

Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy

Fredrik CARLSSON*

Technical Report TRITA-MAT-2007-OS3

Department of Mathematics

Royal Institute of Technology

October 2007

Abstract

A method for generating a sequence of intensity-modulated radiation therapy step-and-shoot plans with increasing number of segments is presented. The objectives are to generate high-quality plans with few, large and regular segments, and to make the planning process more intuitive.

The proposed method combines segment generation with direct step-and-shoot optimization, where leaf positions and segment weights are optimized simultaneously. The segment generation is based on a column generation approach. The method is evaluated on a test suite consisting of five head-and-neck cases and five prostate cases, planned for delivery with an Elekta SLi accelerator.

The adjustment of segment shapes by direct step-and-shoot optimization improves the plan quality compared to using fixed segment shapes. The improvement in plan quality when adding segments is larger for plans with few segments. Eventually, adding more segments contributes very little to the plan quality but increases the plan complexity. Thus, the method provides a tool for controlling the number of segments and, indirectly, the delivery time. This can support the planner in finding a sound trade-off between plan quality and treatment complexity.

Key words. intensity-modulated radiation therapy, step-and-shoot delivery, optimization, column generation

1. Introduction

The advantage in dose conformity of intensity-modulated radiation therapy (IMRT) plans compared to conventional plans comes at the expense of increased treatment complexity. For the widely used step-and-shoot IMRT delivery technique, this complexity can be measured in terms of the number of multi-leaf collimator (MLC) segments, the number of monitor units (MUs) and the jaggedness of the segments. A challenge in step-and-shoot IMRT is to design “simple” plans, i.e. plans with few, large and regular segments, that fulfill the prescribed treatment goals. This paper is motivated by this challenge in combination with a desire to make the planning process more intuitive.

There are many reasons for generating simple step-and-shoot plans. Such plans have few monitor units and low integral dose, which reduces the risk of secondary cancers [12]. Simple plans

*Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden, (fcar@kth.se); and RaySearch Laboratories, Sveavägen 25, SE-111 34 Stockholm, Sweden. The research was supported by the Swedish Research Council (VR).

have shorter delivery times than complex plans, which, in addition to increasing the throughput of patients, avoids the risk of reduction in radiobiological effects due to prolonged delivery times [16, 27]. A simple plan is less sensitive to geometrical uncertainties and can be delivered with higher accuracy than a plan with many small and irregular segments [8, 23].

It has been demonstrated that rather simple intensity patterns can generate high-quality plans [2], indicating that few and regular segments can produce satisfactory plans if chosen wisely. This has indeed been shown in [3, 13], where approximately five segments per beam seem to suffice for most cases. The number of MUs may sometimes be reduced significantly with only small changes in plan quality, see [10].

Encouraged by these findings, we present a method for generating a sequence of step-and-shoot plans with increasing number of segments. The objective is twofold. First, to generate simple and high-quality step-and-shoot plans. Second, to support the planner in exploring the trade-off between plan quality and complexity. The proposed method combines a segment generation module with direct step-and-shoot (DSS) optimization. This combination allows for controlling the number of segments while fine-tuning their shapes and weights. The method combines the flexibility of altering the set of segments with the practical benefits of DSS optimization.

In DSS optimization, which is often referred to as *direct aperture optimization*, the segment shapes and weights are optimized simultaneously. Similar approaches are described in [3, 4, 9, 11, 14, 15, 24, 25]. A major advantage of these approaches compared to the traditional two-step IMRT approach is that the optimized plan is deliverable. There is therefore no need for any post-process that might degrade the quality of the final plan. The degeneracy inherent in IMRT problems is not an issue in these direct approaches, contrary to the two-step approach where regularization is needed to obtain smooth fluence profiles, see e.g. [6]. In addition, a clinical study has found that DSS optimization simplifies plans without reducing plan quality, compared to plans generated with the two-step approach [26]. Our segment generation module is based on the column generation approach for IMRT presented in [22].

The paper is organized as follows: The solution process of the method is described in Section 2, with emphasis on DSS optimization and segment generation. The computational study is described in Section 3. The results are given in Section 4 and a concluding discussion is carried out in Section 5.

2. Method

In the proposed method, a sequence of step-and-shoot plans with increasing number of segments is generated. This is done by alternating between generation of new segments and optimization of segment shapes and weights. No leaf-sequencing is necessary and all plans in the generated sequence are deliverable, so no post-processing is needed.

The initial segments, one per beam, are generated from the projection of the target(s) onto the isocenter plane of each beam. If more than one segment is needed for reconstructing a projection, e.g. if the projection is not contiguous along the leaves, a segment is created from the largest contiguous part of the projection. Then, a number of loops are performed according to Figure 1.

We now describe one loop in the solution process, where the words emphasized in bold face correspond to boxes in Figure 1. In the first stage of the loop, a **direct step-and-shoot optimization** is performed; see Section 2.1 for details. The optimization is run for few iterations

and will, in general, not reach the optimal solution. Next, the leaf positions are fixed and a **segment weight optimization** is performed to ensure that the segment weights are optimal or close to optimal so the plan can be **saved**. If the planner is satisfied with the plan or if the maximum number of loops is reached, the solution process is terminated. Otherwise, the **segment generation** module is entered. This module uses the gradient of the objective function with respect to the fluence to create the segments that are most likely to improve the objective function value and thus improve the plan. For details, see Section 2.2. To adjust the segment weights of the former and the new segments, another **segment weight optimization** is performed directly after the segment generation module. Segments with weights at the lower monitor unit limit set by the planner might then be removed in the **segment removal** module. The motivation of this module is to get rid of the segments, often generated early in the solution process, that contribute very little to the plan. The following rules are applied for the segment removal: The segments generated in the current loop are immune and cannot be removed. Further, at most $k - 1$ segments can be removed if k segments were just generated and no more than one segment per beam can be removed in one loop.

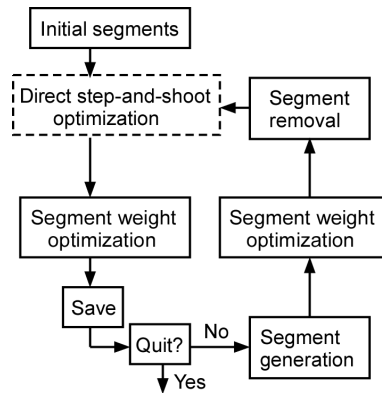


Figure 1: Solution process.

2.1. Direct step-and-shoot optimization

The direct step-and-shoot optimization problem is given by

$$\begin{aligned}
 & \underset{x,w}{\text{minimize}} && F(d(x,w)) \\
 & \text{s.t.} && Ax \geq b \\
 & && w \geq w_0,
 \end{aligned} \tag{2.1}$$

where leaf-position variables are denoted by x , segment weight variables are denoted by w and the dose distribution is denoted by d . The constraints $Ax \geq b$ represent MLC requirements such as interdigitation, minimum gaps and minimum segment areas. The bounds on w are included to ensure that all segments fulfill their lower monitor unit limit w_0 . These constraints ensure that the solution to (2.1) is deliverable and that no post-processing is needed. A segment weight optimization problem is equivalent to (2.1) with x fixed. This restricted problem is much easier to solve since the number of variables is smaller and since d is linear in w , as opposed to the non-convexity introduced by the leaf position variables.

In our study, the objective function F is a weighted sum of standard IMRT penalty functions such as one- and two-sided quadratic penalty functions and $gEUD$, see e.g. [7] for mathematical formulations of the former and [17] for the latter. The parameters of the penalty functions are set up according to RTOG protocols; see Section 3.2 and Table 1 for details. No so-called hard nonlinear constraints are used.

2.2. Segment generation

Our segment generation strategy follows the one presented in [22]. Initially, a discretization of the cross-sections of the beams into beamlets is performed. The beamlet weights, i.e. the elements of the discretized fluence, are denoted by τ . The beamlets are aligned with the collimator for each beam and the beamlet size is set to the leaf width across the leaves and to five millimetres along the leaves. This size gives a good balance between computational time and quality of the generated segments. Next, the gradient of F with respect to τ is calculated. New segments are then generated by combining beamlets into feasible segments so that the sum of the gradient components of the included beamlets is as small as possible. This can be formulated as the integer programming problem

$$\begin{aligned} & \underset{z}{\text{minimize}} && \left(\frac{\partial F}{\partial \tau}\right)^T z \\ & \text{s.t.} && z \in \mathcal{Z}, \end{aligned} \tag{2.2}$$

where z is a binary vector specifying the beamlets that are exposed by a segment and \mathcal{Z} is the set of beamlet regions corresponding to feasible segments with respect to the MLC used. Note that (2.2) is separable per beam.

We do not solve (2.2) using general-purpose integer programming methods. Instead, (2.2) is solved by constructing a layered graph and formulating it as a shortest-path problem per beam as suggested in [5]. For an introduction to shortest-path problems, see e.g. [18]. Every node in the graph represents a leaf pair configuration and the weights of all arcs incident on a node equals the sum of the gradient components for the exposed beamlets in the corresponding leaf pair configuration. The leaf positions of the generated segments are thus restricted by the beamlet size. Since the beamlet size along the leaves is as small as five millimetres and since the leaf positions are fine-tuned afterwards when solving (2.1), this restriction should not be an issue.

With the shortest-path formulation, MLC requirements such as connectivity, interdigitation and minimum gaps are handled in the graph construction, see [22]. Our algorithm also incorporates a penalizing strategy to avoid irregular segment shapes. This strategy is flexible in that it can be tuned so that the segments generated are anywhere between the jagged segments that are optimal with respect to (2.2), to purely rectangular-shaped segments.

Once segments have been generated for all beams, they are ranked and some, or all, are added to the set of segments. The weights of the added segments are set to the lower monitor unit limit specified by the user. Forthcoming optimizations will then adjust their weights.

The segment generation process is illustrated in Figure 2. A gradient map is shown in (a). Blue beamlets have negative gradient components, white beamlets have zero gradient components and red/yellow beamlets have positive gradient components. Ideally, the exposed beamlets of the generated segment should coincide with the negative gradient components. This is however not possible, in general, since the MLC requirements do not allow such a segment. The optimal solution to (2.2), given by the optimal solution to the corresponding shortest-path problem, results in the exposed beamlet pattern shown in (b). This pattern follows the central contiguous

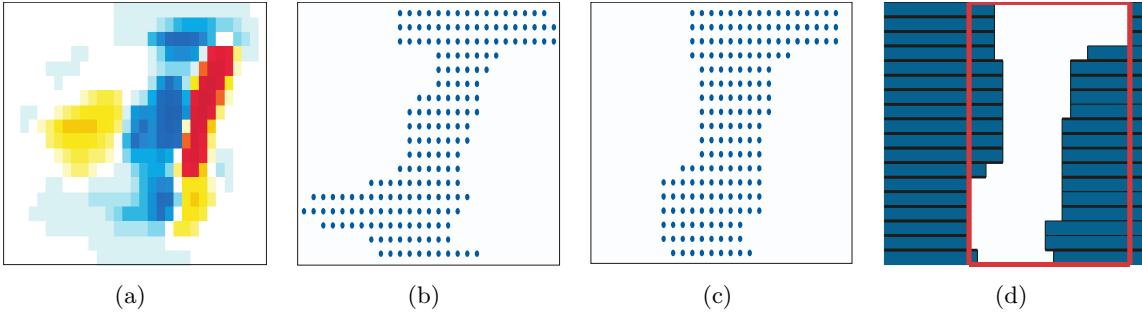


Figure 2: The process of generating a feasible and regular segment from the gradient with respect to beamlet weights. The gradient is shown in (a). The blue beamlets have negative gradient components and the red beamlets have positive gradient components. The beamlet regions denoted by dots in (b) and (c) are the solutions to the corresponding shortest-path problem without and with the penalizing strategy, respectively. The generated segment is illustrated in (d), where the red box gives the jaw positions.

region with negative gradient components closely. A few peripheral regions with negative gradient components are excluded in order to avoid some of the regions with positive gradient components. With the penalizing strategy, the solution to the shortest-path problem has a much more regular beamlet pattern, seen in (c). The cost of this regularity is to exclude some regions with negative gradient components located far from the center of the segment. These excluded regions may be included later, either through the movement of the leaves of the current segment or as part of a new segment. Either (b) or (c) can be used to generate a feasible segment. Here, the regular shape in (c) is transformed into the segment shown in (d), where the jaw positions are specified by the red box.

2.3. Illustration of the solution process

An illustration of the first four loops of our method on a test case is given in Figure 3. The figure follows the evolvment for one beam. The top row shows, from left to right, the initial projection-based segment and the first three generated segments in order of creation. The lower row shows the gradient maps of the beam prior to solving the shortest-path problem at each stage. The black lines outline the created segments. The gradients have identical color scales to illustrate that the gradient approaches zero as the method proceeds. Again, blue areas have negative gradient components and red areas have positive gradient components. Following the arrows, we see how segments are created and how the gradients change in every loop. All three segments generated include one or more component with positive gradient. In this way, more negative gradient components can be included and the total sum of the included gradient components can be decreased. The impact of the penalizing strategy can also be seen in the figure; some regions of negative gradient components on the sides of the created segments are omitted. The increased jaggedness of including them is judged to be worse than their contribution to the total sum of the included gradient components by our penalizing strategy.

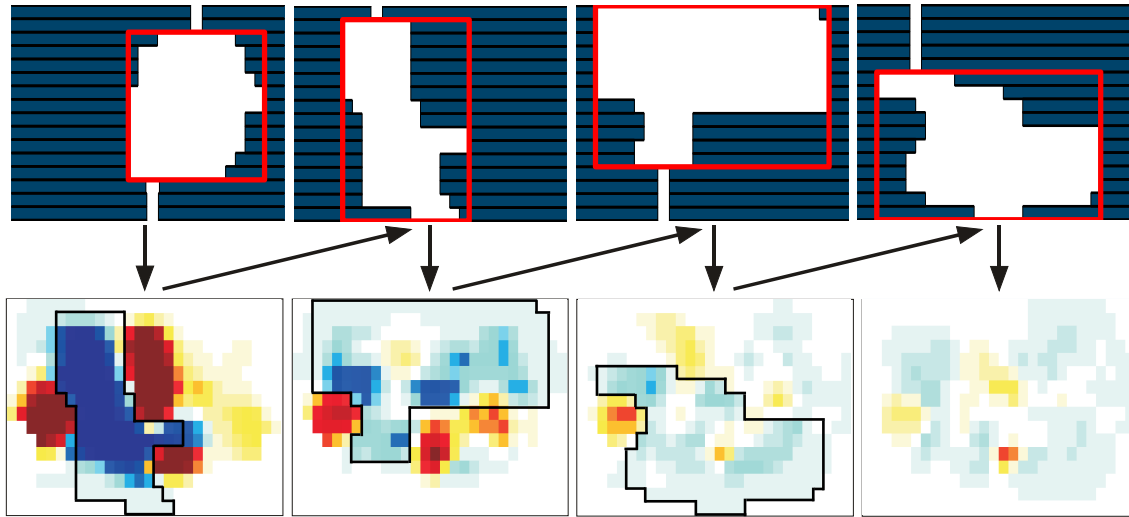


Figure 3: Top: The initial projection-based segment and the first three generated segments for one beam in order of creation. The red boxes show the jaw positions. Bottom: The gradient maps prior to solving the corresponding shortest-path problem. Blue beamlets have negative gradient components and red have positive gradient components. The black lines outline the created segments.

3. Computational study

The main comparison performed in this study is between our proposed method, denoted by the *adjustable leaves* approach, and a limited version of it, denoted by the *fixed leaves* approach. The fixed leaves approach means that the dotted box in Figure 1 is skipped. This approach resembles the column generation approach, described in [19, 22]. The main differences to these column generation approaches are that the fixed leaves approach uses a nonzero lower bound on the segment weights and removes segments during the solution process.

In the adjustable leaves approach, 10 segment weight iterations and 10 DSS iterations are performed in each loop. In the fixed leaves approach, 20 segment weight iterations are performed in each loop. The segment weight iterations are split equally between the two segment weight optimizations in every loop. The plans generated with the two approaches are compared to *benchmark* plans. The benchmark plans are generated by optimizing beamlet weights for 10 iterations, applying a leaf-sequencing algorithm to generate 50 segments and performing 90 iterations of DSS optimization. The objective function is the same as in the other approaches.

3.1. Treatment planning software

New functionality, such as the segment generation module, has been merged into ORBIT Workstation [20], which is used for the treatment planning in this study.

To ensure high-speed dose calculations during the optimization without sacrificing the accuracy of the dose in the generated plans, the following dose calculation strategy is employed: After the segment removal in each loop, dose is computed with a collapsed cone convolution algorithm [1]. The fluence engine is based on a two-source head-scatter model, where MLC

transmission and flattening filter are accounted for. At other stages in the loop, the dose is calculated with a faster pencil-beam algorithm. The optimization algorithm used is a quasi-Newton sequential quadratic programming method.

3.2. Patient cases

The test suite consists of five head-and-neck cases and five prostate cases. The head-and-neck cases are set up with seven or nine beams and the prostate cases are set up with five beams. All cases have voxel sizes of $(4 \times 4 \times 4 \text{ mm}^3)$.

The cases are planned for delivery with an Elekta SLi 6 MeV linac (Elekta Oncology Systems, Crawley, UK), equipped with an MLC_i collimator. This MLC, which is widely used for step-and-shoot delivery, does not allow for interdigitation and has a minimum gap requirement. Further, jaws orthogonal to the leaf banks are not allowed to travel beyond the central axis. These requirements make the segment generation more challenging and hence interesting. The treatment planning is set up according to the RTOG 0615 protocol for the head-and-neck cases and is based on the RTOG 0415 protocol for the prostate cases¹. In Table 1, the planning regions and specifications for the head-and-neck and prostate cases are listed. For the prostate cases, the CTV is the prostate itself and the PTV margin is set to five millimetre, motivated by [28]. The width of the rectum and bladder walls is set to four millimetres in the left-right and anterior-posterior directions.

Head and neck		Prostate	
Structure	Specification	Structure	Specification
PTV_{70}	$D_{95} \geq 70$	CTV	$D_{99} \geq 78$
	$D_{99} \geq 65.1$	PTV	$D_{99} \geq 74.1$
	$D_{20} \leq 77$		$D_1 \leq 81.9$
	$D_5 \leq 80$	Rectum/Bladder walls	$D_{30} \leq 70$
$D_{mean} \leq 74$		$D_{50} \leq 53$	
$PTV_{59.4}$	$D_{95} \geq 59.4$	Femoral Heads	$D_5 \leq 53$
	$D_{99} \geq 55.2$	Rectum	$D_{15} \leq 74$
	$D_{20} \leq 77$		$D_{25} \leq 69$
	$D_5 \leq 80$		$D_{35} \leq 64$
Brainstem	$D_{max} \leq 54$		$D_{50} \leq 59$
Spinal cord	$D_{max} \leq 45$	Bladder	$D_{15} \leq 79$
Mandible	$D_{max} \leq 70$		$D_{25} \leq 74$
Either parotid gland	$D_{mean} \leq 26$		$D_{35} \leq 69$
Oral cavity	$D_{mean} \leq 40$		$D_{50} \leq 64$

Table 1: Plan specifications for the head-and-neck cases and the prostate cases. The dose covering $x\%$ of the volume is denoted by D_x and dose levels are given in Gy.

For all cases, the solution process is run for 20 loops and the lower bound on the segment weights is set to 4 MU, in accordance with the findings in [23]. For the head-and-neck cases, four segments distributed over the beams are created in every loop. For the prostate cases, three

¹<http://www.rtog.org>

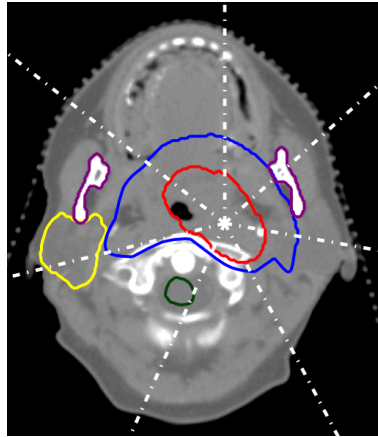


Figure 4: A transversal slice of the HN₂ head-and-neck case, displaying the contours of PTV_{70} (red), $PTV_{59.4}$ (blue), cord (green), left parotid gland (yellow) and mandible (purple).

segments are created in every loop. The reason for this difference is that the head-and-neck cases are set up with more beams than the prostate cases. A transversal slice of one of the head-and-neck cases, referred to as HN₂, is shown in Figure 4.

4. Results

For the sake of brevity, we only present results for the head-and-neck cases. The results for the prostate cases are qualitatively similar and do not change the message we want to convey. Section 4.1 contains the main results, a comparison of the plan quality of the generated sequences in different measures for the head-and-neck cases. Some results from the solution process are given in Section 4.2.

4.1. Plan quality comparisons

In the figures of this section, we show different measures of plan quality versus number of segments for the generated plans for the HN₂ case. The figures compare the adjustable leaves approach (solid lines) with the fixed leaves approach (dotted lines). The HN₂ case is one of the hardest in the test suite and it therefore accentuates the differences between the approaches slightly more than most of the other plans.

Figure 5 shows the conformity index of the target region $PTV_{59.4}$. The conformity index is calculated according to [21] using the D_{95} level of 59.4 Gy. The higher value of the index, the more conform dose distribution to $PTV_{59.4}$. It is clear that the conformity index values for plans with few segments are much higher with the adjustable leaves approach than with the fixed leaves approach. This difference is reduced as the number of segments increases. With 50 segments or more, the two approaches produces plans with similar conformity index values.

The maximum dose levels to brainstem and cord are shown in Figure 6. These dose levels are evaluated as the maximum dose to the regions when the volume equivalent to one voxel with the highest dose has been removed. The horizontal lines show the corresponding criteria specified in Table 1. Both approaches produce plans that fulfill both criteria, even though slightly more

segments are needed for the fixed leaves approach.

The D_{95} and the D_{99} levels of the PTV_{70} target region are shown in Figure 7 together with their criteria. Again, the adjustable leaves sequence reach the plan criteria with fewer segments than the fixed leaves sequence. With more segments, however, the differences vanish and both approaches fulfill both criteria. Figure 8 shows the same dose specification levels for the $PTV_{59.4}$ target region. For the HN_2 case, these criteria are harder to fulfill than the criteria for PTV_{70} , i.e. more segments are needed. The trend is the same as in the previous figures; large differences with few segments and small differences with many segments.

The average regularity of the segments of the generated plans are shown in Figure 9. The higher value in the figure, the more regular segments. This regularity measure is given by the mean value of the quotients of segment area and segment perimeter for all segments in the plan. The first property to observe is that the average regularity decreases with increasing number of segments, for both approaches. This is due to that the regions of negative gradient components get smaller as the solution process proceeds. The figure also shows that the adjustable leaves approach produces less regular segments than the fixed leaves approach. There are two reason for this. One is that the DSS optimization allows the segment shapes to become slightly more irregular. The other is that the gradients of F with respect to τ get closer to zero earlier in the solution process with the adjustable leaves approach. This implies that smaller segments are created in the segment generation module.

The comparisons in Figures 5 – 9 have demonstrated that the impact of fine-tuning the segment shapes with DSS optimization is more pronounced for plans with few segments. Plans with many segments are rather similar, both in terms of regularity of the segments and plan quality.

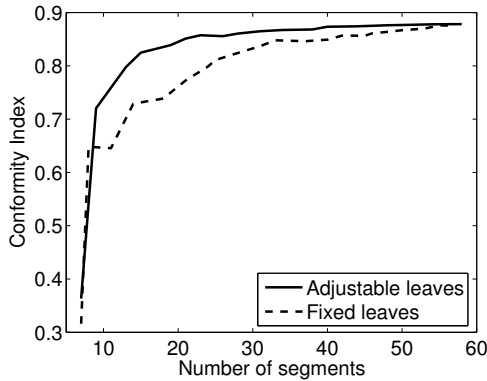


Figure 5: Conformity index of $PTV_{59.4}$ at 59.4 Gy for both approaches.

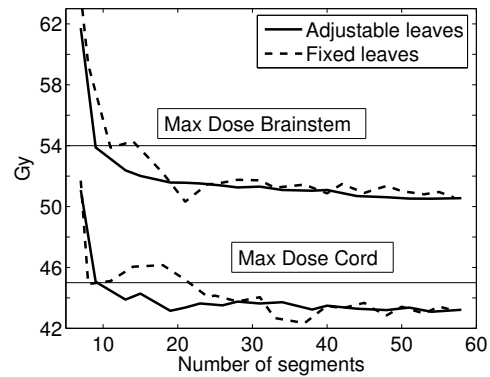


Figure 6: Max dose to brainstem and cord. The horizontal lines represent the RTOG criteria for the organs.

Table 2 shows the mean relative violation for a subset of the generated head-and-neck plans, with the mean taken over all plan criteria in Table 1. The relative violation for a given criterion is given by $\max(0, \frac{d_c - d_p}{d_p})$ if it is an upper bound criterion and by $\max(0, \frac{d_p - d_c}{d_p})$ if it is a lower bound criterion, where d_p is the prescribed dose level and d_c is the current dose level at the specification level. The values in Table 2 are given in parts per thousand. If, for instance, the

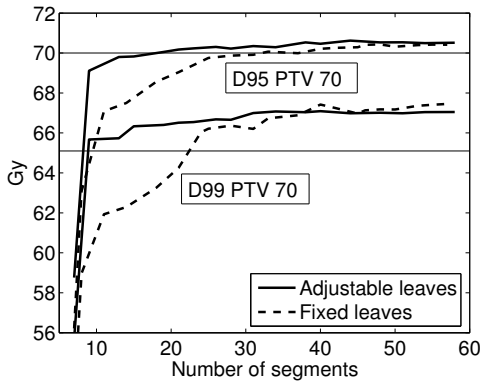
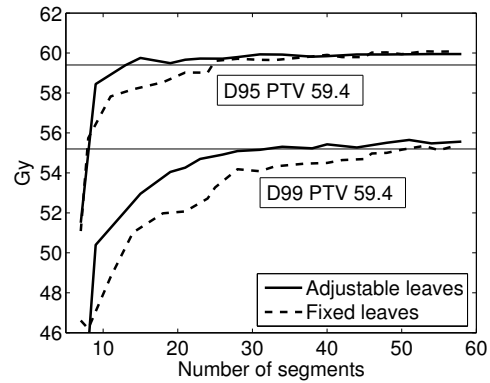
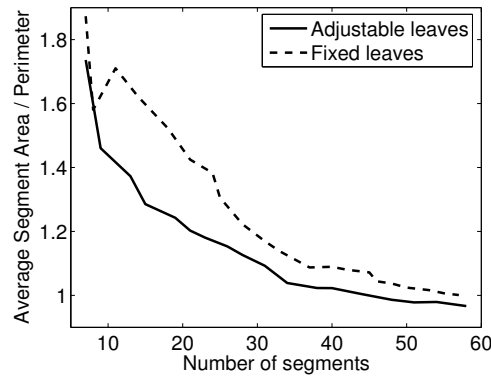
Figure 7: D_{95} and D_{99} levels of PTV_{70} .Figure 8: D_{95} and D_{99} levels of $PTV_{59.4}$.

Figure 9: The average number of the quotients of segment area with segment perimeter for all segments in each plan. The higher number, the more regular segment.

D_{95} level of PTV_{70} is 69 Gy for a plan, the relative violation for this criterion is $1/70$. If all the other 13 criteria are fulfilled, the mean relative violation becomes one part per thousand. To limit the number of columns in the table, eight of the twenty plans in each sequence are selected. The plans selected are those for which the number of segments is closest to 15, 20, \dots , 50. (If two plans in a sequence are equally close, the one with the lowest mean relative violation is chosen.) The rightmost column shows the mean relative violation values for the benchmark plans.

For all five cases, the mean relative violation is lower with the adjustable leaves approach than with the fixed leaves approach. As seen in the figures for the HN_2 case, the differences between the approaches are larger for plans with few segments than with many segments. The benchmark plans have rather similar values of the mean relative violations compared to the 50 segment plans in both approaches. The HN_1 plan is better with the benchmark approach, while the HN_2 and HN_5 plans are worse.

The numbers of segments required to achieve a mean relative violation less than 1.0 and 0.1 parts per thousand for all head-and-neck plans are given in Figure 10. Consequently, fewer segments are required with the adjustable leaves approach, except that no generated plan fulfills

the 0.1 requirement for the HN_1 case. On average, the adjustable leaves approach requires 26 segments and the fixed leaves approach requires 38 segments to fulfill the 1.0 requirement. This implies three to five segments per beam, which is in line with the findings in [3,13]. The figure also shows that the number of required segments varies rather much between the cases. It might therefore be hard to predict the number of required segments for a case prior to planning. This observation motivates our approach to present a sequence of plans to the planner.

The number of MUs increase with increasing number of segments for both approaches and for all cases in this study. Our method thus produces plan sequences where the delivery time of the plans increases in every loop. Therefore, the results presented above would not change qualitatively if we replace number of segments by monitor units or delivery time, i.e. the treatment complexity increases as the method proceeds. The difference in MUs between the two approaches vary from case to case, but the aggregated difference over all cases is rather small. Compared to the benchmark plans, the number of MUs for the 50 segment adjustable leaves plans are higher in four of the five head-and-neck plans. Selecting the adjustable leaves plans that achieve the same mean relative violation as the benchmark plans in Table 2, it turns out that the number of MUs for the benchmark plans and the selected plans are similar.

Figures 11 – 12 show the dose-volume histograms (DVHs) of the $PTV_{59.4}$ and the PTV_{70} target regions for the HN_2 case. The plan criteria are marked with triangles. The figures compare three plans. The first is a 31 segment plan with 611 MU taken from the adjustable leaves sequence, which is the plan with the smallest number of segments fulfilling all criteria for this case. The second is a 31 segment 667 MU plan taken from the fixed leaves approach. The third plan is our benchmark plan, with 50 segments and 682 MU. For both regions, the benchmark plan has the most conform dose distribution and the fixed leaves plan the least conform one. While the adjustable leaves plan fulfill all criteria, the benchmark plan fulfills all but the D_{99} criterion for the $PTV_{59.4}$ region. The low conformity of the fixed leaves plan results in violation of the D_{99} criterion of the $PTV_{59.4}$ region and of both dose limiting criteria of the PTV_{70} region.

Case	Approach	Number of segments								Benchmark
		15	20	25	30	35	40	45	50	50
HN_1	Adjustable	42	15	6.7	3.7	2.8	0.2	0.9	0.7	0.1
	Fixed	44	24	13	8.5	5.1	3.8	1.3	0.9	
HN_2	Adjustable	4.7	1.7	0.6	0.1	0	0	0	0	1.1
	Fixed	22	15	4.2	2.5	1.2	1.1	0.9	0.1	
HN_3	Adjustable	1.3	0.6	0.4	0.2	0.2	0.1	0	0	0
	Fixed	6.0	3.7	1.2	0.7	0.3	0.2	0.1	0.3	
HN_4	Adjustable	16	2.7	0.6	0.1	0	0	0	0	0
	Fixed	33	13	5.7	2.1	0.5	0.5	0	0	
HN_5	Adjustable	4.1	1.2	1.0	0.5	0.5	0.3	0.6	0	0.8
	Fixed	18	8.1	5.2	2.2	0.9	1.0	0.6	0.2	

Table 2: Mean relative violations, in parts per thousand, for a selection of the generated plans for all head-and-neck cases. The plans chosen from the sequences are the plans that have the number of segments closest to the specified number of segments in every column.

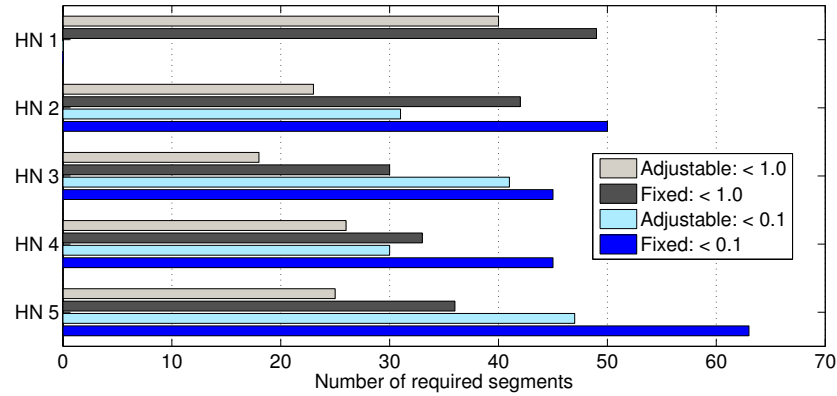


Figure 10: The number of segments required to achieve a mean relative violation less than 1.0 and 0.1 parts per thousand.

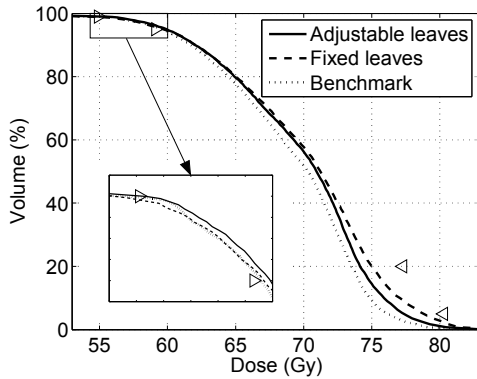


Figure 11: DVH of $PTV_{59.4}$ for the HN_2 case for one plan with 31 segments from both approaches and for the benchmark plan.

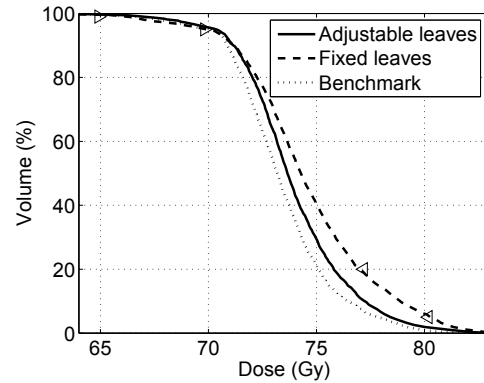


Figure 12: DVH of PTV_{70} for the HN_2 case for one plan with 31 segments from both approaches and for the benchmark plan.

4.2. Solution process results

The leaf movements in four segments of one beam for the HN_2 case with the adjustable leaves approach are shown in Figure 13. The top row shows the initial segment and three generated segments when created, while the bottom row shows the same segments at the end of the solution process, i.e. after 20 loops. The initial segment, in the leftmost column, has been modified by 20 DSS optimizations. The other segments, generated in loops one, two and five, have been modified by 19, 18 and 15 DSS optimizations, respectively. The leaf movements are rather small, but as demonstrated in the plan comparisons of the previous section, this fine-tuning of the segment affects the plan quality considerably.

The process of finding the optimal segment from a gradient map takes around a tenth of a second on a standard computer (2.0 GHz Pentium 4 processor, 2 GB RAM). Using 20 loops with

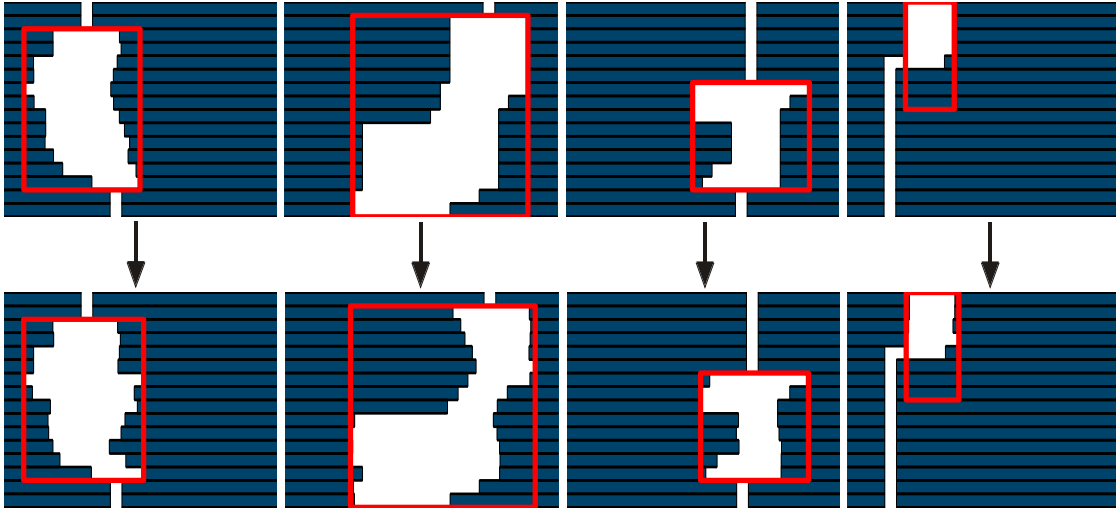


Figure 13: Leaf movements with the adjustable leaves approach. The red boxes show the jaw positions. Top: The initial segment and three generated segments for one beam when created. Bottom: The same segments at the end of the solution process, after 20 DSS optimizations.

7 beams for a case, we need to solve 140 such problems, resulting in a total computation time of approximately 15 seconds for the segment generation part in the solution process. The current implementation of the shortest-path problem has focused more on functionality than speed, so we envisage that this solution time could be shortened with a more efficient implementation. The overall computational time for one head-and-neck case is in the order of 40 minutes, i.e. each loop takes about two minutes.

In the 20 loops performed, 80 segments are generated. This would result in final plans with 87 or 89 segments for the head-and-neck cases if no segments were removed. The mean number of segments for the final plans for the head-and-neck cases is 56, meaning that around 30 segments are removed on average in the solution process.

5. Discussion and conclusion

It has been shown that the rate of plan improvement is much higher with the adjustable leaves approach than with the fixed leaves approach early in the solution process. This is a crucial difference since it implies that the adjustable leaves approach can find adequate plans with fewer segments and probably fewer MUs than the fixed leaves approach. Once a certain number of segments have been generated with the adjustable leaves approach, adding more segments merely increase the treatment complexity. The proposed adjustable leaves approach is therefore suitable for supporting the planner in exploring the trade-off between treatment complexity and plan quality.

One reason for the large differences between plans with few segments could be the lack of flexibility inherent in the fixed leaves approach. The generated segments are based on local gradient information. As the solution process proceeds, the segments therefore need to deform to adjust to changes in the dose distribution. With no leaf-position optimization, this adjustment

cannot be done. This lack of flexibility might be especially problematic for plans with few segments. Another reason could be that the adjustable leaves plans have slightly more jagged segments, which allows for more conform dose distributions.

In this study, the initial segments are constructed from the projection of the target(s). There are two other ways of generating the initial segments in the current implementation. One is to perform a few iterations of beamlet weight optimization and then create the initial segments using a leaf sequencing algorithm, i.e. to mimic the first part of the solution process for the benchmark plans. With this approach, more than one initial segment per beam can be generated and the number of loops can be reduced. The other way is to start with no segments at all and to generate the initial ones in the segment generation module. Then, some beams will have no segments initially. This approach might be useful in the context of gantry angle optimization; define a plan with redundant beams and run the method for some loops, then discard the beams that have no segments. This could provide a good initial estimate for important beam directions.

In the proposed method, the number of segments in the plan is controlled, but we cannot directly affect the number of MUs. (All we know is that the number of MUs increases throughout the sequence of plans.) Trying to incorporate monitor units into the solution process, either directly by including them into the optimization problem, or indirectly by controlling the size of the segments, are topics for future work. Another future research direction is to tailor the shortest-path formulation to the MLC used. Different graph constructions and penalizing schemes for different collimators might improve the quality of the generated segments. Reducing the overall optimization time and including hard constraints are other topics to explore. A more mathematical description of the method will be presented in a forthcoming paper.

The flexibility of generating segments throughout the solution process combined with the ability to alter the segment shapes in the DSS optimization gives a promising method for generating simple yet satisfactory step-and-shoot plans. The method has potential for supporting the planner in finding satisfactory plans with fewer segments than what is traditionally used in step-and-shoot radiotherapy.

6. Acknowledgements

I thank Anders Forsgren for many helpful discussions on the mathematical aspects of this problem. I also thank Göran Sporre and Henrik Rehbinder for their careful reading of the manuscript and a number of helpful comments. The inspiring comments from Edwin Romeijn and Steve Webb are highly appreciated. Finally, thanks to Tim Craig for clinical input.

References

- [1] A. Ahnesjö. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Medical Physics*, 16(4):577–592, 1989.
- [2] M. Alber, G. Meedt, F. Nüsslin, and R. Reemtsen. On the degeneracy of the IMRT optimization problem. *Medical Physics*, 29(11):2584–2589, 2002.
- [3] J. L. Bedford and S. Webb. Constrained segment shapes in direct-aperture optimization for step-and-shoot IMRT. *Medical Physics*, 33(4):944–958, 2006.
- [4] A. M. Bergman, K. Bush, M.-P. Milette, I. A. Popescu, K. Otto, and C. Duzenli. Direct aperture optimization for IMRT using Monte Carlo generated beamlets. *Medical Physics*, 33(10):3666–3679, 2006.
- [5] N. Boland, H. W. Hamacher, and F. Lenzen. Minimizing beam-on time in cancer radiation treatment using multileaf collimators. *Networks*, 43(4):226–240, 2004.

-
- [6] F. Carlsson and A. Forsgren. Iterative regularization in intensity-modulated radiation therapy optimization. *Medical Physics*, 33(1):225–234, 2006.
- [7] F. Carlsson, A. Forsgren, H. Rehbinder, and K. Eriksson. Using eigenstructure of the Hessian to reduce the dimension of the intensity modulated radiation therapy optimization problem. *Annals of Operations Research*, 148:81–94, 2006.
- [8] C. W. Cheng, I. J. Das, and M. S. Huq. Lateral loss and dose discrepancies of multileaf collimator segments in intensity modulated radiation therapy. *Medical Physics*, 30(11):2959–2968, 2003.
- [9] C. Cotrutz and L. Xing. Segment-based dose optimization using a genetic algorithm. *Physics in Medicine and Biology*, 48(18):2987–2998, 2003.
- [10] D. Craft, P. Süß, and T. Bortfeld. The tradeoff between treatment plan quality and required number of monitor units in intensity-modulated radiotherapy. *International Journal of Radiation Oncology, Biology, Physics*, 67(5):1596–1605, 2007.
- [11] W. D. Gersem, F. Claus, C. D. Wagter, B. V. Duyse, and W. D. Neve. Leaf position optimization for step-and-shoot IMRT. *International Journal of Radiation Oncology, Biology, Physics*, 51(5):1371–1388, 2001.
- [12] E. Hall. Intensity-modulated radiation therapy, protons, and the risk of second cancers. *International Journal of Radiation Oncology, Biology, Physics*, 65(1):1–7, 2006.
- [13] Z. Jiang, M. A. Earl, G. W. Zhang, C. X. Yu, and D. M. Shepard. An examination of the number of required apertures for step-and-shoot IMRT. *Physics in Medicine and Biology*, 50(23):5653–5663, 2005.
- [14] Y. Li, J. Yao, and D. Yao. Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy. *Physics in Medicine and Biology*, 48(20):3353–3374, 2003.
- [15] J. Löf. *Development of a general framework for optimization of radiation therapy*. PhD thesis, 2000, Stockholm University.
- [16] X. Mu, P.-O. Löfroth, M. Karlsson, and B. Zackrisson. The effect of fraction time in intensity modulated radiotherapy: theoretical and experimental evaluation of an optimisation problem. *Radiotherapy Oncology*, 68:181–187, 2003.
- [17] A. Niemierko. A generalized concept of equivalent uniform dose EUD. *Medical Physics*, 26:1100, 1999.
- [18] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization*. Dover, New York, 1998. ISBN 0-486-40258-4.
- [19] F. Preciado-Walters, M. P. Langer, R. L. Rardin, and V. Thai. Column generation for IMRT cancer therapy optimization with implementable segments. *Annals of Operations Research*, 148:65–79, 2006.
- [20] H. Rehbinder, J. Andersson, S. Ericson, K. Eriksson, C. Forsgren, B. Hårdemark, E. Korevaar, M. Larsson, A. Liander, A. Lundin, J. Riddersporre, D. Rosseau, C. Sundgren, J. Uhrdin, and J. Löf. A general software framework for investigations in radiation therapy planning. *Medical Physics*, 33(6):2273, 2006.
- [21] A. V. Riet, C. Mak, M. Moerland, L. Elders, and W. van der Zee. A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate. *International Journal of Radiation Oncology, Biology, Physics*, 37(3):731–736, 1997.
- [22] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, and A. Kumar. A column generation approach to radiation therapy treatment planning using aperture modulation. *SIAM Journal on Optimization*, 15(3):838–862, 2005.
- [23] M. B. Sharpe, B. M. Miller, D. Yan, and J. W. Wong. Monitor unit settings for intensity modulated beams delivered using a step-and-shoot approach. *Medical Physics*, 27(12):2719–2725, 2000.
- [24] D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu. Direct aperture optimization: A turnkey solution for step-and-shoot IMRT. *Medical Physics*, 29(6):1007–1018, 2002.
- [25] J. Tervo, P. Kolmonen, T. Lyyra-Laitinen, J. Pintér, and T. Lahtinen. An optimization-based approach to the multiple static delivery technique in radiation therapy. *Annals of Operations Research*, 119:205–227, 2003.
- [26] S. Tung, M. Lii, P. Lai, P. Wong, B. Mason, A. Garden, K. Chao, and X. Zhu. Clinical evaluation of direct machine parameter optimization algorithm for head and neck IMRT treatment. *Medical Physics*, 32(6):1971, 2005.
- [27] J. Wang, X. Li, W. D’Souza, and R. Stewart. Impact of prolonged fraction delivery times on tumor control: A note of caution for intensity-modulated radiation therapy (IMRT). *International Journal of Radiation Oncology, Biology, Physics*, 57(10):543–552, 2003.
- [28] M. Zhang, V. Moiseenko, and M. Liu. PTV margin for dose escalated radiation therapy of prostate cancer with daily on-line realignment using internal fiducial markers: Monte carlo approach and dose population histogram (DPH) analysis. *Journal of Applied Clinical Medical Physics*, 7(2):38–49, 2006.