Fast approximate delivery of fluence maps for IMRT and VMAT

Marleen Balvert\textsuperscript{1} and David Craft\textsuperscript{2}

\textsuperscript{1} Department of Econometrics and Operations Research/Center for Economic Research (CentER), Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands
\textsuperscript{2} Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

E-mail: m.balvert@tilburguniversity.edu

Received 29 June 2016, revised 16 December 2016
Accepted for publication 4 January 2017
Published 23 January 2017

Abstract

In this article we provide a method to generate the trade-off between delivery time and fluence map matching quality for dynamically delivered fluence maps. At the heart of our method lies a mathematical programming model that, for a given duration of delivery, optimizes leaf trajectories and dose rates such that the desired fluence map is reproduced as well as possible. We begin with the single fluence map case and then generalize the model and the solution technique to the delivery of sequential fluence maps. The resulting large-scale, non-convex optimization problem was solved using a heuristic approach. We test our method using a prostate case and a head and neck case, and present the resulting trade-off curves. Analysis of the leaf trajectories reveals that short time plans have larger leaf openings in general than longer delivery time plans. Our method allows one to explore the continuum of possibilities between coarse, large segment plans characteristic of direct aperture approaches and narrow field plans produced by sliding window approaches. Exposing this trade-off will allow for an informed choice between plan quality and solution time. Further research is required to speed up the optimization process to make this method clinically implementable.

Keywords: fluence map, sliding window, VMAT, optimization, dMLC, dose rate

(Some figures may appear in colour only in the online journal)
1. Introduction

The fast delivery of a fluence map (also sometimes referred to as an intensity map) has received some attention over the years (Crooks et al. 2002, Engel 2005, Luan et al. 2008) but the following remains an unresolved problem in general: for a given allotted delivery time, determine a set of leaf trajectories and dose rates versus time to best recreate a given fluence map, given machine characteristics such as dose rate restrictions and maximum leaf speed. Since the efficient delivery of fluence maps is at the heart of intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT), we have returned to this basic question in order to improve IMRT and in particular, VMAT.

Fluence map delivery by a multi-leaf collimator (MLC), the defining hardware component of IMRT and VMAT, is done in either a step-and-shoot or dynamic fashion. In step-and-shoot delivery, the MLC leaves are moved into position while the beam is off, then the beam is turned on for delivery, and then this process is repeated for each segment shape to deliver (Xia and Verhey 1998). In dynamic delivery, the beam is on while the leaves are moving, painting out the fluence map (Yu 1995). In VMAT radiation is delivered in a dynamic fashion while simultaneously moving the gantry around the patient. This allows for a faster delivery than IMRT without compromising the plan quality (see, e.g. Teoh et al. 2011 and references therein). Even faster deliveries can be achieved for both IMRT and VMAT when accepting sub-optimal treatment plans. The trade-off between treatment duration and plan quality thus lies at the heart of IMRT and VMAT treatment planning (Unkelbach et al. 2015).

The starting point of this work is captured by the simple question: given a fluence map and a fixed delivery time, what is the best we can do in terms of matching that fluence map? For the case of infinite leaf speed and assuming the leaves can move across the field with no tip gap (i.e. fully closed), the sliding window algorithm with maximum dose rate is known to be optimal (Bortfeld et al. 1994, Stein et al. 1994). For finite leaf speed a sliding window approach will yield perfect fluence map replication (Stein et al. 1994), though at the cost of high delivery time and will thus often be suboptimal. For example, for a uniform field, the optimal delivery will be to set the leaves open at the field boundaries and irradiate at maximum dose rate. Likewise, for row-wise unimodal fields, a close-in (or open-out) technique will be faster than a sweeping window technique.

These simple ideas, while useful for thinking about optimal dynamic delivery, are not enough to solve the fluence map delivery problem in general. The difficulty is that the rows of a fluence map, while independent from each other regarding leaf motions (assuming that interdigitation is allowed and that we do not try to reduce tongue-and-groove effect), are coupled via the time-varying dose rate, which applies simultaneously to the entire field. An optimal dose rate versus time for one leaf row considered in isolation will in general be unique to that row. We therefore model the problem as a constrained optimization problem and solve for all leaf motions and the dynamic dose rate simultaneously. Solving the model for a range of delivery times provides the desired trade-off curve.

The aim of this work is to develop a method to construct the trade-off curve for dynamically delivered IMRT fields and full arc VMAT treatments and allow the planner to balance treatment time and plan quality. This has, to the best of our knowledge, not been done before. We are the first to optimize the leaf trajectories and dose rates in a single smooth optimization problem.
2. Materials and methods

2.1. Treatment planning model

To isolate the treatment planning problem for VMAT to fluence map delivery, we assume the standard two-step approach to IMRT planning. For this, the 360° arc is divided into $M$ arc segments. For each segment the optimization problem is solved by first optimizing for the fluence maps and then applying a leaf sequencing algorithm for the delivery of those fluence maps (Jelen et al 2005). In order to not deviate from the focus of this work, which is to replicate a given fluence map, we ignore the first optimization and simply assume that the arc segments and fluence maps are given as a result of an IMRT optimization. For an elaborate discussion of fluence map optimization, we refer to Webb (1989), Bortfeld (2006) and Ehrgott et al (2010). We also assume that the treatment delivery device, a linear accelerator (linac) equipped with an MLC, has a continuously variable dose rate and continuously variable MLC leaf speed, both up to some maximum level. We do not model the jaws of the linac and we ignore leaf transmission. We also assume leaves can shut fully and move in this fully shut position. In the Discussion section we describe how to include these features.

Let $f_{ij}^m$ be the fluence map for arc segment $m$ we are trying to produce. The index $i$ indexes the leaf rows and $j$ indexes the columns (position along each row). We assume that the bixels across a row are indexed 1, 2, …, $B$. Physical floating point leaf positions aligned with these bixel locations are such that ‘1’ (e.g. centimeter) on the physical ruler corresponds with the left side of bixel with index $j = 1$. Therefore, physical leaf positions for the left and right leaves range from 1 to $B + 1$, see figure 1.

The key function used to map leaf positions to the fluence map produced is the exposure function $e$, which gives the exposure (a number between 0 and 1) of beamlet $j$ given the left leaf position is $L$ and the right leaf position is $R$. The same function is applicable to all the rows and maps, so there is no row and map index in the following:

$$e(L, R, j) = \begin{cases} 1 & \text{if } L \leq j \text{ and } R \geq j + 1 \text{ [fully exposed]}, \\ j + 1 - L & \text{if } j < L < j + 1 \text{ and } R \geq j + 1 \text{ [fully exp. by right, part blocked by left]}, \\ R - j & \text{if } L \leq j \text{ and } j < R < j + 1 \text{ [fully exp. by left, part blocked by right]}, \\ R - L & \text{if } j < L < j + 1 \text{ and } j < R < j + 1 \text{ [partially blocked by both]}, \\ 0 & \text{if } L \geq j + 1 \text{ or } R \leq j \text{ [fully blocked by one of the leaves]} \end{cases} \quad (1)$$

We choose a fixed discretization for time, $\Delta$. For example $\Delta = 1/3$ of a second. Let $L_i^t$ be the position of the $i$th leaf end at time step $t$, and $R_i^t$ is likewise defined. Note that if $L_i^t = R_i^t$, the leaves are closed. Let $D_t$ be the dose rate at time $t$ (dose rate units are MU s$^{-1}$, where MU stands for monitor units). The fluence map $g_{ij}^m$ (in MU) delivered to arc segment $m$ obtained by a set of leaf trajectories and dose rates is then given by:

$$g_{ij}^m = \sum_{t=s_m}^{e_m} e(L_i^t, R_i^t, j)D_t \Delta \quad (2)$$

where $s_m$ and $e_m$ are respectively the first and the last time step the gantry spends in arc segment $m$. Note that $s_1 = 1$, $s_m = e_{m-1} + 1$ and $e_M = T$, where $M$ is the total number of fluence maps.
The optimization problem of matching the fluence maps \( f^m, m = 1, \ldots, M \), is given by:

\[
\min \sum_{m} \sum_{i} \sum_{j} (f^m_{ij} - g^m_{ij})^2
\]

subject to:

\[
g^m_{ij} = \sum_{t=T} e(L^t_i, R^t_i, j)D_t \Delta
\]

\[
L^t_i \leq R^t_i, \quad \forall \ t, \ i
\]

\[
L^t_i - c \leq L^t_{i+1} \leq L^t_i + c, \quad \forall \ t = 1 \ldots T - 1, \ i
\]

\[
R^t_i - c \leq R^t_{i+1} \leq R^t_i + c, \quad \forall \ t = 1 \ldots T - 1, \ i
\]

\[
L^t_i \geq 1, \quad \forall \ t, \ i
\]

\[
R^t_i \leq B + 1, \quad \forall \ t, \ i
\]

\[
0 \leq D_t \leq D_{\text{max}}, \quad \forall \ t
\]

(3)

where \( c \) is a constant reflecting the maximum leaf speed constraint, and \( e \) is the function given in (1). The \( L, R \) constraints are, in the order they appear: the left leaves must stay to the left of the right leaves, the left leaves cannot travel more than \( c \) cm per time step (a typical maximum leaf speed is 3 cm s\(^{-1}\), so, with a time step of \( 1/3 \) s, \( c = 1 \)), same for the right leaves, the left leaves should never go to the left of position 1, the right leaves should never go to the right of position \( B + 1 \), and the dose rate at each time step should be non-negative and no larger than the maximum dose rate \( D_{\text{max}} \). Note that we allow leaves to move back and forth rather than enforcing unidirectional motion across the field.

Our model allows for non-unidirectional leaf sweeps. Allowing for only unidirectional leaf sweeps simplifies the problem, however unidirectional leaf motions can be suboptimal, which we prove in appendix A.

2.2. Four step solution approach

Solving formulation 3 raises several issues:

(i) The model requires the amount of time spent on each arc segment as an input, whereas only the total delivery time for the full arc would be specified by a treatment planner.
Formulation 3 is a large non-convex optimization model, and thus has a large number of local minima. The non-convexity is due to the non-convex mapping between leaf positions and beamlet exposure and the multiplication of the dose rate and leaf position variables.

We adopt a four-step heuristic approach to overcome these issues. First, the model is solved for each individual arc segment over a range of delivery times by computing, for each map and for each delivery time, many local optima and selecting the best solution among those. This provides a plan quality versus delivery time trade-off curve for each individual arc segment. In the second step, we compute, based on these trade-off curves, the delivery time that will be allocated to each arc segment. This provides a solution to the first issue mentioned above. In the third step the full arc treatment plan is optimized. Note that the second issue is still valid. In order to split the problem into multiple smaller problems, we fix the dose rates that were optimal in the individual arc segment optimization. By doing so the problem decouples across the leaf rows, which greatly reduces the model size. The resulting treatment plan is fine-tuned in the fourth step by iteratively optimizing the dose rates for fixed leaf trajectories, and optimizing the leaf trajectories for fixed dose rates. Each of the four steps is further explained below.

2.3. Step 1: treatment plan optimization for a single arc segment

In the first step, the trade-off between plan quality and delivery time for each arc segment with its corresponding fluence map is constructed using formulation 3 for a single arc segment only. The non-convexity implies the existence of local minima and therefore a global optimization procedure is needed. There is no finite time algorithm that solves general non-convex optimization problems to proveable optimality. Gradient descent methods find local optimal solutions. We run gradient-based minimizations at a diverse set of starting solutions and choose the best overall solution. For the local minimizations, we use the Matlab (The MathWorks, Natick MA, version 7.14) function fmincon with the default interior-point method and a user supplied gradient. The gradient computation is given in appendix B.

2.3.1. Generating diverse starting solutions. For generating diverse starting feasible solutions, we randomize on the following trajectory types:

- Type 1: Left to right leaf sweep
- Type 2: Right to left leaf sweep
- Type 3: Close-in trajectory
- Type 4: Open-out trajectory
- Type 5: Random leaf motion trajectory
- Type 6: For each row, choose one of the above independently of the other rows.
- Type 7: Left to right for large delivery time rows, close-in otherwise

Each of the above trajectories is randomly generated. For example, to generate a random feasible left to right trajectory, we flip a coin at each time step to determine if a leaf will advance to the right at maximum speed or stay still. One could use other distributions to randomly generate leaf steps as well. We ensure feasibility by making sure leaves do not collide or extend out of range. Dose rates are set at $D_{\text{max}}$.

2.3.2. Speeding up the optimization by focusing on tough rows. The fluence often varies strongly among the rows of a fluence map, and for some rows it can be much more difficult to obtain the desired fluence than for others. This difference can be so large that for some
rows a leaf trajectory that gives exactly the desired fluence can easily be found for any sequence of dose rates that is suitable for the tougher rows. Consider the fluence map in figure 2 as an example. Rows 7 and 8 have positive fluence over the full width of the row as well as a high total fluence, which means that they require a relatively large amount of delivery time in order to replicate the map properly. On the other hand, row 11 has a positive and low fluence in only two bixels. The amount of MUs required to deliver the fluence of rows 7 and 8 is thus much higher than what is required for row 11 and any sequence of dose rates that is suitable for rows 7 and 8 will suffice for the fluence in row 11 to be perfectly replicated.

The above observation allows us to decompose the solution approach into two steps. First, the dose rates and leaf positions are optimized for a reduced fluence map, from which the rows with a fluence that can easily be replicated are removed. This greatly reduces the number of variables, which is especially useful for non-convex optimization problems, and thus improves the solution times. In the second step, the dose rates are fixed to the values found in the first step, and the leaf trajectories for each of the easy rows are optimized individually. These are small and simple optimization problems for which only a small number of random starting points are required, so little time needs to be spent on finding a good solution. Decoupling the procedure in this manner cuts down the solution time by a factor of about three for maps of the size shown in figure 2 and an even larger factor for larger maps.

All rows with a very low total fluence compared to the other rows are considered easy. We define ‘very low’ as any total fluence below 10% of the maximum total fluence over all rows. Of the remaining rows, those with a high sum of positive gradients (SPG, (Craft et al 2007)) are considered to be tough, and the others easy. The SPG is the sum of all fluence increments over a row. If the SPG of a row is larger than the mean SPG over all rows, it is considered a difficult row. SPG is a good indicator of row delivery complexity since delivery time for a row using the sliding window technique increases linearly with SPG (Craft et al 2012).

2.4. Step 2: delivery time allocation

The trade-off curves for the individual arc segments obtained in step 1 indicate how complicated the replication of each of the fluence maps is. We use this information to allocate the total delivery time over the arc segments in such a way that, if we were able to optimize for each of the fluence maps individually, the ssdif (ssdif refers to the objective function in formulation 3, the sum of the squared differences between desired and achieved fluence maps) summed over all arc segments is minimized. For this we solve the following small integer programming problem:
\[
\begin{align*}
\min & \quad \sum_{m \in M, t \in T} v_{mt} x_{mt} \\
\text{s.t.} & \quad \sum_{m \in M, t \in T} t x_{mt} \leq T T \\
& \quad \sum_{t \in T} x_{mt} = 1 \quad \forall m \in M \\
& \quad x_{mt} \in [0, 1].
\end{align*}
\]

Here, \( x_{mt} \) is a binary variable that is equal to 1 if \( t \) units of time are assigned to arc segment \( m \), and 0 otherwise. Furthermore, \( v_{mt} \) is the minimum ssdif for individual optimization of arc segment \( m \) when its delivery time is \( t \), which is known from step 1. \( M \) is the set of arc segments, and \( T \) is the set of delivery times for which the single arc segment model was solved in step 1. The total delivery time for the full arc is limited by \( T T \) in the first constraint, where the parameter \( t \) is equal to the subscript \( t \) of \( x_{mt} \). Note that the optimal value of (4) is a lower bound for the true ssdif (the real solution will be worse due to leaf position matching constraints). The time allocation can then be used to determine \( s_{m} \) and \( e_{m} \) for each map. We note that there is no reason to consider times for each sector that are lower than the time it takes the gantry to sweep out the arc sector at maximum speed. A typical maximum gantry speed is 1 revolution per minute, which is equivalent to sweeping a 10-degree sector in 1.67 s.

### 2.5. Step 3: fluence map optimization for the full arc

With the time allocation obtained in step 2, we can solve formulation (3) in the third step. However, as noted before, the resulting model is too large to solve. A natural idea to decompose the problem would be to solve it for each arc segment separately. This requires fixing the start and end positions of the leaves to allow for a smooth transition from one arc segment to the next, as proposed in Chen et al. (2011). Fixing the leaf start and end positions imposes restrictions on what trajectory types can be used, leaving little freedom for the optimization problem. We have tested several decision rules for fixing the leaf positions at the transition points, but none of them produced treatment plans that yield fluence maps similar to the desired ones.

Instead of decomposing the problem into separate optimizations for each arc segment, one can decouple the rows and optimize the trajectories of a single leaf pair for the full arc at once. Although the leaf trajectories of the leaf pairs are mechanically independent (assuming that interdigitation is allowed), the dose rate variable couples the individual rows since the dose rate applies to the entire field. Therefore, in order to allow for separate optimization of the leaf trajectories, we fix the dose rates to those that were optimal in step 1 (from the particular plans selected in step 2).

As in the single map case, the many local minima of this highly non-convex model can be identified using an interior-point method for various starting solutions. Using the same approach of generating random starting solutions however would result in a lack of freedom: if the first map has a left–right trajectory, the leaves end up all the way to the right and the next trajectory can only be a right–left sweep or a random trajectory. In order to allow for more flexibility, each of the arc segments is subdivided into one (so not subdivided), two or three subsegments of random length, for which a trajectory is generated according to a randomly selected trajectory type. The number of subsegments is randomly selected as well. An example of random leaf trajectories with three arc segments is given in figure 3. Like in step 1, we perform interior point minimizations from many of these randomly generated trajectories and select the best local optima as our solution.
2.6. Step 4: Alternate improvement of dose rate schedule and leaf trajectories

The treatment plans obtained in step 3 have leaf trajectories that are optimized with respect to a given dose rate schedule. However, the current dose rate schedule is not necessarily optimal for the current leaf trajectories. Therefore, we re-optimize the dose rates by fixing the leaf trajectories, for which a quadratic programming problem needs to be solved. As the current leaf trajectories may not be optimal for the new dose rate schedule, we then re-optimize the leaf trajectories using the approach in step 3, though with only one start solution, namely the current leaf trajectories. This process is repeated until the improvement per iteration drops below 1%.

2.7. Test cases

We demonstrate the method on two fluence map sets, both generated from the data publicly available via the CORT data set (Craft et al 2014). We use the same fluence map sets for both the single map case and the VMAT case. The first fluence map set is from the prostate patient with lymph nodes and the second is from the head and neck patient. We assume a maximum leaf speed of 3 cm s\(^{-1}\) and a maximum dose rate of 10 MU s\(^{-1}\). The bixel width is 1 cm for the prostate case and 0.5 cm for the head and neck case, and the arc is divided into 36 arc segments for both cases. Fluence maps are obtained by solving a dosimetric convex optimization, see e.g. Chen et al (2012) (the fluence map optimization problem has been described many times; so as to not distract from the fluence map matching problem, we intentionally do not discuss this problem. We do however make the fluence maps publically available). Time steps of 1/3 s are used for the prostate case. Since the fluence maps for the head and neck case are larger and...
thus result in a larger model, time step size was chosen to be 1/2 s to slightly ease the computational burden. The model was solved using the fmincon function in Matlab Release R2014a (The Mathworks, Inc., Natick, USA). Below, we discuss the results for the single map case (step 1) and the full arc case (steps 2–4) separately.

3. Results

3.1. The single map case

For each fluence map we solve the complete fluence matching problem (i.e. generate many starting solutions and minimize each one using fmincon, then pick the overall best objective value) for several values of $T$ in order to generate the trade-off curve of delivery time and fluence map matching quality. An upper bound on the maximum time needed for perfect fluence map delivery is the maximum row delivery time for the leaf-sweep algorithm. The row delivery time is given as (time to sweep the leaves across the row at maximum leaf speed) + (SPG time), where (SPG time) is the SPG of the row (in MU) divided by the maximum dose rate (to convert the units into time) (Craft et al 2012). For the prostate case this value is 6.3 s, and for the head and neck case it is 10.3 s.

Figure 4 shows the delivery time versus fluence map matching trade-off curves, where the solid and dotted lines represent the treatment plans with variable dose rates and dose rates fixed at 10 MU s$^{-1}$, respectively. As is typical in IMRT smoothing studies (Alber and Niusslin 2000, Craft et al 2007, Matuszak et al 2008), we see that the fluence map can be delivered much faster with only negligible (indeed, barely visible) degradation. Dose rates are seen to fluctuate in most cases: they generally drop to move leaves from one place to another without depositing dose. Only for the smallest total allotted time case the dose rate remains at its maximum level in order to get enough fluence through.

It is interesting to note that if given sufficient time, a fluence map can be perfectly delivered with a sliding window technique at constant maximum dose rate, but one can also achieve a near perfect match with a fluctuating dose rate.

For short delivery time solutions, where using the maximum dose rate is optimal, there is no difference between the solutions with variable and fixed dose rates, and for the large delivery time solutions, where sliding window with maximum dose rate is also known to be an exact solution, there is no difference either.

The head and neck case fluence maps are larger since the target is larger and the beamlets are smaller. A typical head and neck map from this data set has dimensions 40 leaf pairs $\times$ 44 columns. Figure 5 depicts the trade-off curve between delivery time and fluence map matching quality for a head and neck case fluence map, with the original map being recreated shown in the top right corner. The optimized solutions, like for the prostate case, all show variable dose rates except for the smallest delivery time solution which uses the maximum dose rate the entire time.

Figure 6 shows two optimal leaf trajectories for different total time allotted solutions from the trade-off curve. The fluence row that is chosen is a difficult row with three prominent peaks and smaller peaks between those. The 5 s solution, with the resulting fluence given by the dotted line, is clearly not enough time to replicate these peaks. The 11 s solution does very well, only slightly truncating the largest peak and approximating the small leftmost fluence peak with a flatter version. For the 11 s solution, the left leaf moves leftward, rightward, and then leftward again (i.e. the leaf trajectory is not unidirectional), and the right leaf uses a similar pattern. Both the 5 s and the 11 s solutions zero the dose rate towards the end in order to position the leaves for a final surge of radiation.
We investigated the sensitivity of the optimal solution for the head and neck case to the leaf speed. Doubling the leaf speed from the default value of 3 cm s$^{-1}$ to 6 cm s$^{-1}$ produces a dramatic improvement in the solution, with the objective function value (the sum of squared differences) dropping from 2300 to 500. Further increasing the leaf speed to 9 cm s$^{-1}$ offers only marginal improvement to 460. Increasing the dose rate above the nominal value does not improve the solution at all in this regime, as can be gleaned from noting in the bottom of figure 5 that for total time $= 7$ s, the default dose rate is not an active constraint (the dose rate for this solution, while allowed to be as high as 10 MU s$^{-1}$, rarely goes above 5 MU s$^{-1}$). If the fluence map was globally scaled up (for example in a hypofractionated case where much more dose is delivered per treatment), then the maximum dose rate would become an influential parameter.

3.2. The full arc case

Leaf trajectories and dose rates are optimized for total delivery times varying from 60 to 180 s with step sizes of 15 s for the prostate case, and from 75 to 345 s with step sizes of 90 s for the head and neck case. For both cases, we show the time versus fluence map resemblance (ssdif) trade-off curves (figures 7 and 11). The dotted line shows the total ssdif if the treatment plans for each arc segment were optimized individually, which coincides with the optimal objective values of the time allocation problem in step 2. This is a lower bound for the actual full arc.
trade-off, as generating a full arc treatment plan is more restrictive than optimizing the arc segments individually. The solid and dashed lines show the results for the full arc treatment plan optimization using optimized dose rate schedules and constant maximum dose rates (i.e. all dose rates equal to 10 MU s$^{-1}$), respectively. For these, the gray and black lines represent the trade-off curves obtained at steps 3 and 4, respectively. This implies that the dashed black line represents treatment plans where the dose rates are variable, but which were obtained using the plans corresponding to the gray dashed line (with dose rates equal to 10 MU s$^{-1}$) as a starting point for step 4.

In addition to the trade-off curve, we show the fluence maps corresponding to one arc segment for each of the cases (figures 8 and 12). This allows for comparison of the fluence maps obtained with individual arc segment optimization (first row), full arc optimization with optimized dose rates obtained at steps 4 and 3 (second and third row, respectively) and full arc optimization with constant maximum dose rates (fourth row). Note that the latter corresponds to the dashed gray line in the trade-off curves. Each of the columns corresponds to a different full arc delivery time ($TT$). The maps in the first row are optimized using the time allocated to this arc segment for the full arc delivery time corresponding to that column, namely for 6s, 9s, 15s, 18s and 18s for the prostate case, and 4s, 6s, 16s and 20s for the head and neck case. The corresponding ssdifs are indicated below each of the maps. Note that an increase in $TT$ may yield a lower fluence map resemblance for an individual arc segment. This is because quality of reproduction of an individual arc segment may be sacrificed for an overall gain across all maps.

3.2.1. Prostate case. The trade-off curve for the prostate case (figure 7) has a convex shape. This implies that for short delivery times, much can be gained by lengthening the treatment, whereas this is less so for treatments of average length. The same trend is visible in the fluence maps (figure 8).
The limitations of a constant maximum dose rate are evident from figure 8: the gains from applying step 4 to the plans with constant dose rate are large. Furthermore, in order to achieve the same level of fluence map resemblance with constant dose rates as can be obtained with variable dose rates, the total delivery time needs to be increased, sometimes even by 45 s. Furthermore, allowing for a varying dose rate in step 3 gives better solutions in step 4.

The level of fluence map replication shown in figures 7 and 8 reflects the quality of DVH replication, as can be seen from figure 9. The DVHs for a total delivery time of 150 s and 180 s are slightly worse than those for the sliding window delivery (for which 240 s of delivery time are required), while the DVH for a delivery time of 60 s shows a major deterioration.

Figure 10(a) shows the leaf trajectories for treatment durations varying from 60 (left-most plot) up to 180 s (right-most plot) with 30 s increments. In the 1 min treatment plan the leaves are generally either (almost) completely closed or they leave a large opening to deliver dose to many bixels at the same time. For the latter there is often a small close-in or open-out movement visible that results in the delivery of a peak-like fluence. As delivery time increases, the opening between the leaves generally shrinks and the trajectories become more similar to a sliding window approach. For a fixed dose rate, plans converge to the sliding window approach even faster (figure 10(b)).

Note that even though the leaves remain closed in some arc segments (see figures 10(a) and (b)) as no fluence needs to be delivered here, a rather large amount of time is allocated to this arc segment. This is because another row of that fluence map requires a large amount of time for delivery.
3.2.2. Head and neck case. While the trade-off curve for the prostate case is convex, it is nearly linear for the head and neck case (figure 11). The benefits from allowing for a varying dose rate are even more pronounced than for the prostate case: the delivery times can be reduced by up to 100s when using varying instead of constant dose rates, without reducing ssdif. Additionally, treatment plans with constant dose rates do not provide a good starting solution for step 4: the dashed black trade-off curve lies far above the trade-off curve obtained at step 3 when allowing for varying dose rates.

Figure 12 shows the fluence maps for arc segment 15 obtained with four different optimization approaches. The maps corresponding to a low delivery time ($TT = 150s$ and $TT = 330s$, where respectively 4s and 6s are assigned to this map) show the different choices the optimization model makes when using a variable versus a constant maximum dose rate. The short delivery times limit the number of bixels a leaf can traverse. As a result, the leaves can either be placed close to the boundaries of the field to deliver fluence to a large part of the row, or they can be located close to one another, which implies that they block many of the bixels. In the case of a variable dose rate, the model chooses the first option and turns the dose rate down in order to avoid an excessively high amount of fluence to the bixels that are exposed. However, when fixing the dose rate to its maximum, one cannot avoid this excessively high amount of fluence, so the model is forced to place the leaves closer to each other, thus blocking out a large part of the row. This also becomes apparent from figures 13(a) and (b), showing the leaf trajectories for row 20 using optimized and constant maximum dose rates, respectively.

4. Discussion

In this article we have developed a method to construct the delivery time versus fluence map fidelity trade-off curve for VMAT. The core of our method is an optimization model that
optimizes the leaf trajectories and dose rates for a given delivery time such that the optimal fluence maps are replicated as well as possible. Solving this optimization model for a sequence of delivery times gives the complete trade-off curve, and allows the planner to balance delivery time and fluence map resemblance.

**Figure 8.** Fluence maps for arc segment 11 of the prostate case, optimized using a total arc delivery time \( TT \) of 60s, 90s, 120s, 150s and 180s. Fluence maps obtained with individual arc segment optimization are optimized with delivery times equal to the time allotted to this arc segment in the full arc treatment plans (6s, 9s, 15s, 18s and 18s, respectively). The corresponding ssdif is indicated below each map.

**Figure 9.** DVH plots for the PTV (black) (PTV_56 from the CORT data set) and rectum (gray) for the plans obtained with various total times \( TT \), in seconds) and the sliding window approach.
The trade-off curves for the prostate case presented in this article show that the gain in fluence map resemblance from an additional second of delivery time decreases as the total delivery time increases. More importantly, using a shorter delivery time than what is required for the sliding window approach yields only a minor reduction in plan quality. A further reduction may result in too low fluence map resemblance. For the head and neck case, there seems to be a more linear relation between delivery time and fluence map resemblance.

**Figure 10.** The leaf trajectories for row 8 of the prostate case using variable (a) and constant maximum dose rates (b) for treatment durations from 1 (left-most plot) up to 3 min (right-most plot) with 30 s increments. The horizontal lines demarcate the time allotted to each arc segment.

**Figure 11.** Trade-off between delivery time and ssdif for the head & neck case. Results obtained at step 3 with optimized dose rates (solid) and dose rates equal to 10 MU s\(^{-1}\) (dashed) are presented in gray. Improvements from these results obtained in step 4 are indicated in black. The lower bound based on individually optimized arc segments is shown by the dotted line.

The trade-off curves for the prostate case presented in this article show that the gain in fluence map resemblance from an additional second of delivery time decreases as the total delivery time increases. More importantly, using a shorter delivery time than what is required for the sliding window approach yields only a minor reduction in plan quality. A further reduction may result in too low fluence map resemblance. For the head and neck case, there seems to be a more linear relation between delivery time and fluence map resemblance. Construction of
the trade-off curve allows the planner to obtain these insights and choose the preferred balance between the delivery time and the level of fluence map replication.

While the majority of the current literature assumes a constant dose rate, we include optimization of the dose rate scheme as well. Comparing the trade-off curves of plans with a varying dose rate to those with a constant maximum dose rate shows that the same level of plan quality can be achieved with shorter delivery times when using variable dose rates. A similar conclusion was found by Palma et al (2008). It is also not sufficient to optimize the dose rates as a final step: optimizing dose rates already as a part of step 3 leads to better final plans.

Several other works have looked into the conversion of fluence maps to a deliverable treatment plan. Some of these use a sliding window delivery since this allows the maps to be replicated to high accuracy (Craft et al 2012, 2014), though it inevitably results in a high treatment time as the leaves have to traverse the whole field for each fluence map. This has led to the development of several arc sequencing algorithms that allow for other types of leaf trajectories as well. Shepard et al (2007) minimize the sum of absolute differences between the ideal and planned fluence maps using simulated annealing. The authors optimize the aperture shapes and dose rates at every $10^\circ$ (which can be chosen to be more dense). Their objective

Figure 12. Fluence maps for arc segment 15 of the head and neck case, optimized using a total arc delivery time ($TT$) of 150s, 330s, 510s and 690s. Fluence maps obtained with individual arc segment optimization are optimized with delivery times equal to the time allotted to this arc segment in the full arc treatment plans (4s, 6s, 16s and 20s, respectively). The corresponding ssdif is indicated below each map.
is to minimize the sum of absolute differences between the ideal and planned fluence maps. The optimization problem is similar to ours as both methods directly aim to minimize the difference between ideal and planned fluence maps, though the model is solved using a different solution method (simulated annealing versus an interior point method). Bokrantz (2012) also uses optimized fluence maps, but does not consider directly replicating these. Instead, his goal is to find a treatment plan that minimizes the difference between the DVHs corresponding to the optimized and the planned maps. A comparison of the performance of these methods is only possible when using the same set of patients.

Unkelbach et al (2015) mention several disadvantages of treatment planning via fluence maps compared to direct aperture optimization (DAO), which are mostly overcome by our method. First, the formulation of a DAO model exactly reflects the DICOM specifications and thus always yields a deliverable treatment plan. In our model, we use the same decision variables as are generally used in DAO approaches, and the resulting treatment plan can thus be described according to DICOM specifications. Second, Unkelbach et al (2015) state that a highly accurate replication of fluence maps is expected to yield inefficient treatment delivery. This is exactly the issue we are addressing here by constructing the delivery time versus replication accuracy trade-off curve. A third and final drawback is that the distribution of apertures over an arc sector inherent to arc sequencing generally results in a dose degradation. This drawback does hold for our method. As shown in Craft et al (2012), the degradation is minimal when arc sectors of 10° or less are used. We developed our work to bridge the gap between the coarse, large field, and fast delivery solutions typical of direct aperture approaches and the smaller field, higher monitor unit, and more time consuming sliding window approaches.

This work is the first to assess the trade-off between delivery time and fluence map resemblance for arc sequencing, but it is not the first to assess the general question of the trade-off between delivery time and plan quality in VMAT. Craft et al (2012) explore quality versus time for by optimizing fluence maps at a 2° angular spacing and then successively merging fluence maps based on their similarity. Each merge decreases treatment time but also decreases plan quality. In this approach, the fluence maps are replicated perfectly using sliding window delivery. This prompts two further ideas to achieve delivery time versus plan quality
trade-offs: either by adding a smoothing term to the original fluence map optimization—an approach that has recently been investigated by Gaddy and Papp (2016)—or by speeding up the delivery of the fluence maps by departing from strict sliding window, as we do herein. The optimal way to combine all of these approaches—merging, smoothing, and approximate delivery—remains to be worked out.

The main limitation of our method is the size of the full arc optimization model, which requires us to use a heuristic approach. This has several consequences. First, due to the highly non-convex nature of the model, we have no guarantee on the optimality of our solution. However, as the objective function value plateaus as the number of starting points for which the model is solved increases, we believe that the optimal ssdif found in step 4 lies close to the true optimal objective value of this model. Second, a large amount of time is required to solve the model, which makes it unsuitable for implementation in the clinic in its current form. The size of the problem increases with the size of the fluence maps and the total delivery time. Therefore, future research should aim at reducing the solution times. For this, one can take advantage of the highly parallelizable nature of the method. Furthermore, for some rows there are several consecutive maps where the fluence is (close to) zero, which may allow us to ignore those segments in the first optimization, and optimize their leaf trajectories in a post optimization. Lastly, heuristics to find better starting solutions, or other techniques to sample the global solution space, will reduce the number of local searches that need to be performed.

We chose to formulate the dynamic leaf sequencing problem as a continuous optimization problem since simpler approaches, such as manually determining leaf trajectories, proved too complex given a variable dose rate which couples all the leaf rows together. Since the resulting formulation is necessarily non-convex, due in part to the fundamental step-function like relationship of leaf position and fluence transmission to a bixel, the solution procedure needs to have a global search aspect. Global search algorithms such as particle swarm, simulated annealing, differential evolution, and genetic algorithms could be attempted, but in order to keep the focus on the problem being solved, and knowing that gradient descent is a useful strategy for smooth optimization problems, especially when an analytical gradient is available, we opted for a straightforward approach that could use this information. In order to gain confidence in the near-optimality of our approach, we let the random search and gradient descent optimizer run for days for certain problem instances to show that after a few hours (problem size dependent of course) the solution quality plateaus. Testing alternate global search strategies will be an interesting future study.

The procedure of step 1 is applicable to static beam IMRT performed with dynamic delivery. For treatment planning systems such as Monaco, RayStation, and Pinnacle that perform IMRT optimization as a two step process: (1) fluence map computation and (2) leaf sequencing and refinement, the procedure can be used as the leaf sequencing procedure either for a single fixed time per fluence map or to generate a set of delivery time/plan quality trade-off curves as shown in figures 4 and 5. Fluence map smoothing is also often incorporated into the fluence map optimization step, and how best to utilize both of these smoothing techniques (i.e. directly incorporating smoothing into the optimization and implicitly smoothing in the sequencing procedure described herein by choosing a small enough delivery time) is a topic for future investigation.

Restricting large sudden changes in the dose rates can be done with linear constraints on the dose rate variable. Incorporating leaf gaps into the optimization can be done by altering the inequality $L_i \leq R_i$ with a given minimum clearance leaf gap, say 1 millimeter. Maximum leaf tip separations can also be handled with linear constraints. If we assume that the jaws of the linac are fixed, leaf transmission can be accounted for by replacing the 0 term in the $e(L,R,j)$ equation, representing leaf blocking, with a transmission term. Moving jaws, either
independent from the moving leaves or serving as a carriage on which the leaves move relative to, can also be modeled, although the exposure function $e$ becomes more challenging.

However, given the general modeling approach taken here, provided that for a given set of jaw and leaf positions an exposure function can be written down, we believe most or all linacs can be modeled with this technique.

5. Conclusion

The trade-off between delivery time and fluence map replication accuracy lies at the heart of IMRT and VMAT treatment planning. In this work we have presented a method to construct the trade-off curve between these two conflicting objectives by optimizing treatment plans for various delivery times. Future research is needed to speed up the optimization process before it can be clinically implemented. Once this becomes possible, the treatment planner can use the trade-off curve to select a plan that balances delivery time and plan quality.

Appendix A. The need for non-unidirectional leaf trajectories

Several works on treatment plan optimization for VMAT only consider unidirectional leaf trajectories, i.e. leaves can only move from left to right or right to left (Chen et al 2011, Papp and Unkelbach 2014). If one uses a fixed dose rate, this is a suitable assumption, since any non-unidirectional leaf trajectory can be rewritten into a unidirectional trajectory without changing the fluence. In order to see this, consider the leaf trajectories in figure A1(a), where the dose rate is constant. The total fluence deposited at position $j$ is indicated by the dashed lines, and is thus given by $L_j^1 - R_j^1 + R_j^2 - L_j^2$, where $L_j^k$ and $R_j^k$ denote the time at which the left and right leaf, respectively, pass position $j$ for the $k$th time. The same fluence can be achieved with a unidirectional leaf trajectory, where $L_j = L_j^1 - L_j^2$ and $R_j = R_j^1 - R_j^2$. This gives the unidirectional leaf trajectory in figure A1(b). The fluence deposited with the unidirectional trajectories is $\sum_{j} \min(L_j - R_j)$, which is equal to the fluence deposited in the bidirectional case. This approach can be extended to nonunidirectional maps where the leaves move in more than two, say $n$, directions by using the transformation $L_j = L_j^1 - L_j^2 + ... + L_j^n$ if $n$ is even, and $L_j = L_j^1 - L_j^2 + ... + (T - L_j^n)$ if $n$ is odd, and $R_j$ is likewise defined. Note that the above transformations only hold when the leaves first move from left to right, which can be assumed without loss of generality.

The constraints on leaf positions and leaf speed remain valid. For leaf speed, this is intuitive: the time a leaf uses to move from one position to another will never be reduced, but it may be increased. Thus, if the multidirectional plan satisfies the leaf speed constraints, then the unidirectional plan satisfies those as well. In order to see that the left leaf remains on the left of the right leaf, we need to show that the right leaf traverses a position $j$ before the left leaf does, i.e. $L_j > R_j$. Recall that we assume the leaves first move from left to right, then from right to left, and so on, which implies $R_j < L_j^1$, $L_j^2 > R_j^2$, $R_j^3 < L_j^3$, etc. This implies that $L_j = L_j^1 - L_j^2 + ...$ is at least $R_j = R_j^1 - R_j^2 + ...$, and hence $L_j > R_j$.

When the dose rate is allowed to vary, unidirectional leaf trajectories may not be optimal as is shown in the following example. Keep in mind that the fluence at a bixel is visible in a leaf trajectory graph as the surface enclosed by the leaf trajectories and the boundaries of the bixel, see figure A2 where the fluence to the second bixel is equal to the surface of the grey area. Suppose that for a given maximum number of time steps $T = 8$, we aim to find
leaf trajectories that yield the fluence $[5 5.5 1 5.5 10 5.5 1 5.5 5]$ for the first leaf pair, with a maximum dose rate of $10 \text{ MU s}^{-1}$ and a maximum leaf speed of 1 bixel per second. Note that the leaf trajectory and dose rate schedule in figure A3 give the exact fluence map. This is the only treatment plan that can do so. In order to see this, note that the low fluence in bixels 3 and 7 can either be achieved by letting the leaves pass these bixels at maximum leaf speed with a small distance between them, or by lowering the dose rate. Moving the leaves past these bixels in close proximity to each other is not possible due to the time restriction, so turning the dose rate down is the only possibility.

Now suppose we have another leaf pair in the same map with fluence $[10 17 2 17 10 0 0 0 0]$. Given the dose rates enforced by the first leaf pair, there exist non-unidirectional leaf trajectories that can perfectly replicate this map (see figure A4(a) for an example). This fluence cannot be achieved with unidirectional leaf trajectories for the given dose rates, which we show by aiming to obtain such a leaf trajectory. Due to the symmetry, we deliver the dose to the third bixel halfway the time period by letting the leaves pass bixel 3 at maximum leaf speed and with a distance of 0.2, allowing for a delivery of 2 MU. Note that the fluence to this bixel cannot be delivered in one of the time periods with low dose rate, since this does not allow enough time to deliver the fluence in bixels 1 and 2 or 4 and 5. Given the leaf trajectory for bixel 3, we let the left leaf pass bixels 1 and 2 as late and as fast as possible, and let the right leaf start at the end of bixel 2 (see figure A4(b)). This gives the maximum possible dose

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figureA1.png}
\caption{An example of bidirectional (a) trajectories for the left and right leaf that can be transformed into unidirectional (b) trajectories without changing the fluence map.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figureA2.png}
\caption{The total fluence delivered to a bixel is equal to the area enclosed by the leaf trajectories.}
\end{figure}
to bixel 2, which is lower than the desired fluence. Thus, a perfect replication of the fluence map cannot be delivered with unidirectional leaf trajectories.

Appendix B. Gradient and Hessian

The gradient of the objective function $F = \sum_i (f_i - g_i)^2$, where $g_i(L, R, D)$ is substituted by its definition, is given by:
\[ \frac{\partial F}{\partial D_{ij}} = -2 \sum_{i,j} (f_{ij} - \sum_{s=1}^{T} e(L_{ij}^s, R_{ij}^s, j) \Delta D_{ij} e(L_{ij}^s, R_{ij}^s, j) \Delta ) \]  

(B.1)

\[ \frac{\partial F}{\partial L_{ij}^s} = -2 \sum_{i,j} (f_{ij} - \sum_{s=1}^{T} e(L_{ij}^s, R_{ij}^s, j) \Delta D_{ij} e(L_{ij}^s, R_{ij}^s, j) \Delta D_{ij} ) \]  

(B.2)

\[ \frac{\partial F}{\partial R_{ij}^s} = -2 \sum_{i,j} (f_{ij} - \sum_{s=1}^{T} e(L_{ij}^s, R_{ij}^s, j) \Delta D_{ij} e(L_{ij}^s, R_{ij}^s, j) \Delta D_{ij} ) \]  

(B.3)

In the above, the functions \( e_L(L_{ij}^s, R_{ij}^s, j) \) and \( e_R(L_{ij}^s, R_{ij}^s, j) \) are the derivatives of \( e(L_{ij}^s, R_{ij}^s, j) \) with respect to \( L \) and \( R \), respectively:

\[
e_L(L, R, j) = \begin{cases} 
0 & \text{if } L < j \text{ and } R > j, \text{ or } R \leq j \text{ or } L > j + 1 \\
-1 & \text{if } L < j + 1 \text{ and } R > j \\
[-1, 0] & \text{if } L = j \text{ or } L = j + 1
\end{cases}
\]  

(B.4)

\[
e_R(L, R, j) = \begin{cases} 
0 & \text{if } L < j + 1 \text{ and } R > j + 1, \text{ or } L \geq j + 1 \text{ or } R < j \\
1 & \text{if } L < j + 1 \text{ and } j < R < j + 1 \\
[1, 0] & \text{if } R = j \text{ or } R = j + 1
\end{cases}
\]  

(B.5)

From this, we can compute the Hessian using the following second order derivatives:

\[ \frac{\partial^2 F}{\partial D_{ij}^2} = 2 \Delta^2 \sum_{i,j} e(L_{ij}^s, R_{ij}^s, j) \]  

(B.6)

\[ \frac{\partial^2 F}{\partial (L_{ij}^s)^2} = 2 \Delta^2 \sum_{i,j} e_L(L_{ij}^s, R_{ij}^s, j)^2 D_{ij}^2 \]  

(B.7)

\[ \frac{\partial^2 F}{\partial (R_{ij}^s)^2} = 2 \Delta^2 \sum_{i,j} e_R(L_{ij}^s, R_{ij}^s, j)^2 D_{ij}^2 \]  

(B.8)

\[ \frac{\partial^2 F}{\partial D_{ij} \partial L_{ij}^s} = 2 \Delta \sum_{i,j} e_L(L_{ij}^s, R_{ij}^s, j) (g_{ij} - f_{ij} + e(L_{ij}^s, R_{ij}^s, j) D_{ij}) \]  

(B.9)

\[ \frac{\partial^2 F}{\partial D_{ij} \partial R_{ij}^s} = 2 \Delta \sum_{i,j} e_R(L_{ij}^s, R_{ij}^s, j) (g_{ij} - f_{ij} + e(L_{ij}^s, R_{ij}^s, j) D_{ij}) \]  

(B.10)

\[ \frac{\partial^2 F}{\partial L_{ij}^s \partial R_{ij}^s} = 2 \Delta^2 \sum_{i,j} e_L(L_{ij}^s, R_{ij}^s, j) e_R(L_{ij}^s, R_{ij}^s, j) D_{ij}^2. \]  

(B.11)

The solution time of the interior-point optimization can be strongly reduced by giving the gradient to the solver. However, in our case supplying the Hessian does not yield improvements. We observed that reducing the time the solver spends on computing the Hessian yields an overall time speed-up, indicating further that the Hessian is not helpful for this optimization problem. We include the Hessian calculation for reference and in the event that is useful for other optimization approaches.
References

Alber M and Nüsslin F 2000 Intensity modulated photon beams subject to a minimal surface smoothing constraint Phys. Med. Biol. 45 N49–52
Bokrantz R 2012 Multicriteria optimization for volumetric-modulated arc therapy by decomposition into a fluence-based relaxation and a segment weight-based restriction Med. Phys. 39 6712–25
Bortfeld T R 2006 IMRT: a review and preview Phys. Med. Biol. 51 R363
Bortfeld T R, Kahler D L, Waldron T J and Boyer A L 1994 X-ray field compensation with multileaf collimators Int. J. Radiat. Oncol. Biol. Phys. 28 723–30
Chen D, Luan S and Wang C 2011 Coupled path planning, region optimization, and applications in intensity-modulated radiation therapy Algorithmica 60 152–74
Chen W, Unkelbach J, Trofimov A, Madden T, Kooy H, Bortfeld T R and Craft D L 2012 Including robustness in multi-criteria optimization for intensity modulated proton therapy Phys. Med. Biol. 57 591–608
Craft D, Bangert M, Long T, Papp D and Unkelbach J 2014 The CORT dataset GigaSci Database 3 1
Craft D, McQuaid D, Wala J, Chen W, Salari E and Bortfeld T 2012 Multicriteria VMAT optimization Med. Phys. 39 686
Craft D, Süß P and Bortfeld T 2007 The tradeoff between treatment plan quality and required number of monitor units in IMRT Int. J. Radiat. Oncol. Biol. Phys. 67 1596–605
Crooks S M, McAvan L F, Robinson D F and Xing L 2002 Minimizing delivery time and monitor units in static imrt by leaf-sequencing Phys. Med. Biol. 47 3105–16
Ehrgott M, Gülçer Ç, Hamacher H W and Shao L 2010 Mathematical optimization in intensity modulated radiation therapy Ann. Oper. Res. 175 309–65
Engel K 2005 A new algorithm for optimal multileaf collimator field segmentation Discrete Appl. Math. 152 35–51
Gaddy M R and Papp D 2016 Technical note: improving the VMERGE treatment planning algorithm for rotational therapy Med. Phys. 43 4093–7
Jelen U, Sohn M and Alber M 2005 A finite size pencil beam for IMRT dose optimization Phys. Med. Biol. 50 1747–66
Luan S, Wang C, Cao D, Chen D, Shepard D and Yu C 2008 Leaf-sequencing for intensity-modulated arc therapy using graph algorithms Med. Phys. 35 61–9
Matuszak M M, Larsen E W, Jee K W K and Fraass D L D L M A 2008 Adaptive diffusion smoothing: a diffusion-based method to reduce IMRT field complexity Med. Phys. 35 1532–46
Palma D, Vollans E, James K, Nakano S, Moiseenko V, Schaffer R, McKenzie M, Morris J and Otto K 2008 Volumetric modulated arc therapy for delivery of prostate radiotherapy: comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy Int. J. Radiat. Oncol. Biol. Phys. 72 996–1001
Papp D and Unkelbach J 2014 Direct leaf trajectory optimization for volumetric modulated arc therapy planning with sliding window delivery Med. Phys. 41 011701
Shepard D, Cao D, Afghan M and Earl M 2007 An arc-sequencing algorithm for intensity modulated arc therapy Med. Phys. 34 464–70
Stein J, Bortfeld T R, Dörschel B and Schlegel W 1994 Dynamic x-ray compensation for conformal radiotherapy by means of multi-leaf collimation Radiother. Oncol. 32 163–73
Teoh M, Clark C, Wood K, Whitaker S and Nisbet A 2011 Volumetric modulated arc therapy: a review of current literature and clinical use in practice Br. J. Radiol. 84 967–96
Unkelbach J et al 2015 Optimization approaches to volumetric modulated arc therapy planning Med. Phys. 42 1367–77
Webb S 1989 Optimisation of conformal radiotherapy dose distributions by simulated annealing Phys. Med. Biol. 34 1349–70
Xia P and Verhey L 1998 Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments Med. Phys. 25 1424–34
Yu C X 1995 Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy Phys. Med. Biol. 40 1435–49