to the intensity modulation of the IMRT plan. The quality of the solutions is largely dependent on the quality of the predefined segments. However, the model achieves high similarity in intensity modulation when supplied with suitable segments. Unfortunately, high similarity in intensity modulation does not necessarily imply high similarity in dose distribution. In order for the model to generate VMAT plans with acceptable dose distributions the leaf travel between adjacent control points needs to be kept low. The model shows some promising features, but improvements, especially regarding implementation, need to be made in order for the model to be useful.
Acknowledgements

I would like to thank my supervisor at RaySearch Laboratories, Björn Hårdemark, for great guidance and for giving me the opportunity to write this thesis. I would also like to thank my supervisor at KTH, Anders Forsgren, for feedback and guidance. Finally, I would like to thank Olof Bergvall, for patience and support.

Stockholm, November 2011

Elin Hynning
## Contents

1 **Introduction** ................................................................. 1  
   1.1 Cancer Incidence and Treatment ....................................... 1  
   1.2 Equipment Used for Radiation Therapy Treatment ................. 2  
   1.3 Terminology and Objectives in Radiation Therapy Treatment Planning ............................................. 4  
   1.4 Treatment Constraints Arising from the Method of Delivery ...... 5  
   1.5 The Importance of Treatment Delivery Time ................... 6  

2 **Problem Description** .................................................. 9  

3 **Mathematical Model** .................................................. 13  
   3.1 Introduction of Variables ............................................. 13  
   3.2 The Objective Function ............................................... 15  

4 **Solving the Model** ....................................................... 19  
   4.1 Creating the IMRT Plan ............................................... 19  
   4.2 Reducing Computational Complexity .......................... 20  
   4.3 Implementation ........................................................ 22  
   4.4 The Importance of the Segments Used .................... 23  

5 **Results** ........................................................................ 27  
   5.1 Results for the Prostate Case ................................... 29  
   5.2 Results for the Head-and-Neck Case ............................ 34  
   5.3 Results for the Pelvic Case ...................................... 40  

6 **Discussion** ..................................................................... 49  
   6.1 Results in Fluence Domain ...................................... 49  
   6.2 Results in Dose Domain .......................................... 52  

7 **Further Outlook** ......................................................... 55  
   7.1 Providing Better Segments ..................................... 55  
   7.2 Include Objective Weight in Fluence Maps .................. 56
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.3 Variation of the Model Parameters</td>
<td>56</td>
</tr>
<tr>
<td>7.4 Reduce Complexity Associated with the Objective Function</td>
<td>57</td>
</tr>
<tr>
<td>7.5 Heuristic Solution Methods</td>
<td>57</td>
</tr>
<tr>
<td>8 Conclusion</td>
<td>59</td>
</tr>
<tr>
<td>References</td>
<td>61</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

1.1 Cancer Incidence and Treatment

Cancer is a widespread disease that has been incident throughout times. Due to the vastly prolonged lifespan in industrialized countries, there has been an increase in cancer incidence over the last century. Statistically, one in three people in Sweden will be diagnosed with cancer sometime during their lifetime [1].

Radiation therapy has been used for treating cancer for over 100 years, basically since the discovery of X-rays. In radiation therapy, highly energetic radiation beams are used to irradiate the tumor and thereby causing damage to the DNA of the cancerous cells and killing them. Radiation therapy exploits the fact that cells which regenerate often, such as cancer cells, are more sensitive to radiation than other cells. However, the radiation will harm irradiated healthy cells as well as cancer cells. Thus, it is not possible to irradiate the tumor with the desired dose and disregard the side effects on healthy cells [5].

The goal of radiation therapy is to deliver the desired deadly radiation dose to the tumor while minimizing the damage to healthy cells. Unfortunately, this goal may be very difficult to achieve.

One key factor to the success of radiation therapy is the fact that received dose adds up linearly over short periods of time (up to about 15 minutes). Thus, it is possible to irradiate a tumor from different directions and obtain the same tumor dose as one would have if the tumor would have been irradiated from only one direction. By irradiating a tumor from different directions, the unavoidable irradiation of healthy cells is spread out, thereby reducing the risk for unwanted side effects. Moreover, different kinds of blocking devices can be used in order to spare healthy tissue as much as possible.

For the past 25 years, radiation therapy research and development has been focused on intensity modulated radiation therapy. The technique is gen-
erally referred to as IMRT and this acronym will be used throughout this report.

Intensity modulation means that instead of having a uniform intensity of radiation over the radiation field, the intensity changes with the position in the field. Thus, we may have high intensity in the parts of the field that irradiate the tumor and lower intensity in the parts of the field that irradiate sensitive organs. This technique makes it possible to irradiate tumors situated close to sensitive organs without causing unacceptable damage to these organs.

The use of IMRT has introduced a need for optimization in treatment planning of radiation therapy. While with older techniques it was possible for physicians to pick out directions and field shapes just by anatomical knowledge and experience, the degrees of freedom introduced by IMRT makes it impossible to attain optimal treatment plans without the aid of advanced mathematical algorithms. A summary of research done in the field of IMRT optimization can be found in e.g. [3].

An IMRT treatment plan usually consists of irradiation from 7-11 directions, i.e. with 7-11 optimized beams. As one might expect, the use of a large number of beams allows for more freedom in the optimization and thereby the possibility of creating better treatment plans. In recent years, researchers have tried to exploit this fact by developing techniques for continuous radiation delivery. Instead of delivering radiation from static directions, the radiation beam is swept around the patient, thereby irradiating the patient from all directions. Techniques for continuous radiation delivery go under a few different names. In this report, the acronym VMAT (volumetric modulated arc therapy) will be used.

1.2 Equipment Used for Radiation Therapy Treatment

The most common way to deliver radiation therapy is by a photon beam generated by a linear accelerator. The photon beam emanates from a beam head which is mounted on a gantry. The gantry can be moved 360° and the patient is placed on a couch in such a way that the gantry can move all the way around the patient. A linear accelerator with a patient positioned on the treatment couch is shown in Figure 1.1(a).

The device used for creating intensity modulation is the multi-leaf collimator, which will be referred to as the MLC. The MLC is placed in the gantry head and is used to shape the radiation field. It consists of around 40 tungsten leaf pairs that can be moved back and forth more or less independently of one another. The leaves are usually 5 to 7 cm thick, so almost all radiation is blocked by the leaves. The leaves are 5-10 mm wide, which enables the creation of highly diverse shapes. A configuration of the leaves of the MLC
that is used for creating a beam shape is referred to as a segment. An MLC is shown in Figure 1.1(b).

Fig. 1.1 Equipment used for radiation therapy.

For static IMRT delivery, i.e. non-VMAT delivery, there are two major techniques for intensity modulation. The first one is referred to as “step-and-shoot”, or “SMLC”, which is short for Static Multi-Leaf Collimator. With this technique, each angle is irradiated by a number of different segments, all allowed to be irradiated with different amounts of fluence. When the fluence of one segment has been delivered, the beam is switched off and the leaves of the MLC move to their positions in the next segment. The accumulated result of the irradiation by different segments will be that different parts of the field have received different amounts of fluence. Thus, intensity modulation is achieved.

The second technique for static IMRT delivery is referred to as “sliding window”, or “DMLC”, which is short for Dynamic Multi-Leaf Collimator. When intensity modulation is created with this technique, the leaves of the MLC move continuously while the field is irradiated. By moving different leaves with different speeds, different parts of the radiation field will receive different amounts of fluence, and thus we have intensity modulation.

In VMAT delivery the beam moves continuously while the patient is irradiated. Thus, intensity modulation cannot be created with any of the above described techniques, since only one segment can be delivered from each direction. Instead, the intensity modulation is achieved by summing up fluence from adjacent angles. By moving the leaves of the MLC while the beam is moving, the segments at adjacent angles differ and we get intensity modulation. The fact that we may view the sum of fluence delivered from adjacent angles as intensity modulation is shown in [10].
1.3 Terminology and Objectives in Radiation Therapy

Treatment Planning

When setting up a plan for radiation therapy, we need to be able to define which parts of the patient that we want to irradiate and which parts that need to be spared. Any region of a patient that needs consideration in treatment planning is referred to as a “region of interest”, ROI.

We differentiate between the parts that should be irradiated, i.e. the tumor and regions where there might be unseen cancerous cells, and the parts that should be spared. The latter are referred to as “organs at risk”, OAR. Regarding the tumor and its surrounding, we distinguish between the “gross tumor volume”, GTV, the “clinical target volume”, CTV, and the “planning target volume”, PTV. The GTV is the volume where there is known tumor infiltration, whereas the CTV is the volume of known and suspected tumor infiltration. Thus, the CTV contains the GTV and any parts of its surroundings where there is a certain probability that cancerous cells have spread. The GTV typically consists of the primary tumor and metastasis. If it is possible, different GTVs are usually defined for the primary tumor and the metastasis. The PTV is a geometrical concept that is used to ensure that the prescribed dose is delivered to all parts of the CTV with a clinically acceptable probability. The concept of the PTV is intended to counteract for example uncertainties from organ motion and patient setup [6].

Depending on what kind of ROI is considered, different objective functions are used in the treatment planning. Here, some of the most common objective functions that are used in treatment planning will be introduced.

The simplest objective functions are maximum and minimum dose. If a maximum/minimum dose objective is used on an ROI, then any voxel in the volume that absorbs more/less than the specified dose is penalized. The maximum dose function is typically used to avoid harmful levels of irradiation of OARs, while the minimum dose function is used to make sure that a CTV or PTV is irradiated with the prescribed dose level.

A uniform dose objective function is a combination of the maximum and minimum dose functions. Here, all voxels that absorb dose that deviates from the reference dose are penalized. Uniform dose objectives are usually used on CTVs/PTVs to avoid hot spots and cold spots in the tumor.

A dose falloff objective function is used to specify how the dose should fall off from a higher level to a lower level. A dose falloff function is defined and the voxels in the ROI that deviates from this function are penalized. The dose falloff function usually specifies an upper dose level, a lower dose level and a distance which indicates the maximum allowed distance to lower the dose level from the upper level to the lower level. These objectives are typically used for ROI positioned close to a CTV/PTV.

The last objective function that will be mentioned here is the equivalent uniform dose function, EUD. The EUD function is introduced to consider bi-
1.4 Treatment Constraints Arising from the Method of Delivery

Depending on which technique is used to deliver the radiation therapy treatment, we will be faced with constraints associated with the delivery technique. In this report, it is assumed that the radiation therapy treatment is delivered with a linear accelerator and that the fluence modulation is obtained using an MLC. The most important constraints imposed on the treatment by this delivery technique will therefore be discussed.

We will start by describing the constraints imposed by the multi-leaf collimator. The MLC is a versatile device, but there are still limitations on the segments the MLC can create. Firstly, since there is only one leaf pair shaping each channel, it is not possible to have more than one open slot in each channel. Thus, the MLC cannot construct shapes with more than one opening in the dimension in which the leaves move. This constraint is quite intuitive and holds for all types of MLCs. However, there are a number of model-specific constraints as well.

In order to protect the leaves from damage incurred by collision, some models of MLCs prohibit complete closing of a channel, i.e. there is a limit on how close two opposing leaves are allowed to be positioned. This constraint will be referred to as the minimum gap constraint.

Another constraint imposed in order to avoid leaf damage is the interdigitation constraint. This constraint is similar to the minimum gap constraint, with the difference that it limits how close opposing leaves from neighboring pairs are allowed to be positioned. This constraint makes it impossible to create shapes where any leaf from the left bank is positioned to the right of any leaf from the right bank and vice versa.
Due to design features, certain MLC models impose restrictions on how far away leaves emanating from the same bank may be positioned. Regardless of MLC design, there is always a limitation on how fast the leaves of the MLC can move. The numerical speed limitations differ from model to model, but for most models the maximum speed is 2-3 cm/s. This restriction is especially important in the case of VMAT delivery, since it limits the amount the leaves can move between adjacent angles and thus the intensity modulation that can be achieved.

The movement of the gantry also affects the treatment. This is particularly important for delivery of VMAT, since then the gantry moves continuously during treatment delivery. The maximum gantry speed is the ultimate limit on how fast VMAT treatment can be delivered. However, due to the limited leaf speed, treatment is often not delivered at maximum gantry speed. The acceleration of the gantry also needs to be taken into account. Since the gantry has a large angular momentum, large accelerations can cause uncertainties in speed and position of the gantry and should therefore be kept to a minimum.

Another limiting feature emanating from the linear accelerator is the upper and lower limit of the dose rate. These limits may force the gantry to slow down, if there is an interval where we need to deliver more fluence than is possible with maximum dose rate at the current gantry speed. The lower limit on the other hand may cause over-irradiation in certain intervals, if we want to deliver less fluence than is possible at minimum dose rate with maximum gantry speed.

### 1.5 The Importance of Treatment Delivery Time

One of the goals of VMAT is to shorten the time of treatment delivery. There are a few important reasons to why it is important to keep the delivery time as short as possible.

Commercially, shorter treatment times are beneficial since such will allow treatment facilities to treat a larger number of patients each day. Thus, more patients can get the treatment they need and the cost for each treatment is reduced. However, since a treatment session involves more than just irradiating the patient, the treatment time needs to be reduced substantially in order to reduce the time of a treatment session.

When it comes to patient comfort and treatment accuracy, it is very important to keep the treatment time as short as possible. Receiving radiation therapy is in itself not uncomfortable, since the patient does not feel the radiation. However, since IMRT techniques generate very precise treatment plans, it is very important that the patient is positioned exactly the way intended when the treatment plan was created. Thus, the patient is required to be still throughout the treatment and depending on the location of the
tumor, the position the patient is required to keep may be quite uncomfortable. For e.g. head-and-neck cancers, it is not uncommon that the patient is required to wear a mask that fixates their head. Thus, a shorter treatment time will reduce discomfort for the patient. Moreover, it may be very hard for the patient to be completely still for a longer period of time. Therefore, shorter treatment time will also reduce the effects of uncertainties in patient position.

There is also a biological aspect of the treatment time. In general, a cell can recover from most damages within half an hour [5]. Therefore, if the treatment time starts to approach considerable fractions of this time, then the cells can start to heal during the treatment session. Hence, if the treatment times are too long, the treatment will not have the anticipated effect on the tumor cells. However, the healthy cells will also start to heal, which will improve the sparing of healthy tissue. Thus, it is not obvious from a biological point of view that shorter treatment time is better. With intensity modulated radiation therapy, the researchers try to shorten the treatment times so that they are comparable to the treatment times of simpler techniques. In this way, it is at least ensured that the treatment time does not have a negative impact on the treatment compared to other techniques.

In conclusion, there is a lot to gain from shorter treatment times, although not all positive effects will be observed unless the reduction is substantial.
Chapter 2
Problem Description

There are numerous techniques developed for optimizing IMRT treatment plans. However, there are not that many available techniques for optimizing VMAT treatment plans. This is mainly due to the limited leaf and jaw speed, which constrains the difference between adjacent segments and thereby obstructs the possibility for intensity modulation. Some examples of existing VMAT optimization techniques can be found in e.g. [11], [8] and [9].

Since optimized IMRT plans can be obtained quite easily with sophisticated treatment planning systems, such plans may be used as a starting point for creation of optimized VMAT plans.

The aim of this thesis is to find a method for conversion from an optimized IMRT plan to a starting guess for an optimized VMAT plan. The reason why we are only looking for a starting guess for an optimized VMAT plan and not a fully optimized plan is that IMRT and VMAT are such different delivery techniques, and the chances that we shall be able to create an optimal VMAT plan by only using information obtained from an optimized IMRT plan are therefore quite slim.

The reason why we want to find a good starting guess is that the problem of optimizing VMAT treatment plans is a non-convex optimization problem. Thus, if we use local search, depending on where we start, we may end up with different locally optimal solutions. Hence, if we start with a solution which is close to the global optimum, the chances of finding a globally optimal solution increases.

Thus, this thesis will be about using information from an optimized IMRT plan to create a VMAT plan that is sufficiently good to be used as a starting guess for further optimization.

Since this is a conversion problem, only the information gathered from the IMRT treatment plan will be used to create the starting guess for the VMAT plan. Thus, the problem will be posed as creating a VMAT plan that delivers fluence as similar as possible as the fluence delivered in the optimized IMRT plan. When creating the starting guess, we will only use computation
of the dose as a means for evaluating the created plans, not for any direct optimization.

Before introducing the basic ideas for the conversion method, a few considerations will be mentioned.

When creating a radiation therapy treatment plan using a treatment planning system, the plan is specified by leaf positions and segment weight (the amount of fluence that should be delivered through the segment) at different control points. For a step-and-shoot IMRT plan, the control points specify the leaf positions of the segments and the segment weight for each beam. Thus, there is one control point for every segment of every beam and all control points are positioned at an angle for which there is an optimized beam. In VMAT delivery on the other hand, the control points are equidistantly spaced along the arc. The angular distance between adjacent control points is usually between 2 and 8 degrees. Thus, in order for the solution of the optimization problem to be easily transferable to a treatment planning system, we will pose the problem as specifying leaf positions and segment weights at control points.

We also need to take into consideration that when we want to create an IMRT plan that is supposed to be used for conversion, we need to create a plan with a larger number of beams than we would for a normal IMRT treatment. For normal IMRT treatments, it is usually enough to use 7-11 equidistant beams. Thus, the beams are placed more than $30^\circ$ apart. This spacing is too sparse if we want to convert the IMRT plan into a VMAT plan. The reason is that in order to optimize segments and segment weights for the control points, we need to be able to model the fluence of the control points as fluence given at the angles for which the fluence maps are optimized. With the beam spacing of a normal IMRT plan, we will have control points that we cannot model as though their fluence contributes to fluence at an optimized angle without large errors. At these control points, we do not know what fluence we want and we will therefore be unable to optimize their segments and segment weights. Therefore, we need to generate IMRT plans with a larger number of beams than what is usually used for clinical IMRT treatment.

The idea behind the conversion technique considered in this thesis is simple; instead of delivering the segments for the optimized IMRT beams at the same angle for which they are optimized, we spread the segments over adjacent angles. Now, we cannot use all segments for each IMRT beam, since there are usually about 10 segments generated for each beam. If we place the beams sufficiently tight, we will not be able to deliver all these segments without using an unreasonable number of arcs. Thus, we need to choose which segments to use in some way. We will therefore express our optimization problem as choosing, for each control point, one segment among the segments optimized for IMRT beams, and a weight associated with each segment. The objective of the choice is that fluence delivered with these segments and associated weights should be as similar as possible to the fluence
maps generated for IMRT delivery. Of course, we can choose between segments from a different set than the segments optimized for the IMRT plan, but this set of segments is a natural starting point.

By expressing our optimization problem as choosing segments from a set of predefined segments, we can disregard a lot of the delivery constraints regarding the MLC. Unfortunately, there is no escape from the leaf travel constraint, but as long as we make sure that the predefined segments satisfy the remaining MLC constraints, this is the only MLC constraint we need to consider in our optimization problem.

In order to create plans that are closer to being deliverable from a leaf travel point of view, it is also important to make sure that the leaves do not move too far between adjacent control points. We will therefore, in addition to minimizing the difference between delivered fluence and the IMRT fluence maps, also minimize the sum of the distance the leaves move between each control point.

The ideas described above constitute the basic model that will be used for conversion. A formal mathematical description of the model will be given in the subsequent chapter.
Chapter 3
Mathematical Model

We will now formulate the idea for conversion described in Chapter 2 as a mathematical model.

It is assumed that we want to generate a plan where the control points are equidistantly spaced along the arc. For each control point, a segment needs to be chosen from the set of available segments and the weight of this segment also needs to be specified.

3.1 Introduction of Variables

In order to formulate the model, we will start by introducing the variables we will use. The necessary constraints in order for the variables to represent the things we intend them to represent will also be introduced.

To model the choice of segments for each control point, binary variables, $x_{i,j}$, that represent the following choices are introduced:

$$x_{i,j} = \begin{cases} 
1 & \text{if segment number } j \text{ is chosen at control point } i \\
0 & \text{otherwise}
\end{cases}$$

Here, $i = 1, \ldots, I$, where $I$ is the number of control points and $j = 1, \ldots, N_i$, where $N_i$ is the number of segments made available at control point $i$.

We are only allowed to choose one segment for each control point. Thus, the following constraint must hold:

$$\sum_{j=1}^{N_i} x_{i,j} = 1, \forall i.$$

Since we do not want to select fractional segments, we will also require all $x$-variables to be binary:

$$x_{i,j} \in \{0,1\}, \forall i, j.$$
Variables for the weights of each segment are also needed. These will be denoted \( w_{i,j} \) and they represent the weight of segment number \( j \) at control point \( i \). Since there are limits on the dose rate of a linear accelerator, there is a lower and an upper limit that the segment weight must be kept between. These limits will be denoted \( w_{\text{min}} \) and \( w_{\text{max}} \). Since \( w_{\text{min}} \) is positive, we only want to enforce these limits on the weights of the segments that are chosen, because we want the weights of the segments that are not chosen to be zero. We will express this with the following constraints on the segment weight variables:

\[
\begin{align*}
\frac{w_{\text{min}}}{x_{i,j}} \leq w_{i,j} \leq \frac{w_{\text{max}}}{x_{i,j}}.
\end{align*}
\]

In order to keep track of the distance that the leaves travel, we need variables that measure the distance the leaves travel between control point \( i \) and control point \( i + 1 \). These variables will depend on which segments are chosen for control point \( i \) and \( i + 1 \) respectively. Note that it is only the leaf that travels farthest between control point \( i \) and control point \( i + 1 \) that is of interest to us, since it is this distance that will decide the time it takes to move from control point \( i \) to control point \( i + 1 \), and thus the delivery time of the plan.

In order to express these distance variables, we will start by computing distance matrices \( D_i, i = 1, \ldots, I - 1 \). \( D_i \) is the distance matrix associated with the distance between control point \( i \) and control point \( i + 1 \). Element \((j, k)\) of \( D_i \) corresponds to the distance traveled by the leaf that travels farthest between segment number \( j \) from control point \( i \) and segment number \( k \) from control point \( i + 1 \).

To compute the distance the leaves travel between two adjacent control points, we need to introduce some auxiliary variables, \( u_{i,j,k} \). We want \( u_{i,j,k} \) to be 1 if segment \( j \) is chosen at control point \( i \) and segment number \( k \) is chosen at control point number \( i + 1 \). Otherwise, we want \( u_{i,j,k} \) to be 0. We can express this by the following constraints:

\[
\begin{align*}
u_{i,j,k} \geq 0, \\
u_{i,j,k} \geq x_{i,j} - (1 - x_{i+1,k}).
\end{align*}
\]

Note that even though we do not have an interpretation of a fractional value of \( u_{i,j,k} \) we do not need to require \( u_{i,j,k} \) to be binary. The reason for this is that the second constraint requires \( u_{i,j,k} \) to be larger than or equal to -1, 0 or 1, depending on the values of \( x_{i,j} \) and \( x_{i+1,k} \). The first constraint makes all variables non-negative. Thus, these constraints are enough to make sure that \( u_{i,j,k} \geq 1 \) when we want \( u_{i,j,k} = 1 \) and that \( u_{i,j,k} \geq 0 \) when we want \( u_{i,j,k} = 0 \). Since these variables are auxiliary variables used for calculating the distance between two control points, and the distance is something that we want to minimize, these variables will take on their smallest allowed value during optimization. Thus, the \( u \)-variables will not take on a value that is not 0 or 1 even though we do not require them to be binary.
3.2 The Objective Function

We may now introduce the variables for the distance between control point \(i\) and control point \(i + 1\). These will be denoted by \(d_i\) and are given by the following expression:

\[
d_i = \sum_{j=1}^{N_i} \sum_{k=1}^{N_{i+1}} u_{i,j,k} \cdot D_i(j, k).
\]

Here, the notation \(D_i(j, k)\) is used to denote element \((j, k)\) of the matrix \(D_i\). It should also be noted that we do not necessarily sum over the same set for the indexes \(j\) and \(k\), since the set of segments we are allowed to choose from may differ between control point \(i\) and control point \(i + 1\).

These are the variables we need to model the problem. However, if we want to express the problem as a mixed binary linear programming problem, we will need to include extra variables in order to express the objective function as a linear function. This will be explored further in the next section.

3.2 The Objective Function

In Chapter 2 we stated that we want to minimize the difference between the desired fluence, i.e. the fluence specified in the fluence maps from the IMRT plan, and the delivered fluence. In order to minimize this difference, we need a way of computing the delivered fluence and measure the difference between this fluence and the desired fluence.

The desired fluence is expressed as fluence delivered at specific angles. The fluence of our VMAT plan will be delivered at control points. In general, the positions of the control points do not coincide with the angles for which the fluence maps are optimized. Thus, we need to model how fluence delivered at the control points correspond to fluence delivered at the optimized angles. Here, we will use a quite crude approximation to compute the fluence delivered at an optimized angle.

We will assume that fluence at an optimized angle is the accumulated fluence from all control points within an angular distance of \(\pm 10^\circ\). If the fluence from a control point adds to the fluence at more than one optimized angle, then the contribution of fluence to each optimized angle is computed from the distances between the control point and the optimized angles. We will compute the contributions in the following way: If the distance from fluence map \(i\) to the control point is \(v\) and the distance from fluence map \(i + 1\) to the control point is \(w\), then the fraction of the fluence that adds to fluence map \(i\) is \(\frac{10-v}{10-v+(10-w)}\) and the fraction that adds to fluence map \(i + 1\) is \(\frac{10-w}{10-v+(10-w)}\).

Thus, we now have a way of relating fluence delivered at control points to the optimized fluence maps. It remains to measure the difference between the delivered fluence and the optimized fluence.
Both the delivered fluence and the desired fluence will be expressed as matrices. A way of measuring their difference could therefore be any vector or matrix norm. However, our model includes binary variables as well as continuous variables, while all our constraints are linear. Thus, we would like the objective function to be nice, preferably linear, in order to be able to use methods developed for integer programming.

If we want a linear objective function, we can measure the difference between the matrices as the sum of the absolute values of the difference of all the matrix elements. This is the 1-norm of the matrix, if we view the matrix as a vector. However, in order to formulate the absolute value as a linear function, we will need to introduce an auxiliary variable and two constraints for each matrix element, which unfortunately means the introduction of a vast number of extra variables.

In order to formulate a mathematical expression for our objective function, we need to introduce some notation for the data of our problem.

We need a matrix representation of the optimized fluence maps, in order to be able to compute the difference between the delivered fluence and the desired fluence. These will be referred to as the fluence matrices. We also need a matrix representation of the unit fluence through a segment. These will be referred to as segment matrices.

Let $F_l, l = 1, \ldots, L$, denote the $l$th optimized fluence matrix. Here, $L$ is the number of beams used in the IMRT plan. Each entry in $F_l$ specifies the fluence that should be given to the corresponding bixel.

Let $S_{i,j}$ denote the segment matrix of segment number $j$ at control point $i$. $S_{i,j}$ is a matrix of the same size as $F_l$, where an entry is 1 if the corresponding bixel is irradiated by segment number $j$ at control point $i$ and 0 otherwise. If the positions of the leaves of the MLC do not coincide with the edges of the bixels, we will get bixels at the edges of each row of openings that are partially irradiated. In the segment matrices, the corresponding entries will be the proper fractional number between 0 and 1.

Let $f_{i,l}$ denote the fraction of the fluence at control point $i$ that is added to the fluence corresponding to the $l$th optimized fluence map. This quantity will be computed with regards to how many fluence matrices that get fluence from control point $i$ and the angular distance between control point $i$ and fluence map $l$, as explained in section 3.1.

In order to set up a linear objective function, we will need to introduce auxiliary variables to express the absolute value as a linear function. We will need one variable for each matrix element of the fluence matrices. The easiest way to express these variables is to group them as matrices. Thus, we will get $L$ matrices of auxiliary variables and we will denote these by $V_l, l = 1, \ldots, L$. These variables will need to satisfy the following constraints element-wise:

$$
V_l \geq F_l - \sum_{i=1}^{I} f_{i,l} \cdot \sum_{j=1}^{N_i} w_{i,j} S_{i,j},
$$

$$
V_l \geq \sum_{i=1}^{I} f_{i,l} \cdot \sum_{j=1}^{N_i} w_{i,j} S_{i,j} - F_l.
$$
The objective function will then be the sum of all the matrix elements for $l = 1, \ldots, L$. Since we want the leaf travel to be kept to a minimum, we will add a term in the objective function to account for this. This term will simply be the sum of all $d$-variables. Not only do we add this term to keep the leaf travel at a minimum, we also need this term in the objective function in order to make sure that the $u$-variables take on their intended values, as explained in section 3.1. Depending on whether agreement in fluence or short leaf travel is most important, the terms accounting for similarity in fluence and the terms accounting for leaf travel can be weighted differently.

Now we may state our complete model with linear objective function. We assume that the optimized fluence matrices, and thus also the matrices of auxiliary objective variables, are of size $P \times Q$. We get the following model:

$$
\begin{align*}
\min & \quad \sum_{l=1}^{L} \sum_{p=1}^{P} \sum_{q=1}^{Q} V_l(p, q) + \sum_{i=1}^{I} d_i \\
\text{s.t.} & \quad V_l \geq F_l - \sum_{i=1}^{I} f_{i,l} \cdot \sum_{j=1}^{N_i} w_{i,j} S_{i,j}, \quad \forall l, \\
& \quad V_l \geq \sum_{i=1}^{I} f_{i,l} \cdot \sum_{j=1}^{N_i} w_{i,j} S_{i,j} - F_l, \quad \forall l, \\
& \quad w_{i,j} \leq w_{\max} \cdot x_{i,j}, \quad \forall i, j, \\
& \quad w_{i,j} \geq w_{\min} \cdot x_{i,j}, \quad \forall i, j, \\
& \quad d_i = \sum_{j=1}^{N_i} \sum_{k=1}^{N_{i+1}} u_{i,j,k} \cdot D_i(j, k), \quad \forall i, \\
& \quad u_{i,j,k} \geq x_{i,j} - (1 - x_{i+1,k}), \quad \forall i, j, k, \\
& \quad u_{i,j,k} \geq 0, \quad \forall i, j, k, \\
& \quad \sum_{j=1}^{N_i} x_{i,j} = 1, \quad \forall i, \\
& \quad x_{i,j} \in \{0, 1\}, \quad \forall i, j.
\end{align*}
$$

We should note that it is possible to introduce more constraints to this model. For example, including the leaf travel distance in the objective function might not be enough to get an acceptable value of the leaf travel distance. To deal with this, we can include extra constraints on these variables. For example, we can specify an upper bound for the sum of the distance variables, or limit the values of the individual variables, i.e. limiting the maximum distance the leaves are allowed to travel between two adjacent control points. Whether or not any extra constraints are necessary differs from case to case and also depends on what we want to achieve with the plan.

We have now formulated a model with both binary and continuous variables, linear objective function and linear constraints. We want to proceed by implementing it and analyze the results.
In this chapter, considerations taken when implementing the model and creating IMRT plans for conversion, are described and discussed.

4.1 Creating the IMRT Plan

The IMRT plans that are used for conversion in this project will all be generated using RayStation (RaySearch Laboratories, Stockholm, Sweden). As explained in Chapter 2, we want the beams to be positioned close enough to make sure that every control point along the arc contributes with fluence to at least one optimized fluence map.

In our mathematical model, we assume that a fluence matrix is influenced by control points positioned within an angular distance of $10^\circ$. Thus, we need the beams to be positioned with a maximum angular distance of $20^\circ$. In other words, we need to have at least 18 beams in our IMRT plan. However, with 18 beams, there will be no control points that influence more than one fluence map, so the problem reduces to choosing a number of segments for each beam and spread them out.

In order to use the full potential of our model, we consider the use of 36 beams. With this number of beams, we have overlap at every control point except for the control points whose positions coincide with the position of an optimized fluence map. To use 36 beams seems to be a good balance between a rich model and a computationally manageable model. Therefore, we will use IMRT plans with 36 beams when the model is implemented.

When generating IMRT plans that are supposed to be used for conversion, we do not want to generate plans in the same way that we would generate plans intended to be used directly as IMRT plans. The reason for this is that the fluence maps generated for clinical IMRT plans tend to be highly modulated with lots of sharp peaks, which requires a large number of segments for delivery. In our VMAT setting, it is hard to produce the level of modulation
required to get fluence similar to these fluence maps. Thus, when we generate IMRT plans for conversion, we will not optimize the plans to the same degree as we would for a normal IMRT plan.

If we want to generate an IMRT plan for IMRT treatment, we usually let the optimization process run for about 40 iterations. In order to avoid highly modulated fluence maps when creating the IMRT plans for conversion, we will let the optimization process run for 10 iterations [4]. Note that these iteration specifications are only used for RayStation.

The control points of the VMAT plan will be spaced with an angular distance of $4^\circ$. When choosing the spacing of the control points we need to take three things into consideration. Firstly, if we choose a spacing that is too sparse, then we will not be able to specify enough segments to get sufficient fluence modulation. Secondly, if we choose a spacing that is too dense, then the segments must have very little leaf travel in between them, since there is not much time for leaf travel if adjacent control points are very close to one another. Thirdly, with a too dense spacing, the computational complexity of the problem becomes too large. A spacing of $4^\circ$ seems to be a good choice. Then we are allowed to choose 4 or 5 segments that add to each fluence map for each arc, without ending up with a problem that is too large to solve properly. However, this spacing does not allow for very much leaf travel. We will deal with this by modifying the segments rather than increasing the spacing of the control points, since a sparser spacing will not allow us to choose enough segments for each fluence map to obtain proper intensity modulation.

This spacing will yield 91 control points for single arc and 182 control points for dual arc. The extra control point for each arc is introduced since RayStation introduces a control point at angle $359^\circ$ and we want to make the solutions of the model easily transferable to RayStation.

### 4.2 Reducing Computational Complexity

As will be explained in further detail later on in this chapter, we will try implementing the model with a few different kinds of segments. The segments will be generated beam-wise. Thus, for each fluence map, a set of segments will be generated for delivery of that specific fluence map. There will be between 5 and 15 segments generated for each fluence map, which means that there will be roughly 360 segments generated in total. If we would allow the model to choose from all of these segments at every control point, we would need 360 binary variables to model the choice at every control point. We will not be able to solve such a model in a reasonable amount of time. Therefore, at every control point, we will only allow the model to choose between segments optimized for the fluence maps that the considered control point contributes to. This restriction should not be very limiting, since it seems rea-
sonable that we want to choose a segment optimized for a fluence map that the control point will influence even when we are allowed to choose between all possible segments.

We have reduced the computational complexity considerably by only allowing the choice of segments optimized for fluence maps that the considered control point contributes to. However, the complexity of the model may be reduced a bit further. If we were to use all the segments generated for each beam in the setting with 36 beams and 91 control points, we would get about 20 segments to choose between at every control point not coinciding with a fluence map. It is possible to solve models with this complexity, but we can make some simplifications that will allow us to reduce the computational complexity without considerable loss of accuracy.

Firstly, we note that since we may only choose 182 segments (if dual arcs are used) in total, there is no way that all 360 segments will be chosen. In fact, we will probably not even choose 182 different segments, but rather repeat segments at some adjacent control points. This does not mean that we should disregard half of the generated segments, but we may disregard a couple of segments per beam without substantial risk of degrading the plan. For most beams, there are segments that are considerably smaller than the rest of the segments. These may be necessary to obtain exactly the right fluence modulation of a certain high fluence area, but when the number of segments to use is limited, it is unlikely that the small segments be the optimal choice, especially for control points where the fluence is shared between different fluence maps. Thus, in order to reduce the number of segments we can choose between at each control point, and thereby reducing the number of binary variables, we can disregard the smallest segments from each beam. How many segments we should disregard depends on what balance between computational complexity and goodness of solution we want.

When the segments from the IMRT plan are used, we do not need to choose which segments to make available based on segment size. Instead of disregarding segments based on their size, which may be a bit misleading for certain cases, we may disregard segments based on their weight in the IMRT plan. The reasoning behind this is that if a segment has low weight in the IMRT plan, then there are other segments that are more important to obtain the right fluence modulation. We may therefore choose to disregard segments on this criterion as well. Which criterion that gives the best plans may vary from case to case.

If we assume that dual arcs are used, we can reduce the choices at each control point even more, without disregarding any more segments. Since the control points are positioned at the same angles in both arcs, it is not necessary that all segments are available at both arcs. Thus, there will be fewer segments to choose between at each control point, which means fewer binary variables, but it is still possible to include all interesting segments in the model and get them delivered at the right angle. However, since we are interested in minimizing leaf travel as well, we must keep in mind that
we limit the model by restricting the segments to certain arcs. When we
decide which segments should be available at which arcs, it is important to
remember that which segments are made available at which arc will affect
how short delivery time we can achieve. We should therefore try to develop
some method for deciding which segments that should be made available at
which arc, e.g. by letting the segments whose leaves are positioned more to
the left be sorted into one arc and the segments whose leaves are positioned
more to the right be sorted into the other arc.

4.3 Implementation

The fluence maps and some of the segments that will be used are generated
using RayStation. The fluence maps are generated for 36 equidistant beams
along the arc and the optimization process is run for 10 iterations. When
the fluence maps are generated, no segmentation is done in the optimization
process. The fluence maps are then exported from RayStation with the aid of
a python-script, that stores the data in text files.

When the fluence maps have been generated, we start a new optimization
process for the 36 beams in order to generate the segments. Again, we let the
optimization process run for 10 iterations, but this time, segmentation is car-
ried out in the last iteration. The segments are also exported from RayStation
to text files using a python-script.

The implementation of the model is done in MATLAB (MathWorks, Natick,
MA, USA). Thus, the fluence maps will be represented by matrices. The size
of the matrices depends on the size of the treatment field and will therefore
vary with different cases. For each case, we will have 36 matrices of equal
size that represents the fluence maps of the 36 IMRT beams that we want to
mimic. Each matrix element corresponds to a bixel of size 0.5 cm × 0.5 cm.

The fluence matrices will be converted into integer matrices with ten lev-
els in order to reduce computational complexity. The conversion works as
follows: the “height” of a step is computed as the maximum MU a bixel in
any of the 36 fluence matrices should receive divided by 9.5. Each matrix
element is then rounded to the nearest multiple of a step. Thus, all matrices
will be converted according to the same step height and since the maximum
number of MU each fluence matrix is supposed to receive varies, not all con-
verted matrices will have an element of value 10.

The model will be run both with and without constraints on leaf travel. When
no constraints on leaf travel are enforced, the objective function will
simply be the sum of the differences between the fluence matrices and the
delivered matrices. When leaf travel constraints are used, the objective func-
tion will be the sum of the function that measures the difference between the
matrices and the function that measures the distance the leaves travel. The
two functions will be given equal weights. The choice of using equal weights
is somewhat arbitrary. It is necessary to include the distance variables in the objective function, as explained in Section 3.1. However, the leaf travel is limited by the distance constraints, so the appearance of the distance variables in the objective function is not that important to constrain the leaf motion. In the cases tried in this project, the value of the function that measures the similarity in fluence has been at least five times as large as the value of the distance function, so weighting the two functions equally will keep focus on the similarity in fluence.

In this project, the model has only been implemented for dual arc delivery. The number of segments that will be made available for each control point will differ from case to case. If there are more segments generated for a beam than made available in the model, then the segments to use will be chosen either according to their weight in the IMRT plan or according to their size, depending on what has been found best for the considered case. The segments will be sorted according to the sum of the positions of their left leaves in the following way: assume that $n$ segments will be made available for each arc. Then the $n$ segments with the lowest sum will be sorted into the first arc and the $n$ segments with the highest sum will be sorted into the second arc. Note that this sorting only takes place when there are more than $n$ segments generated for a beam. Otherwise, all segments will be made available for both arcs.

Since the model is a mixed binary linear programming model, we need sophisticated optimization software to solve the model. In this project, CPLEX (IBM, Armonk, NY, USA) is the optimization software used. CPLEX is called from MATLAB using the modelling language YALMIP [7], in order to avoid reformulating the model on standard form.

Since mixed binary programming problems are NP-complete, we do not solve the model to optimality. CPLEX uses a branch-and-bound method for solving integer programming problems, which makes it easy to formulate a stopping criterion based on how far from optimality we allow our objective value to be. Choosing tolerance levels for the objective function value is a trade-off between goodness of solution and time required to produce a solution.

4.4 The Importance of the Segments Used

From the way that the model is formulated, as a choice between different predefined segments, it is obvious that the segments we are allowed to choose between play a major role in the quality of the solution of the model. To obtain a good solution, it is therefore very important that the predefined segments are shaped in such a way that it is possible to use them to obtain fluence modulation similar to the desired fluence modulation and deliver them at adjacent angles without too much leaf travel.
It is hard to determine the quality of the proposed conversion algorithm without evaluating the results obtained with different segments to choose between. We will expect that the solution we obtain with segments not at all suited for delivery of the desired fluence maps will give poor results, but how good must the segments be in order to obtain a good solution? Moreover, what is meant by a “good” set of segments and a “good” solution?

In order to make a better evaluation of the model, it will be implemented using three different kinds of segments. The first kind of segments used will be the segments from the 36 beam IMRT plan. It should be noted that the IMRT segments are modified slightly before they are used in the model. The modification consists of moving closed leaves, or rather, leaves whose gap is smaller than 0.5 cm, to a more suitable position for VMAT delivery. This modification does not change the shape of the segments, but it decreases the leaf travel for closed leaves. Since these segments are generated to deliver the fluence specified in the fluence maps we want to mimic, it will be possible to use these segments to obtain fluence modulation similar to the desired fluence modulation. However, the leaf travel required to move between these segments is not considered when they are created and we must therefore assume that quite a large amount of leaf travel will be necessary in order to deliver these segments at adjacent angles. Thus, we would like to try solving the model using segments that require less leaf motion when moving from one segment to another.

In order to accomplish this, the second kind of segments we will use will be generated as follows: we start by choosing a number of segments from the segments generated for the IMRT plan. For the chosen segments, we generate intermediate segments between every pair of IMRT segments. We will use an implementation where two intermediate segments are generated between every pair of IMRT segments. The intermediate segments will be generated by linear interpolation of the leaf positions and creating two equidistantly spaced segments between each pair of IMRT segments. The introduction of intermediate segments largely increases the number of segments generated for each fluence map, and thereby the complexity of the problem. For computational reasons, it will therefore not be possible to use more than two or three IMRT segments when generating intermediate segments. This will give us either 4 or 9 segments per arc in total. The intermediate segments will not be optimized for delivery of the desired fluence modulation. Therefore, we do not expect this kind of segments to be able to achieve as good similarity in fluence as by only using IMRT segments. In return, we expect that a lower amount of leaf travel will be required between adjacent control points when this kind of segments is used.

The third kind of segments we will try are segments generated in a DMLC-fashion. The reason for using such segments is that DMLC segments are generated to deliver the desired fluence map by letting the MLC leaves sweep from one side to the other. Thus, the DMLC segments should be well suited for delivery of fluence modulation similar to the desired fluence modulation.
and require a small amount of leaf motion between adjacent segments. Since there is currently no DMLC converter available in RayStation, these segments will be generated from scratch. Thus, when the DMLC segments are used, the only information we get from the IMRT plan is the fluence maps.

The algorithm for generating the DMLC segments is quite simple. It is assumed that the fluence matrix is converted to a corresponding integer matrix. For each row of the matrix, the openings needed to deliver the fluence profile of that particular row is generated in the following way: all bixel edges where the fluence increases are found and stored as “up”-moves. Similarly, all bixel edges where the fluence decreases are found and stored as “down”-moves. The openings are now generated by moving from the left to the right and pairing each “up”-move with the next “down”-move. The “up”-move gives the position of the left leaf and the “down”-move gives the position of the right leaf. To generate the openings that span all the rows, the openings for each row is simply put together. In order to do this, we need the same number of openings for each leaf pair. If we just use the above algorithm, it is unlikely that we will end up with the same number of openings for each row. If we lack openings for certain rows, we will introduce so-called Steiner points. These points are simply bixel edges where we introduce additional “up”-moves and “down”-moves in order to generate extra openings. This is the basic algorithm used for generating the DMLC segments. Figure 4.1 shows which openings are generated to deliver a fluence profile when no Steiner points are used and Figure 4.2 shows which openings are generated to deliver the same fluence profile when a Steiner point is introduced.

When sharp peaks occur in the fluence maps, the DMLC segments generated to deliver the fluence profile tend to be narrow and a large amount of segments need to be used in order to deliver the fluence. This is undesirable.
26 4 Solving the Model

(a) A fluence profile to be delivered. A Steiner-point is introduced at \( x = 3 \). The up-moves and down-moves are shown.

(b) The openings generated to deliver the fluence profile with the introduced Steiner point. Blocks of the same color constitute an opening.

**Fig. 4.2** DMLC openings generated for a fluence profile with an introduced Steiner point.

for our method, since no more than 8-10 segments can be used to approximate each fluence map. Thus, we would like to smooth out the fluence maps. We will do this by convolution of the fluence maps in the x-direction. The kernel that will be used is derived from the Gaussian function \( e^{-x^2} \) rounded to work over five matrix elements. It becomes \([0.042, 0.25, 0.417, 0.25, 0.042]\).

All implementation details are now sorted out, and we may move on to show the results.
Chapter 5
Results

In order to evaluate the model, we have used it to create VMAT plans for three different clinical cases, one case of prostate cancer, one case of cancer in the head-neck region and one case of pelvic cancer. As mentioned in section 4.4, the model has been implemented with three different kinds of segments for each case.

Both the results from the optimization, i.e. similarity of fluence and leaf travel distance, will be presented, as well as results for the dose that the plan gives rise to.

The results of the optimization of the model for the different kinds of segments will be displayed in tables. Here follows an explanation of how the table entries should be interpreted. “Segments per arc” denotes the number of segments made available from each beam at each arc. “Distance constraint” denotes the constraint enforced on leaf travel distance in the model. The leaf travel distance is given in units of matrix elements. Since one matrix element corresponds to a bixel of width 0.5 cm, we just need to divide the distance given by two in order to get the leaf travel distance in cm. In the solutions presented here, the only distance constraint that is used is an upper bound on the total leaf travel distance. “Optimality limit” gives the optimality limit in %. It is the stopping criterion used in the optimization. An optimality limit of $k$ % means that the objective value of the solution is at most $k$ % larger than the lower bound of the objective value. “Objective fluence” is the value of the part of the objective function that measures the difference between the fluence of the created plan and the desired fluence, i.e. the sum of the absolute value of the difference of the matrices. The value given in the parenthesis is the objective value divided by the sum of the matrix elements in the fluence matrices. “Distance” denotes the total leaf travel distance in the plan, again given in units of matrix elements. “Computational time” denotes the time needed to solve the model in the specified case. It is given in seconds.

The dose results for the corresponding solutions that will be shown are the values of the objective functions for the considered ROIs. Note that the objective values depend on the weight given to that specific objective. For
readability, not all objectives will be shown in the tables. They will however be included in the corresponding dose volume histogram (DVH) plots. A dose volume histogram displays volume percentage of an ROI as a function of received dose level.

There are some differences in how the model is run for the different patient cases. Different numbers of segments from each beam for each arc have been made available in different cases, and the optimality limit also differs, both between different cases and depending on the number of segments that have been used. The reason for this is simply that we want to obtain results within a reasonable time frame. Since the complexity differs among the different cases, so will the number of segments we can work with and the optimality limit.

In order to be able to evaluate the goodness of the optimization solutions, the best possible optimization solutions for the different cases and kinds of segments have also been computed. The best possible optimization solutions for the different kinds of segments are the results we get when we, for each fluence map, optimize the similarity between desired and delivered fluence when all segments associated with that fluence map is allowed to be used and the only constraint on the segment weights is that they need to be non-negative. These results have been computed for two different definitions of “associated segments”. First, we consider the associated segments to be the segments generated specifically for that fluence map. In the second case, we consider the associated segments to also include the segments of the previous fluence map and the subsequent fluence map. The reason for evaluating the second notion is that the segments associated with the previous and the subsequent fluence map will be available at control points that influence the fluence map itself. Results have been computed both for the case where all generated segments have been available and when we have limited the number of available segments, which is the case for most of the models solved.

We have also generated dose results in order to be able to evaluate the dose results for each case. Results for a 36 beam IMRT plan and for a VMAT plan has been generated in RayStation. Since the plans generated using the proposed method only optimizes on fluence, the most reasonable plans for comparison are plans where no optimization is carried out after the segmentation. The 36 beam IMRT plans considered are optimized for 10 iterations and the segmentation takes place after the ninth iteration. The reason for this is that this IMRT plan corresponds to the plan that has been used in the model. The VMAT plan is run for 7 iterations, and the segmentation takes place during the seventh iteration. The reason is that this is the normal procedure when VMAT plans are generated. Of course, then further optimization is carried out after segmentation, but this is not of interest for comparison.
5.1 Results for the Prostate Case

In the prostate case, we have two ROIs that we want to irradiate, namely the prostate and its PTV. We also have three OARs; the rectum, the bladder and external tissue. Both the prostate and the PTV are assigned two objective functions; a minimum dose of 81 Gy and a maximum dose of 83 Gy. All of these objectives have weight 100. Both the rectum and the bladder are given objectives expressed in terms of maximum EUD. Both these organs have the objective that the maximum EUD should be 60 Gy, with weight 20. The objective of the external tissue is a dose fall-off function with high level 80 Gy, low level 0 Gy and a low dose distance of 8 cm. The weight of this objective function is 40. The different objective functions are described in Section 1.3.

The fluence matrices exported from RayStation are represented as $18 \times 26$-matrices in MATLAB. The maximum MU for the fluence matrices varies between 8.76 and 16.61. The sum of all the entries in the converted fluence matrices is 33 224. This number is interesting when we want to evaluate the similarity between the desired fluence matrices and the fluence matrices we obtain in the model. Since the objective function is the sum of the difference in absolute value in all the matrix elements, we need to compare this number to the sum of all the matrix elements in order for the objective value to be meaningful. This number also gives an indication of the computational complexity of this specific case.

We will not display the objective value associated with the external tissue and the bladder in the dose result tables. However, the dose distribution of these OARs will be included in the DVHs.

For this case, there are at most 7 IMRT segments generated for any beam. Thus, in most of the considered cases we do not need to disregard any segments. When we do need to disregard segments, the choice of which segments to use are based on the weight of the segments in the IMRT plan. Of course, this does not apply to the DMLC segments. In that case, the choice is based on the size of the segments.

When solving the model using segments from the IMRT plan, three different cases have been considered. One where no constraints on leaf travel have been enforced and two cases with different constraints on leaf travel. The results are shown in Table 5.1. The corresponding DVHs are shown in Figures 5.1, 5.2 and 5.3. Figure 5.1 shows the DVHs for the plans with no distance constraint, Figure 5.2 shows the DVHs for the plans with maximum distance 1000 and Figure 5.3 shows the DVHs for the plans with maximum distance 1500.

When solving the model using both IMRT segments and intermediate segments, two different levels for the maximum distance have been used. The results are shown in Table 5.2. The corresponding DVHs are shown in Figures 5.4 and 5.5. Figure 5.4 shows the DVHs when 4 segments per arc are used and Figure 5.5 shows the DVHs when 9 segments per arc are used.
Table 5.1: Results of the optimization and the corresponding dose objective values for the prostate case when flattened DML segments are used.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Recum</th>
<th>Distance (cm)</th>
<th>Objective (Optimally)</th>
<th>Objective Distance (cm)</th>
<th>Fluence (Optimally)</th>
<th>Dose (cm²)</th>
<th>Time [s]</th>
<th>Segments</th>
<th>Fluence (min)</th>
<th>Dose (min)</th>
<th>Time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 5.2: Results of the optimization and the corresponding dose objective values for the prostate case when IMRT segments are used.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Recum</th>
<th>Distance (cm)</th>
<th>Objective (Optimally)</th>
<th>Objective Distance (cm)</th>
<th>Fluence (Optimally)</th>
<th>Dose (cm²)</th>
<th>Time [s]</th>
<th>Segments</th>
<th>Fluence (min)</th>
<th>Dose (min)</th>
<th>Time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0.91</td>
<td>0.92</td>
<td>0.90</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
</tbody>
</table>
5.1 Results for the Prostate Case

Fig. 5.1 DVHs for the ROIs of the prostate case when IMRT segments are used with no distance constraints. The solid lines show the solution with optimality limit 30 %, the dashed lines show the solution with optimality limit 20 % and the dotted lines show the solution with optimality limit 15 %.

Fig. 5.2 DVHs for the ROIs of the prostate case when IMRT segments are used with maximum distance 1000. The solid lines show the solution with optimality limit 100 % and the dashed lines show the solution with optimality limit 50 %.
Fig. 5.3 DVHs for the ROIs of the prostate case when IMRT segments are used with maximum distance 1500. The solid lines show the solution with optimality limit 70 %, the dashed lines show the solution with optimality limit 50 % and the dotted lines show the solution with optimality limit 40 %.

Solving the model using DMLC segments has been tried with different numbers of segments made available per arc and different leaf travel constraints. The results are shown in Table 5.3. The corresponding DVHs are shown in Figures 5.6 and 5.7. Figure 5.6 shows the DVHs for the case when 4 segments per arc are used and Figure 5.7 shows the DVHs when 5 or 6 segments per arc are used.

For comparison, the best possible results in fluence domain for the different kinds of segments are shown in Table 5.4. Moreover, Table 5.5 shows the values of the objective functions for the regions of interest for an IMRT plan with 36 beams generated in RayStation right after segmentation and for a VMAT plan generated in RayStation right after segmentation. The top row shows the results for the IMRT plan and the bottom row shows the results for the VMAT plan. The corresponding DVHs are shown in Figure 5.8.
5.1 Results for the Prostate Case

Fig. 5.4 DVHs for the ROIs of the prostate case when IMRT segments and intermediate segments are used. 4 segments per arc are used. The solid lines show the solution with optimality limit 40 % and maximum distance 1000, the dashed lines show the solution with optimality limit 50 % and maximum distance 1500, and the dotted lines show the solution with optimality limit 40 % and maximum distance 1500 and the starred lines show the solution with optimality limit 35 % and maximum distance 1500.

<table>
<thead>
<tr>
<th>Objective</th>
<th>max PTV</th>
<th>min PTV</th>
<th>max Prostate</th>
<th>min Prostate</th>
<th>Rectum</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMRT all</td>
<td>0.1657</td>
<td>0.0532</td>
<td>0.0316</td>
<td>0.0011</td>
<td>0.0022</td>
</tr>
<tr>
<td>Intermediate 4 seg/arc</td>
<td>0.3126</td>
<td>0.0561</td>
<td>0.0306</td>
<td>0.0011</td>
<td>0.0022</td>
</tr>
<tr>
<td>Intermediate 9 seg/arc</td>
<td>0.0532</td>
<td>0.0618</td>
<td>0.0306</td>
<td>0.0011</td>
<td>0.0022</td>
</tr>
<tr>
<td>DMLC all</td>
<td>0.1657</td>
<td>0.0532</td>
<td>0.0316</td>
<td>0.0011</td>
<td>0.0022</td>
</tr>
<tr>
<td>DMLC 4 seg/arc</td>
<td>0.3126</td>
<td>0.0561</td>
<td>0.0306</td>
<td>0.0011</td>
<td>0.0022</td>
</tr>
<tr>
<td>DMLC 5 seg/arc</td>
<td>0.1657</td>
<td>0.0532</td>
<td>0.0316</td>
<td>0.0011</td>
<td>0.0022</td>
</tr>
</tbody>
</table>

Table 5.4 Value of the best possible objective functions for the different kinds of segments in the prostate case.

Table 5.5 The values of the objective functions for the different regions of interest for a 36 beam IMRT plan directly after segmentation (top row) and a VMAT plan directly after segmentation (bottom row).
Fig. 5.5 DVHs for the ROIs of the prostate case when IMRT segments and intermediate segments are used. 9 segments per arc are used. The solid lines show the solution with optimality limit 50% and maximum distance 1000, the dashed lines show the solution with optimality limit 55% and maximum distance 1500, the dotted lines show the solution with optimality limit 50% and maximum distance 1500 and the starred lines show the solution with optimality limit 45% and maximum distance 1500.

5.2 Results for the Head-and-Neck Case

The head-and-neck case is a bit more complicated than the prostate case. Here, we have five CTVs. One for the primary tumor, two at the right side of the neck and two at the left side of the neck. Moreover, we have five OARs that we want to spare, namely the spinal cord, the right and left parotid gland, the brain stem and external tissue. The objective of the primary CTV is uniform dose of 66 Gy, the objective of the upper CTVs in the side of the neck is a uniform dose of 55 Gy and the objective of the lower CTVs in the side of the neck is a uniform dose of 50 Gy. The objectives of all the CTVs have a weight of 100. The brain stem and the spinal cord both have the same objective, namely a maximum dose of 30 Gy with weight 10. Both of the parotid glands have the same objective, namely a maximum EUD of 25 Gy. The weight of these objectives is 10. The objective of the tissue is formulated in terms of dose fall-off with a high value of 50 Gy and a low value of 15 Gy. The low dose distance is 1.5 cm. The objective functions and the terminology is described in Section 1.3.
5.2 Results for the Head-and-Neck Case

Fig. 5.6 DVHs for the ROIs of the prostate case when DMLC segments are used. 4 segments per arc are used. The solid lines show the solution with optimality limit 70% and maximum distance 1000, the dashed lines show the solution with optimality limit 50% and maximum distance 1000, the dotted lines show the solution with optimality limit 70% and maximum distance 1500 and the starred lines show the solution with optimality limit 60% and maximum distance 1500.

The size of the fluence matrices for this case is $40 \times 51$ and the sum of the entries of the converted fluence matrices is 55,804. The maximum MU of the fluence matrices ranges from 10.84 to 22.14.

In the head-and-neck case, we have removed the objectives of the tissue, the parotid glands and the right and left lower CTVs from the dose result tables. However, their dose distributions are still displayed in the DVH plots.

For this case, when not all generated segments are made available, the choice of which segments to make available is based on the size of the segments.

When IMRT segments are used, the model has been solved both with and without distance constraints. The model has also been solved using 4, 5 and 6 segments per arc. The results are shown in Table 5.6. The corresponding DVHs are shown in Figures 5.9 and 5.10. Figure 5.9 shows the DVHs for the solutions with 4 segments per arc and Figure 5.10 shows the solutions with 5 and 6 segments per arc.

When IMRT segments are used together with generated intermediate segments, the model has been solved using both 4 and 9 segments per arc. Two levels of maximum distance have been used. The results are shown in Table 5.7. The corresponding DVHs are shown in Figure 5.11.
Fig. 5.7 DVHs for the ROIs of the prostate case when DMLC segments are used. The maximum distance is 1500. The solid lines show the solution with optimality limit 70% and 5 segments/arc and the dashed lines show the solution with optimality limit 90% and 6 segments/arc.

Fig. 5.8 DVHs for the ROIs of the prostate case for a 36 beam IMRT plan after segmentation (solid lines) and a VMAT plan after segmentation (dashed lines).
### Table 5.6
Results of the optimization and the corresponding dose objective values for the head-and-neck case when IMRT segments are used.

<table>
<thead>
<tr>
<th>Segments per arc</th>
<th>Distance constraint</th>
<th>Optimality limit [%]</th>
<th>Objective fluence</th>
<th>Distance</th>
<th>Computational time [s]</th>
<th>Objective dose</th>
<th>CTV primary</th>
<th>CTV upper right</th>
<th>CTV upper left</th>
<th>Brain stem</th>
<th>Spinal cord</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>none</td>
<td>50</td>
<td>18716 (0.3354)</td>
<td>2487</td>
<td>219</td>
<td>10.5458</td>
<td>2.8984</td>
<td>1.8788</td>
<td>1.8072</td>
<td>1.9 - 10^-5</td>
<td>0.0966</td>
</tr>
<tr>
<td>4</td>
<td>none</td>
<td>40</td>
<td>18193 (0.326)</td>
<td>2551</td>
<td>297</td>
<td>11.5598</td>
<td>3.5979</td>
<td>1.6317</td>
<td>1.4503</td>
<td>0.1 - 10^-5</td>
<td>0.0416</td>
</tr>
<tr>
<td>5</td>
<td>none</td>
<td>50</td>
<td>20173 (0.3615)</td>
<td>2487</td>
<td>277</td>
<td>24.7444</td>
<td>6.8052</td>
<td>5.6289</td>
<td>4.894</td>
<td>1.9 - 10^-4</td>
<td>1.0956</td>
</tr>
<tr>
<td>5</td>
<td>none</td>
<td>40</td>
<td>17146 (0.3073)</td>
<td>2759</td>
<td>1249</td>
<td>21.7502</td>
<td>2.3219</td>
<td>5.0021</td>
<td>4.0615</td>
<td>0.0298</td>
<td>2.1303</td>
</tr>
<tr>
<td>6</td>
<td>none</td>
<td>40</td>
<td>19400 (0.3476)</td>
<td>2848</td>
<td>586</td>
<td>26.7401</td>
<td>6.3462</td>
<td>6.6493</td>
<td>5.1006</td>
<td>0.0277</td>
<td>2.2445</td>
</tr>
<tr>
<td>4</td>
<td>max 1500</td>
<td>50</td>
<td>20061 (0.3595)</td>
<td>1399</td>
<td>1303</td>
<td>9.6049</td>
<td>2.8881</td>
<td>1.0846</td>
<td>0.837</td>
<td>1.8 - 10^-4</td>
<td>1.8 - 10^-4</td>
</tr>
</tbody>
</table>

### Table 5.7
Results of the optimization and the corresponding dose objective values for the head-and-neck case when IMRT segments are used together with generated intermediate segments.

<table>
<thead>
<tr>
<th>Segments per arc</th>
<th>Distance constraint</th>
<th>Optimality limit [%]</th>
<th>Objective fluence</th>
<th>Distance</th>
<th>Computational time [s]</th>
<th>Objective dose</th>
<th>CTV primary</th>
<th>CTV upper right</th>
<th>CTV upper left</th>
<th>Brain stem</th>
<th>Spinal cord</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>max 1500</td>
<td>30</td>
<td>19132 (0.3428)</td>
<td>1265</td>
<td>135</td>
<td>8.9793</td>
<td>0.9059</td>
<td>1.9589</td>
<td>1.4438</td>
<td>0.0022</td>
<td>0.0066</td>
</tr>
<tr>
<td>9</td>
<td>max 1500</td>
<td>50</td>
<td>19250 (0.345)</td>
<td>1471</td>
<td>2457</td>
<td>6.6542</td>
<td>1.176</td>
<td>0.8579</td>
<td>0.6439</td>
<td>0.0089</td>
<td>0.0017</td>
</tr>
<tr>
<td>9</td>
<td>max 2000</td>
<td>40</td>
<td>18718 (0.3354)</td>
<td>1682</td>
<td>754</td>
<td>6.3105</td>
<td>0.4324</td>
<td>1.3827</td>
<td>1.0032</td>
<td>0.0042</td>
<td>0.0095</td>
</tr>
</tbody>
</table>

### Table 5.8
Results of the optimization and the corresponding dose objective values for the head-and-neck case when generated DMLC segments are used.

<table>
<thead>
<tr>
<th>Segments per arc</th>
<th>Distance constraint</th>
<th>Optimality limit [%]</th>
<th>Objective fluence</th>
<th>Distance</th>
<th>Computational time [s]</th>
<th>Objective dose</th>
<th>CTV primary</th>
<th>CTV upper right</th>
<th>CTV upper left</th>
<th>Brain stem</th>
<th>Spinal cord</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>max 1500</td>
<td>50</td>
<td>25895 (0.464)</td>
<td>1466</td>
<td>209</td>
<td>8.8085</td>
<td>1.4309</td>
<td>1.4095</td>
<td>2.3608</td>
<td>4.0 - 10^-6</td>
<td>0.0027</td>
</tr>
<tr>
<td>4</td>
<td>max 2000</td>
<td>40</td>
<td>22236 (0.3985)</td>
<td>1981</td>
<td>518</td>
<td>13.2702</td>
<td>2.1116</td>
<td>1.6618</td>
<td>2.2546</td>
<td>0.0073</td>
<td>0.0518</td>
</tr>
<tr>
<td>5</td>
<td>max 1500</td>
<td>70</td>
<td>27918 (0.3003)</td>
<td>1497.2</td>
<td>4806</td>
<td>22.7194</td>
<td>6.311</td>
<td>3.276</td>
<td>5.2903</td>
<td>0</td>
<td>8.0 - 10^-5</td>
</tr>
<tr>
<td>5</td>
<td>max 2000</td>
<td>70</td>
<td>26112 (0.4679)</td>
<td>1898.4</td>
<td>573</td>
<td>9.9717</td>
<td>0.5057</td>
<td>2.4025</td>
<td>4.3542</td>
<td>2.5 - 10^-6</td>
<td>0.0018</td>
</tr>
</tbody>
</table>
With DMLC segments, we have solved the model using both 4 and 5 segments and using two different levels of maximum distance. The results are shown in Table 5.8. The corresponding DVHs are shown in Figure 5.12.

For comparison, we show the best possible results for fluence optimization with the different kinds of segments in Table 5.9. We also include the values of the objective functions for the different regions of interest for a 36 beam IMRT plan generated in RayStation right after segmentation and for a VMAT plan generated in RayStation right after segmentation. The results are shown in Table 5.10. The top row shows the results for the 36 beam IMRT plan and the bottom row shows the results for the VMAT plan. The corresponding DVHs are shown in Figure 5.13.
5.2 Results for the Head-and-Neck Case

Fig. 5.10 DVHs for the ROIs of the head-and-neck case when IMRT segments are used without distance constraint. The solid lines show the solution with optimality limit 50% and 5 segments/arc, the dashed lines show the solution with optimality limit 40% and 5 segments/arc and the dotted lines show the solution with optimality limit 40% and 6 segments/arc.

Table 5.9 Value of the best possible fluence objective functions for the different kinds of segments in the head-and-neck case.

<table>
<thead>
<tr>
<th></th>
<th>Fluence map only</th>
<th>Fluence map and adjacent</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMRT all</td>
<td>9262</td>
<td>9078</td>
</tr>
<tr>
<td>IMRT 4 seg/arc</td>
<td>10425</td>
<td>1016.1</td>
</tr>
<tr>
<td>IMRT 5 seg/arc</td>
<td>9434</td>
<td>9226</td>
</tr>
<tr>
<td>Intermediate 4 seg/arc</td>
<td>17337</td>
<td>15650</td>
</tr>
<tr>
<td>Intermediate 9 seg/arc</td>
<td>13216</td>
<td>12587</td>
</tr>
<tr>
<td>DMLC all</td>
<td>6091.9</td>
<td>6091.9</td>
</tr>
<tr>
<td>DMLC 4 seg/arc</td>
<td>14284</td>
<td>12836</td>
</tr>
<tr>
<td>DMLC 5 seg/arc</td>
<td>9911.8</td>
<td>9495.3</td>
</tr>
<tr>
<td>DMLC 6 seg/arc</td>
<td>7960.6</td>
<td>7217.4</td>
</tr>
</tbody>
</table>
5.3 Results for the Pelvic Case

The pelvic case consists of three PTVs and three OARS. The first PTV is in the vaginal wall with the objective of a minimum dose of 56 Gy and weight 25. The second PTV covers the lymph node. The objective of this is also a minimum dose of 56 Gy with weight 25. The third PTV covers both the other PTVs and their surroundings. This PTV has two objectives; a minimum dose of 50.4 Gy with weight 50 and a maximum DVH of 52.5 Gy to 5% of the volume with weight 1. The OARs are the bladder, the rectum and
5.3 Results for the Pelvic Case

Fig. 5.12 DVHs for the ROIs of the head-and-neck case when DMLC segments are used. The solid lines show the solution with optimality limit 50 %, 4 segments/arc and maximum distance 1500, the dashed lines show the solution with optimality limit 40 %, 4 segments/arc and maximum distance 2000, the dotted lines show the solution with optimality limit 70 %, 5 segments/arc and maximum distance 1500 and the starred lines show the solution with optimality limit 70 %, 5 segments/arc and maximum distance 2000.

Fig. 5.13 DVHs for the ROIs of the head-and-neck case for a 36 beam IMRT plan after segmentation (solid lines) and a VMAT plan after segmentation (dashed lines).
external tissue. The bladder has two objectives; a maximum dose of 50 Gy and a maximum EUD of 35 Gy, both with weight 1. The rectum also has two objectives; a maximum dose of 50 Gy and a maximum EUD of 30 Gy, both with weight 1. The external tissue has a dose fall-off objective with high level 30 Gy, low level 10 Gy and a low dose distance of 5 cm. The weight of this objective is 1.

The size of the fluence matrices for this case are $45 \times 45$. The maximum MU of the fluence matrices ranges from 13.06 to 30.32. The sum of all the elements in the integer fluence matrices is 83,088.

In the pelvic case, the objective values of the external tissue, the bladder and the rectum are not included in the tables of the dose result. However, they are included in the DVH plots.

For this case, when less IMRT segments than generated are made available in the model, the choice of which segments that should be made available is based on the weight of the segments in the IMRT plan. This does not apply to the case when the DMLC segments are used. In that case, the choice of which segments to use is based on the size of the segments.

When we solve the model using IMRT segments, the model has been solved both with and without distance constraints. The results are shown in Table 5.11. The corresponding DVHs are shown in Figures 5.14 and 5.15. Figure 5.14 shows the DVHs for the solutions with no distance constraints and Figure 5.15 shows the DVHs for the solutions with maximum distance 2500.

With IMRT segments together with generated intermediate segments, the model has been solved using both 4 and 9 segments per arc. Two levels of maximum distance has been used. The results are shown in Table 5.12. The corresponding DVHs are shown in Figure 5.16.

When DMLC segments have been used, the model has been solved using 4 segments per arc and two different levels of maximum distance have been used. The results are shown in Table 5.13. The corresponding DVHs are shown in Figure 5.17.

For comparison, the best possible optimization results for the different kinds of segments have been generated. They are shown in Table 5.14. Dose results have also been generated for comparison. Table 5.15 shows the values of the objectives for a 36 beam IMRT plan directly after segmentation (top row) and a VMAT plan directly after segmentation (bottom row). The corresponding DVHs are shown in Figure 5.18.
### Table 5.11 Results of the optimization and the corresponding dose objective values for the pelvic case when IMRT segments are used.

<table>
<thead>
<tr>
<th>Segments per arc</th>
<th>Distance constraint</th>
<th>Optimality limit [%]</th>
<th>Objective fluence</th>
<th>Distance</th>
<th>Computational time [s]</th>
<th>Objective dose</th>
<th>PTV Vag Wall</th>
<th>PTV node</th>
<th>min PTV large</th>
<th>max DVH PTV large</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>none</td>
<td>50</td>
<td>21490 (0.2586)</td>
<td>3547</td>
<td>589</td>
<td>0.5513</td>
<td>0.0034</td>
<td>0</td>
<td>0.4159</td>
<td>0.0081</td>
</tr>
<tr>
<td>5</td>
<td>none</td>
<td>50</td>
<td>23991 (0.2887)</td>
<td>3387</td>
<td>785</td>
<td>1.1136</td>
<td>0</td>
<td>0.0071</td>
<td>0.9425</td>
<td>0.0054</td>
</tr>
<tr>
<td>6</td>
<td>none</td>
<td>50</td>
<td>23037 (0.2773)</td>
<td>3594</td>
<td>972</td>
<td>1.1056</td>
<td>0</td>
<td>0.0024</td>
<td>0.9085</td>
<td>0.0082</td>
</tr>
<tr>
<td>4</td>
<td>max 2500</td>
<td>50</td>
<td>25813 (0.3107)</td>
<td>2285</td>
<td>3418</td>
<td>0.425</td>
<td>0.0295</td>
<td>0.0013</td>
<td>0.3201</td>
<td>0.0026</td>
</tr>
<tr>
<td>5</td>
<td>max 2500</td>
<td>50</td>
<td>24575 (0.2955)</td>
<td>2500</td>
<td>4859</td>
<td>0.2922</td>
<td>0</td>
<td>0.0013</td>
<td>0.1909</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

### Table 5.12 Results of the optimization and the corresponding dose objective values for the pelvic case when IMRT segments and generated intermediate segments are used.

<table>
<thead>
<tr>
<th>Segments per arc</th>
<th>Distance constraint</th>
<th>Optimality limit [%]</th>
<th>Objective fluence</th>
<th>Distance</th>
<th>Computational time [s]</th>
<th>Objective dose</th>
<th>PTV Vag Wall</th>
<th>PTV node</th>
<th>min PTV large</th>
<th>max DVH PTV large</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>max 2000</td>
<td>50</td>
<td>28822 (0.3469)</td>
<td>1728.5</td>
<td>1226</td>
<td>0.6471</td>
<td>0.0369</td>
<td>6.9 · 10⁻⁴</td>
<td>0.5594</td>
<td>3.0 · 10⁻⁴</td>
</tr>
<tr>
<td>4</td>
<td>max 2500</td>
<td>50</td>
<td>28857 (0.3473)</td>
<td>2214</td>
<td>301</td>
<td>1.1978</td>
<td>9.0 · 10⁻⁴</td>
<td>0.0022</td>
<td>1.1249</td>
<td>0.0019</td>
</tr>
<tr>
<td>9</td>
<td>max 2000</td>
<td>50</td>
<td>29475 (0.3547)</td>
<td>1932</td>
<td>4588</td>
<td>0.8026</td>
<td>0.0038</td>
<td>0.003</td>
<td>0.6996</td>
<td>0.0013</td>
</tr>
<tr>
<td>9</td>
<td>max 2500</td>
<td>70</td>
<td>31092 (0.3742)</td>
<td>1869</td>
<td>5476</td>
<td>1.4609</td>
<td>0.015</td>
<td>1.8 · 10⁻⁶</td>
<td>1.3449</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

### Table 5.13 Results of the optimization and the corresponding dose objective values for the pelvic case when generated DMLC segments are used.

<table>
<thead>
<tr>
<th>Segments per arc</th>
<th>Distance constraint</th>
<th>Optimality limit [%]</th>
<th>Objective fluence</th>
<th>Distance</th>
<th>Computational time [s]</th>
<th>Objective dose</th>
<th>PTV Vag Wall</th>
<th>PTV node</th>
<th>min PTV large</th>
<th>max DVH PTV large</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>max 2000</td>
<td>70</td>
<td>41348 (0.4976)</td>
<td>1882.3</td>
<td>3988</td>
<td>5.5295</td>
<td>0.8676</td>
<td>1.7629</td>
<td>2.8832</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>max 2500</td>
<td>60</td>
<td>33406 (0.4021)</td>
<td>2420.1</td>
<td>647</td>
<td>1.0014</td>
<td>0.0278</td>
<td>0.1219</td>
<td>0.896</td>
<td>5.7 · 10⁻⁴</td>
</tr>
</tbody>
</table>
**Fig. 5.14** DVH curves for the ROIs of the pelvic case for solutions using IMRT segments and no distance constraints. The solid lines show the solution with optimality limit 50% and 4 segments/arc, the dashed lines show the solution with optimality limit 50% and 5 segments/arc, and the dotted lines show the solution with optimality limit 50% and 6 segments/arc.

**Fig. 5.15** DVH curves for the ROIs of the pelvic case for solutions using IMRT segments and maximum distance 2500. The solid lines show the solution with optimality limit 50% and 4 segments/arc, and the dashed lines show the solution with optimality limit 50% and 5 segments/arc.
5.3 Results for the Pelvic Case

Fig. 5.16 DVH curves for the ROIs of the pelvic case for solutions using IMRT and intermediate segments. The solid lines show the solution with optimality limit 50 %, 4 segments/arc and maximum distance 2000, the dashed lines show the solution with optimality limit 50 %, 4 segments/arc and maximum distance 2500, the dotted lines show the solution with optimality limit 50 %, 9 segments/arc and maximum distance 2000 and the starred lines show the solution with optimality limit 50 %, 9 segments/arc and maximum distance 2500.

<table>
<thead>
<tr>
<th>Method</th>
<th>Fluence map only</th>
<th>Fluence map and adjacent</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMRT all</td>
<td>14676</td>
<td>14362</td>
</tr>
<tr>
<td>IMRT 4 seg/arc</td>
<td>15873</td>
<td>15285</td>
</tr>
<tr>
<td>IMRT 5 seg/arc</td>
<td>14856</td>
<td>14497</td>
</tr>
<tr>
<td>Intermediate 4 seg/arc</td>
<td>25822</td>
<td>21940</td>
</tr>
<tr>
<td>Intermediate 9 seg/arc</td>
<td>19196</td>
<td>16381</td>
</tr>
<tr>
<td>DMLC all</td>
<td>3750.3</td>
<td>3750.3</td>
</tr>
<tr>
<td>DMLC 4 seg/arc</td>
<td>14024</td>
<td>11771</td>
</tr>
</tbody>
</table>

Table 5.14 Value of the best possible objective functions in fluence domain for the different kinds of segments in the pelvic case.
Fig. 5.17 DVH curves for the ROIs of the pelvic case for solutions using DMLC segments. 4 segments/arc have been used. The solid lines show the solution with optimality limit 70 % and maximum distance 2000 and the dashed lines show the solution with optimality limit 60 % and maximum distance 2500.

<table>
<thead>
<tr>
<th>Objective</th>
<th>PTV Vag Wall</th>
<th>PTV node</th>
<th>min PTV large</th>
<th>max DVH PTV large</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1098</td>
<td>9.3 · 10^{-4}</td>
<td>4.1 · 10^{-6}</td>
<td>0.0103</td>
<td>0.0015</td>
</tr>
<tr>
<td>0.2759</td>
<td>0</td>
<td>0</td>
<td>0.0501</td>
<td>0.0058</td>
</tr>
</tbody>
</table>

Table 5.15 Values of the objective functions for the pelvic case for a 36 beam IMRT plan generated in RayStation right after segmentation (top row) and for a VMAT plan generated in RayStation right after segmentation (bottom row).
5.3 Results for the Pelvic Case

Fig. 5.18 DVH curves for the ROIs of the pelvic case for a 36 beam IMRT plan after segmentation (solid lines) and a VMAT plan after segmentation (dashed lines).
Chapter 6
Discussion

In this chapter, the results shown in the previous chapter will be discussed. We will divide the discussion into one section concerning the results in the fluence domain and one section that concerns the results in the dose domain as well as how the results in fluence domain corresponds to the results in dose domain.

6.1 Results in Fluence Domain

We start the discussion of the results in fluence domain by noting a few things from the results that are not very surprising.

We see that regardless of what kind of segments are used in the model, if the distance constraints are slacked or removed, the value of the objective function decreases. This is expected, since making a constraint less strict will lead to a larger feasible region. Considering that in this case, a larger feasible region means that we are allowed more freedom in choosing the segments, it is quite reasonable that we get better similarity in fluence when we release the distance constraints.

We also note that when more segments are introduced, the computational time increases. This is not surprising either, since each segment has an associated binary variable. Since binary programming is NP-complete, we expect the computational time to increase exponentially with increased input. Thus, substantial increase in computational time when the number of segments made available are increased is expected.

If we compare the computational time for the different cases, we see that the prostate case requires the lowest amount of time to find solutions whereas the pelvic case requires the highest amount of time to find solutions. It is not unexpected that the prostate case is faster than the other two, since the fluence matrices are a lot smaller in the prostate case, which leads to fewer auxiliary objective variables. The matrices of the head-and-neck case...
and the pelvic case are of comparable size, the head-and-neck matrices are even a bit bigger. However, the sum of the elements in the fluence matrix is roughly 50 % larger in the pelvic case, thus making the optimization problem larger and explaining the need for more time to find the solutions.

When the results in fluence domain are compared to the best possible results we can get with the given segments, we see that for the IMRT segments and the IMRT and intermediate segments, the objective value is within 100 % of the best possible value for all three cases. However, for the DMLC segments, the objective value is between 2 and 3 times as large as the best possible value for all cases. The exception is one DMLC solution for the pelvic case, where the objective value is 3.5 times as large as the best possible value. Thus, for the IMRT segments and the IMRT and intermediate segments, the model is quite good at recreating the desired fluence as compared to what we can expect with the given segments. It is a bit surprising that the objective values are so much worse when the DMLC segments are used. The explanation for this might lie in the fact that the generated DMLC segments tend to be quite narrow. Thus, when we are only allowed to use a few DMLC segments to recreate a fluence map, the segments are not big enough to cover the entire region that we want to recreate. Of course, when we have computed the best possible results for DMLC with 4 segments per arc, then we are only allowed to use 8 segments in total, but then we do not need to take into account that we have to share segments between adjacent fluence maps. Also, when we are only allowed to choose a few of the DMLC segments per arc, the distance the leaves have to travel between segments becomes quite large for certain fluence maps. Thus, it can be hard to achieve high similarity in fluence while satisfying the distance constraints.

The fact that the model is able to generate objective values within 100 % of the best possible values for IMRT segments and IMRT and intermediate segments is promising. This leads us to expect that if we give the model segments even better suited for delivery of the desired fluence, then the model will produce solutions with even better similarity in fluence. However, as we saw with the DMLC segments, the model does not perform well with all kinds of segments and it is therefore important to investigate which properties that are required from the segments in order for the model to perform well.

For the IMRT segments and the IMRT and intermediate segments, we also see that we are closer to the best possible objective value for cases where there are less segments available per arc. Since we are allowed to use all available segments to recreate each fluence map when the best possible value is computed and only allowed to use 8 or 10 segments for each fluence map in the model, it is expected that the difference in the values should be larger when there are more than 8 or 10 segments in total. Thus, this behavior is not unexpected.

For almost all cases and with all kinds of segments, we see that we get better results in the optimization when using fewer segments. This is somewhat unintuitive, since a larger number of available segments means a larger fea-
sible region and thus, the optimal solution should be at least as good when more segments are made available. However, we do not solve to optimality, so a larger feasible region does not imply that solutions using more segments are better than solutions using less segments. One might think that when the same optimality limit is used, a larger feasible region should imply at least as good objective value. This is not true, since the optimality limit only ensures that the objective value is at most \( k \% \) larger than the lower bound. In fact, we may be a lot closer to the lower bound, and therefore, the reasoning about larger feasible region does not apply. We may conclude that when fewer segments are used, the branch-and-bound solver tends to find solutions closer to optimality than required by the stopping criterion.

In general, we have not been able to generate results closer to optimality than roughly 50%. This is not a very good result and indicates that the current implementation of the model is not very useful. One problem is that the lower bound in CPLEX is generated using LP relaxation, i.e. the binary variables are relaxed to continuous variables on the interval \([0, 1]\). This means that we are allowed to choose several segments at each control point. Thus, the result we get for the relaxed problem will be far better than anything we can achieve in reality, thereby constituting a poor lower bound.

Another problem with just feeding data into optimization software is that the software is unaware of which variables that are important. In our model, once the binary variables are chosen, the distance variables follow immediately. With both binary variables and weight variables chosen, the remaining variables are locked. However, this is not evident without knowledge of how the optimization problem is constructed. Therefore, optimization software will not be able to explore this.

Attempts to solve the model using an initial guess have been carried out. We have tried to use a solution for a more slack optimality limit to find a solution with a tighter optimality limit, but this has neither given shorter computational time nor better objective value. This might be due to problems regarding implementation, so the use of an initial guess should be further explored.

We have also tried to solve the model using local distance constraints, i.e. using an upper bound for the leaf travel between adjacent control points. However, it turns out that such constraints either become too limiting or of no use at all. The reason for this is that at certain control points, a large amount of leaf travel is required between every segment. In order to not make the problem infeasible, the upper bound must be slack enough to allow for motion between such segments. It turns out that such upper bounds are too loose to make any difference.
6.2 Results in Dose Domain

When we look at the relations between results in fluence domain and results in dose domain, we note that solutions with large distances give poor results in dose domain. Regardless of the case and the segments used, we always find the worst plans in dose domain among the solutions with highest distance. Moreover, the plans generated with more slacked distance constraints are almost always worse than plans generated with tighter distance constraints. There is a simple explanation for this phenomenon, namely the fact that our model only considers what happens at the control points, whereas dose is delivered continuously. Thus, all segments in between the two segments at the control points will be delivered as well, but these segments are not accounted for in the model. If the segments delivered at adjacent control points differ a lot from one another, then the segments not accounted for will differ a lot from the segments in the model.

Thus, large leaf motion between adjacent control points give rise to fluence not accounted for in the model. Since the model is optimized for similarity in fluence, the intermediate fluence not accounted for will almost always degrade the quality of the plan. The total distance is the sum of the distance that the leaf that travels farthest travel between adjacent control points. Thus, a large total distance most often implies large differences between adjacent segments. We see that it does not matter how good the similarity in fluence is if the distance is too large, since such plans will give rise to dose distributions that do not correspond well to the result in fluence domain.

It should be noted that distance constraints that are too tight also give rise to bad plans. See for example Table 5.13. However, in cases where the distance constraints are too tight, it is most often obvious from the fluence objective value that the plans will be of poor quality.

We should note that since the distance only measures how far the leaf that travels farthest travels, a large total distance does not always imply large difference between adjacent segments. For example, although the leaf that travels farthest has a long way to travel, it is not necessary that all the leaves have a long distance to travel. Thus, the great difference may only apply to a single MLC channel, an impact which will be smaller than if lots of leaves travel far. Therefore, it might be interesting to use other distance measures if the most important purpose of minimizing leaf travel is to minimize delivery of fluence not accounted for. For example, we could use the sum of the distance each individual leaf needs to travel between two segments as the distance between the segments. However, if we want to minimize the delivery time of the plan, we need to consider the leaf that travels farthest, since this will be the leaf that limits how fast the gantry can move.

We can observe another aspect of the distance constraint by looking at e.g. Table 5.3. Here, we see that we obtain better dose results for the solution with optimality limit 70 % than the for the solution with optimality limit 50 %. Thus, a solution that is closer to optimality in our model turns out
to be worse in application. In this case, the worse result in dose domain is probably related to the increase in distance for the result closer to optimality. Regardless, it is not good that approaching optimality in our model means degrading the plan quality. One way to avoid this problem is to use distance constraints that are sufficiently tight to make sure that the generated plans are not greatly affected by dose from intermediate segments.

However, adjacent segments that lie too far apart is not the only problem with the generated plans. For example, Table 5.2 shows a solution closer to optimality which is more similar in fluence and requires shorter leaf motion but still gives rise to a plan with worse dose distribution. This could be due to the weighting of the objective functions in the plan. In the fluence maps, there is nothing that indicates which parts of the maps that are most important, i.e. that affects the highest weighted objectives. Thus, if a solution is more similar in fluence than another solution, it might be more similar in less important regions and less similar in more important regions and thereby give rise to a plan with worse dose objective. It would probably be possible to incorporate different penalties for deviation from desired fluence in different parts of the fluence maps in the model, but then we would need to get data on which parts of the fluence maps that correspond to which objective. At the moment, it is not clear how this could be done.

When we compare the values of the dose objectives of the generated plans with the VMAT plans generated in RayStation for the different cases, we see that for the prostate case and the pelvic case, the generated plans are at best comparable to the plan generated in RayStation, whereas for the head-and-neck case, most plans are better than the plan generated in RayStation. One difference between the prostate plan and the pelvic plan compared to the head-and-neck plan is that in the first two cases, there is a single region with objective weight much higher than the rest of the considered region, whereas in the head-and-neck case, we have five different CTVs, all with the same objective weight. Moreover, we see that when the generated plans get a cumulative objective value that is worse than the objective value in the VMAT plan, but not extremely bad, it is most often due to deviation in dose in the region with large objective weight. In the prostate case, it is the over-irradiation of the prostate and its associated PTV, and in the pelvic case, it is under-irradiation of the large PTV. This phenomenon could be due to the reasons discussed above, i.e. the fact that we do not know which parts of the fluence maps that correspond to the regions with high dose objective weight. In the head-and-neck case, we do not have an isolated region with much higher weight than its surroundings, and better similarity in fluence therefore corresponds more closely to better dose distributions. Thus, the ability to include information of dose objective weights in fluence domain could possibly improve the plan quality for cases where we have isolated regions with much higher objective weight than its surroundings. However, since our only evaluation of plan quality is by comparison to RayStation plans, the dif-
ference might as well depend on the RayStation VMAT algorithm being unfit for plans with multiple high weight objectives.

When we compare the results we get when using the different kinds of segments for the different cases, we see that which kind of segments that give the best dose result varies from case to case. For the prostate case, we get the best results using DMLC segments, whereas for the head-and-neck case, the IMRT and intermediate segments provide the best results and for the prostate case, the IMRT segments give us the best results. Thus, we cannot say which kind of segments that generally perform best and neither are we able to say that any kind of segments should be ruled out of consideration. In order to be able to make a complete evaluation of the suitability of the different kinds of segment we would need solutions closer to optimality than the ones generated in this project.

To summarize the discussion, we can say that the current implementation does not provide solutions close enough to optimality to be particularly useful. We also see that in order to get good correspondence between results in fluence domain and results in dose domain, it is important to keep the leaf travel between adjacent control points small. Moreover, only considering fluence data without any consideration of the weight of the corresponding objectives does not seem to generate sufficiently good plans in the general case.
Chapter 7
Further Outlook

The results show that the model could be of some use. However, we need to make some improvements first, especially regarding implementation. In this chapter we will list a couple of things that might be interesting to investigate in order to improve the model.

7.1 Providing Better Segments

As we have seen, the model does a quite good job at reproducing the desired fluence when it is provided with good segments. Thus, one improvement of the model would be to find a way to generate good segments. For generated segments to be considered as good, they should be able to more or less satisfy three criteria. The segments should, when combined, be able to deliver fluence similar to the desired fluence. The segments should be such that it is possible to move from segment to segment within the set without large amounts of leaf travel. The number of segments required to deliver fluence similar to the desired fluence should not be larger than approximately 8. Finding a way to generate segments that fulfill these criteria could possibly lead to improvements of the model.

The DMLC segments that have been used in this project are generated in a simple manner. Perhaps investigation of more advanced techniques for DMLC conversion of fluence maps could provide more suitable segments. In that case, we would like to be able to limit the number of segments generated for each fluence map, in order to avoid segments that are too narrow, such as the current DMLC segments. We would also like to be able to constrain the leaf travel between adjacent segments, in order to minimize the difference between modeled fluence and delivered dose. However, these two goals are conflicting, since fewer segments tend to lead to more leaf travel.

To find better suited segments will probably require a lot of work. On the other hand, better suited segments could lead to large improvements of
the plans generated with the model and could therefore be interesting to investigate.

7.2 Include Objective Weight in Fluence Maps

In section 6.2, we discussed the fact that the model tends to fail to deliver the right fluence to the most important regions and thereby generating plans that are worse than the plans generated by RayStation.

Since there is no weighting of the regions in the fluence maps, it is not surprising that the model cannot distinguish between important and less important regions. If we could get data on which bixels in each fluence map that corresponds to which objective weight, it would be possible to consider different regions differently in the model. Thus, we could have a model where deviation in a bixel is penalized according to the dose objective value in that bixel. It would be interesting to see how the model would perform given this kind of information. However, it is not clear whether it would be easy to obtain such data or not, and how complicated it would be to assign different objective weights to individual bixels.

7.3 Variation of the Model Parameters

Throughout this project, the model has only been implemented using data from a 36 beam IMRT plan. It would we quite easy to make changes in the model such that IMRT plans with a different number of beams could be used instead. It would be interesting to see if and how the results differ if the model is implemented with a different number of IMRT beams.

The spacing of the control points has also been kept constant throughout this project. As with the number of beams, changing the model to allow for a different control point spacing would be quite easy, and it would be interesting to see what happens with the results if the control point spacing is changed.

Related to the two above notions is the way we have chosen to model how the fluence delivered at control points contribute to fluence maps optimized for adjacent angles. The chosen way to model it is a mere assumption. Further investigations of how fluence delivered at a certain angle contributes to intensity modulation at a nearby angle could help improve the validity of the model.
7.4 Reduce Complexity Associated with the Objective Function

We have chosen to use a linear objective function in the model, in order to get a mixed binary linear model. To be able to express the objective function as a linear function, we had to introduce an auxiliary variable for each matrix element of each fluence matrix. For e.g. the considered head-and-neck case, this means 73,440 auxiliary variables. This is a huge number of extra variables. There are two possible ways to reduce these variables that could be worth investigating.

The first thing we could try is to use a quadratic objective function instead of a linear objective function. To get a quadratic objective function, we could consider the square of the difference between desired and delivered fluence in each matrix element. The objective function is then the sum of the square of all differences. In general, it is computationally more efficient to solve a model with a linear objective function, but since the quadratic objective function does not require any auxiliary variables, it is worth investigating using a quadratic objective function instead of a linear objective function. Some attempts has been made to use a quadratic objective function in this project, but not enough to provide any conclusive answers on the usefulness.

The second thing we could try to investigate is to make fluence matrices smaller. We could do this by replacing groups of $k$ adjacent bixels with their mean value and thereby reduce the size of the fluence matrix with a factor $k$. We would then need to reduce the size of the segments in the same way in order to be able to compare the desired and the delivered fluence. This method could potentially reduce the complexity associated with the objective function drastically. However, it is uncertain how the accuracy would be affected when groups of bixels are replaced with mean values. The potential reduction of complexity would make it worth investigating though.

7.5 Heuristic Solution Methods

In the current implementation, the model is solved using the branch-and-bound solver in CPLEX. If we would like to implement the model in other software, solving the model using CPLEX is not a viable option. Therefore, in order to make use of the model, it would be interesting to investigate heuristic solution methods to solve the model. Moreover, we have been unable to generate solutions close to optimality using CPLEX. As mentioned in section 6.1, just feeding data into optimization software prohibits a lot of exploration of the problem structure at hand.

In order to be able to use our knowledge of the problem at hand, we can investigate heuristic methods. Commonly used heuristic methods for mixed
integer linear programming problems include tabu search and simulated annealing. These methods can also be combined in different ways or used together with e.g. branch-and-bound to find solutions closer to optimality.

Regardless of which heuristic solution method that is chosen, the problem will be easier to solve if the method is provided with a good starting guess. For some methods, the starting guess is crucial for a good outcome. It is therefore very interesting to try to develop a method for generating good starting guesses for the model. With knowledge of the model, it is at least not that hard to generate a feasible starting guess, then we just need to choose a segment for each control point and optimize the segment weights. By finding a clever way to choose the segments, this could provide a good starting point for heuristic solution methods for the model.
Chapter 8
Conclusion

In this thesis, an optimization model for conversion from an IMRT plan to a VMAT plan is developed. The model is based on choosing predefined segments to deliver intensity modulation as similar as possible to the intensity modulation of the IMRT plan. The model is formulated as a mixed-binary linear programming problem.

Since the model is NP-complete, implementation of the model requires assumptions and simplifications that potentially reduce the accuracy of the model.

The similarity in fluence that the model is able to achieve depends heavily on the predefined segments given to the model. However, if the segments are large enough to be able to achieve sufficient fluence modulation using at most about 8 segments and requires a sufficiently small amount of leaf travel, then the model can achieve fluence similarity within 100 % of the best possible fluence similarity we can get with the considered segments.

Similarity in fluence does not correspond directly to similarity in dose distribution. This is related both to how the model works and to lack of information in fluence domain. The model only considers fluence delivered at the control points, whereas in reality, fluence is delivered continuously. Thus, if segments at adjacent control points differ a lot from one another, we will get fluence contribution not accounted for in the model. In fluence domain, there is no information about which parts of the fluence maps correspond to which dose objective. Thus, it is not possible to make the model focus on achieving the highest fluence similarity for the most important dose objectives.

The current implementation of the model is not able to provide solutions for which closeness to optimality can be verified. This seems to be due to the inability to make use of our knowledge of the problem at hand when feeding our data to the optimization software. In order for the model to be useful, it is therefore necessary to explore ways of implementing the model where we can take advantage of our understanding of the model. However, it is not presently known if the optimal solution is significantly better than the one computed in our implementation.
The major reasons for the model's inability to provide better plans than RayStation are probably the optimality gap of the current implementation and the discrepancy between similarity in fluence and similarity in dose distribution.

We may conclude that the model shows some promising features, but improvements need to be made before the model can be of any real use.
References