Optimization approaches to volumetric modulated arc therapy planning

Jan Unkelbach,[a)] Thomas Bortfeld, and David Craft
Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114

Markus Alber
Department of Medical Physics and Department of Radiation Oncology, Aarhus University Hospital, Aarhus C DK-8000, Denmark

Mark Bangert
Department of Medical Physics in Radiation Oncology, German Cancer Research Center, Heidelberg D-69120, Germany

Rasmus Bokrantz
RaySearch Laboratories, Stockholm SE-111 34, Sweden

Danny Chen
Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Ruijiang Li and Lei Xing
Department of Radiation Oncology, Stanford University, Stanford, California 94305

Chunhua Men
Department of Research, Elekta, Maryland Heights, Missouri 63043

Simeon Nill
Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London SM2 5NG, United Kingdom

Dávid Papp
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695

Edwin Romeijn
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Ehsan Salari
Department of Industrial and Manufacturing Engineering, Wichita State University, Wichita, Kansas 67260

(Received 7 November 2014; revised 28 January 2015; accepted for publication 4 February 2015; published 25 February 2015)

Volumetric modulated arc therapy (VMAT) has found widespread clinical application in recent years. A large number of treatment planning studies have evaluated the potential for VMAT for different disease sites based on the currently available commercial implementations of VMAT planning. In contrast, literature on the underlying mathematical optimization methods used in treatment planning is scarce. VMAT planning represents a challenging large scale optimization problem. In contrast to fluence map optimization in intensity-modulated radiotherapy planning for static beams, VMAT planning represents a nonconvex optimization problem. In this paper, the authors review the state-of-the-art in VMAT planning from an algorithmic perspective. Different approaches to VMAT optimization, including arc sequencing methods, extensions of direct aperture optimization, and direct optimization of leaf trajectories are reviewed. Their advantages and limitations are outlined and recommendations for improvements are discussed. [http://dx.doi.org/10.1118/1.4908224]

Key words: radiotherapy, treatment plan optimization, VMAT

1. INTRODUCTION

Arc therapy refers to radiotherapy treatments in which the gantry continuously rotates around the patient while the treatment beam is on and delivers dose to the patient. Using conventional Linacs, conformal arc therapy has long been a delivery mode for a variety of treatment sites. In this case, the treatment field conforms to the projection of the target volume at every gantry angle. For that reason, conformal arcs were typically used for small lesions of roughly spherical shape, which do not require intensity modulation. To extend arc therapy to more complex treatment sites that require intensity-modulated radiotherapy (IMRT), specialized hardware was developed. The design of Tomotherapy machines resembles a helical computed tomography scanner in which the diagnostic x-ray tube is replaced by a Linac. The Linac continuously

 0094-2405/2015/42(3)/1367/11/$30.00

rotates while the patient is transported through the device on a treatment couch. The tumor is irradiated slice-by-slice using a fan beam that is modulated using a binary multileaf collimator (MLC). For a review of the history of Tomotherapy, we refer to the paper by Mackie[1] and the original publication from 1993.[2] In 2002, the first patient was treated with Tomotherapy.

Volumetric modulated arc therapy (VMAT) refers to techniques that deliver intensity-modulated fields through rotational delivery using conventional Linacs equipped with MLCs. In contrast to conformal arcs, the treatment field does not necessarily conform to the target volume at every gantry angle. Instead, an effectively intensity-modulated field is delivered over an arc sector. This rotational form of IMRT delivery has been introduced under different names including intensity-modulated arc therapy (IMAT),[3] VMAT,[4] arc-modulated radiotherapy (AMRT),[5] arc-modulated cone beam therapy (AMCBT),[6] aperture modulated arc therapy (AMAT),[7] and sweeping-window arc therapy (SWAT).[8] For this paper, we adopt the term that became most widely used, VMAT.

The possibility of delivering intensity-modulated arcs on conventional Linacs has been delayed in part due to the absence of treatment planning systems (TPSs) that are able to provide VMAT treatment plans. This changed in 2008 with the marketing of RapidArc (Varian) and SmartArc (Philips). Since then VMAT has experienced a rapid and widespread clinical application (see, e.g., Ref. 9 for a review). Many treatment planning studies have characterized treatment plan quality and delivery times of VMAT for various treatment sites based on the commercial VMAT implementations available at the time.

In contrast, algorithm development for VMAT planning has been scarce.[10] This is despite the fact that VMAT planning represents a more challenging optimization problem than IMRT planning. For most objective functions commonly used, the fluence map optimization (FMO) problem in IMRT represents a large continuous, but convex optimization problem.[11] Thus, established optimization algorithms find the globally optimal solution. This is not the case for VMAT planning as will become apparent in Sec. 2. In November 2013, the authors of this report met at Massachusetts General Hospital for a workshop to review the current state of VMAT optimization and suggest possible improvements. As a result, this paper provides a summary of VMAT planning from an algorithmic perspective. For a review of the clinical implementation issues, we refer the reader to Ref. 9.

Most VMAT planning algorithms utilize methods that were previously developed for IMRT planning, and customize these methods to the VMAT setting. This includes primarily three components:

1. In FMO, the fluence profiles at discrete incident beam angles are optimized.
2. Arc sequencing, where fluence maps are converted to apertures. This step is analogous to the two-step approach to fixed field IMRT planning, however, in VMAT, additional constraints on the shape of the apertures are enforced in order to allow for efficient delivery. For example, apertures of neighboring gantry angles are required to be similar.
3. Direct aperture optimization (DAO). In a DAO approach, a VMAT plan is characterized by a collection of apertures, typically one aperture per 2° arc sector. DAO methods are then used to optimize the shape and intensity of each aperture.

VMAT algorithms differ in the component that they rely on the most, as well as the exact implementation of each step. One type of VMAT algorithms emphasizes the arc sequencing step and attempts to faithfully recreate fluence maps, with the goal of obtaining a final or near final treatment plan (Sec. 3.A). Most of the commercial implementations heavily rely on DAO methods. This includes RapidArc (Varian), which uses a global stochastic optimization approach to DAO, without utilizing FMO and sequencing methods (Sec. 3.C). SmartArc (Philips) as well as MONACO (Elekta) utilize a local gradient based optimization approach to DAO, and adopt FMO and arc sequencing methods to obtain a starting point for DAO (Sec. 3.B).

The remainder of this paper is organized as follows: In Sec. 2, we first provide a general formulation of the VMAT planning problem, followed by a specialized formulation that reflects the DICOM specification of a VMAT plan. In Sec. 3, the published approaches to VMAT optimization are reviewed. Section 4 discusses advantages and limitations of the approaches.

2. PROBLEM FORMULATION

VMAT optimization methods adopt many concepts that have been developed for IMRT optimization. The patient is discretized into voxels which we index by i. The incident fluence is discretized into a 2D beamlet grid which is indexed by the MLC leaf pair index n and the index j for the beamlet position in leaf travel direction. For the main part of the paper, we consider coplanar VMAT delivery, i.e., the couch angle is fixed. A 360° arc is then represented by discrete gantry angles φ. Typically, 180 angles at 2° resolution are used. At gantry angle φ, an IMRT or VMAT plan delivers an effective fluence x^{φ}_{jn}. The dose contribution from an individual beam direction φ can be calculated based on a dose-influence matrix D^{φ}_{ijn}, which describes the dose contribution of a beamlet j in MLC row n to voxel i for unit intensity. If the fluence is measured in monitor units (MU), the natural unit for the dose-influence matrix is Gy per MU. The total dose can then be calculated as

$$d_i = \sum_{\varphi} \sum_{n} \sum_{j} D^{\varphi}_{ijn} x^{\varphi}_{jn}. \tag{1}$$

Treatment planning is based on an objective function $f(d)$, which is a function of the dose distribution d and measures the quality of the treatment plan. For simplicity of notation, we do not consider additional constraints on the dose distribution.

2.A. Ideal benchmark plan

This formulation gives rise to an ideal benchmark solution that a VMAT plan can be compared against. We consider the fluence map optimization problem

$$\underset{x}{\text{minimize}}\ f(d) \tag{2}$$

$$\text{subject to } d_i = \sum_{\varphi}\sum_{n}\sum_{j} D^{\varphi}_{ijn} x^{\varphi}_{jn} \tag{3}$$

$$x^{\varphi}_{jn} \geq 0. \tag{4}$$

Here, we allow arbitrary effective fluence maps at each gantry angle, with the only restriction of non-negative beamlet intensities. Any VMAT plan can at best be as good as the full IMRT benchmark solution. Delivering each fluence map exactly would be very time consuming. Therefore, VMAT planning will seek for treatment parameters that correspond to a plan that is deliverable efficiently, but is close to the benchmark plan in terms of plan quality. This is expected to be possible because a diminishing return is observed when using a large number of coplanar beams in IMRT. In practice, an ideal benchmark plan using 180 beams may only be insignificantly better than say 20 beams. This has been observed empirically, and theoretical support is provided in Refs. 12 and 13. Hence, one can imagine a 360° arc divided into 20 arc sectors of 18° each. Even though VMAT delivers an open field at a single gantry angle,[14] the total fluence delivered over a 18° arc sector can be thought of as an intensity-modulated field. Thus, a well designed VMAT plan has the potential to approximate a very high quality 20-beam IMRT plan that is beyond the current IMRT practice using 7–9 coplanar beams.[15]

2.B. Formulating the VMAT planning problem

When delivering a VMAT plan, the delivered fluence x^{φ}_{jn} is determined through three types of variables:

1. The MLC leaf trajectories, i.e., the positions of the left leaves $L_n(t)$ and the right leaves $R_n(t)$ as a function of time.
2. The gantry angle $\varphi(t)$ as a function of time.
3. The dose rate $\delta(t)$ as a function of time.

Here, we assume that the collimator and couch are at fixed angles, and that the jaws can be positioned in a postprocessing step with minor impact on the treatment plan. Based on these trajectories, the effective fluence x^{φ}_{jn} can be calculated, which is illustrated in Fig. 1 for a single leaf pair. Let us for simplicity assume that the dose rate is constant over the arc sector φ. Then, the effective fluence x^{φ}_{jn} is determined by the time that beamlet j is exposed by the MLC leaves. In Fig. 1, this is given by the area enclosed by the leaf trajectories and the beamlet boundaries. This method to relate leaf positions to fluence involves the following, commonly employed approximation: If at time t, a beamlet is partially exposed by the MLC leaves, the time point's contribution to the beamlet's effective fluence is proportional to the exposed fraction of the beamlet (see further discussion in Sec. 2.C and Fig. 3).

VMAT planning aims at determining the optimal trajectories for MLC leaves, dose rate and gantry angle. Thereby, optimality involves

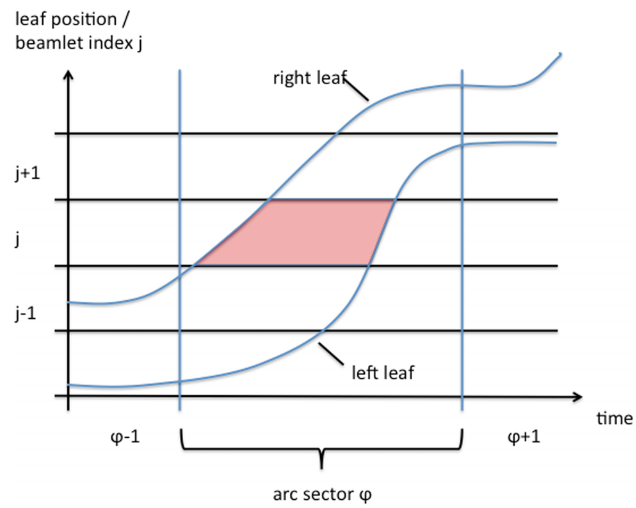

FIG. 1. Illustration of the effective fluence created by a leaf trajectory. For simplicity, we assume constant dose rate such that the fluence of a beamlet j corresponds to the time that the beamlet is exposed by the MLC leaf pair (red area).

1. plan quality measured through the dose based objective function $f(d)$;
2. the delivery time T.

In practice, other goals in addition to plan quality and delivery time may be considered. For example, if two plans are equal in terms of delivery time and plan quality, the plan with a smaller total number of MU is preferred. A small number of MU may further be a surrogate for larger field openings and more accurate dose calculation. The trajectories have to fulfill constraints imposed by machine limitations. The main limitations are as follows.

1. A maximum MLC leaf speed. Typical values are in the range of 3–6 cm/s.
2. Gantry speed constraints: All machines have maximum gantry speed, which is typically 6°/s (1 min for a full rotation). In addition, there are constraints on the acceleration and deceleration of the gantry.
3. Dose rate constraints: All machines have a maximum dose rate. Additional limitations on the dose rate are highly machine dependent. While some Linacs allow continuously varying dose rates, others may allow only discrete values. A typical value for the maximum dose rate is 600 MU/min. However, some machines may have lower values, whereas the use of flattening filter free beams may substantially increase the dose rate.
4. Depending on the MLC model, additional constraints on the leaf motion exist, e.g., a minimum leaf gap for moving leaves or interdigitation constraints for older models.
5. Depending on the treatment machine, additional restrictions on leaf motion exist, e.g., a maximum leaf travel constraint per degree of gantry rotation.

Different approaches to VMAT planning may use different parameterizations of the trajectories in order to formulate the VMAT optimization problem. The most common representation is driven by the DICOM specification of a treatment plan as described below in Sec. 2.C.

It is intuitive that the ideal plan quality of the benchmark plan can be reproduced if the treatment time is allowed to be large, and if dose rate and gantry speed may vary continuously between zero and the upper bound. For short delivery times, sacrifices in plan quality need to be accepted. This is schematically illustrated in Fig. 2. The lower dotted line indicates the optimal objective function value of the benchmark plan. If the objective function $f(d)$ is convex, this value can be computed with high accuracy. Ideally, a VMAT planning algorithm would determine a treatment plan that is Pareto optimal on the trade-off curve between delivery time and plan quality, i.e., guarantee optimal plan quality for a given delivery time. Current VMAT algorithms cannot provide such guarantees so that the true trade-off curve remains unknown.

2.C. DICOM specification of a treatment plan

Today, VMAT planning has to cope with a disconnect of the TPS and the treatment machine control system. As a consequence, the TPS does not specify the trajectory explicitly. Instead, the treatment plan is specified in DICOM format as a sequence of control points (CP). Each CP is defined through a gantry angle, leaf positions, and the cumulative number of MU that is delivered up to this control point. This gives rise to a formulation of VMAT planning as a DAO problem, which is the basis for most current commercial implementations of VMAT planning.

Let us for simplicity of notation assume that one aperture is assigned to each gantry angle. In a DAO formulation of VMAT planning, we aim to determine the leaf positions L_n^φ, R_n^φ, and the aperture weight y^φ at each control point. In this case, the

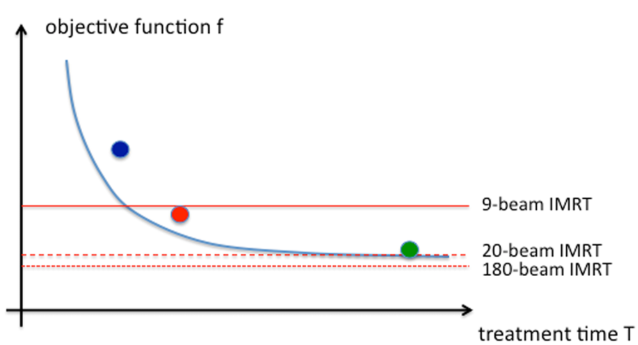

FIG. 2. Schematic illustration of the trade-off between plan quality and delivery time T (blue curve). Plan quality is measured in terms of the objective function value f, i.e., better plan quality corresponds to lower objective function values. A lower bound on the objective function is given by the FMO solution which allows for arbitrary intensity modulation at every angle (dotted line). A FMO plan using 20 beams (dashed line) is typically only slightly inferior to a 180 beam FMO solution, but may noticeably improve on a nine beam FMO plan (solid line). Current VMAT plans are generally not Pareto optimal. For example, sliding window (SW) VMAT plans can achieve high plan quality, but are suboptimal in terms of treatment time (right dot). Typical VMAT plans may exceed the quality of nine beam IMRT plans, but bear potential for improvement in quality or efficiency (middle dot). For complicated geometries or short treatment times, current VMAT plans may be inferior to IMRT plans (left dot).

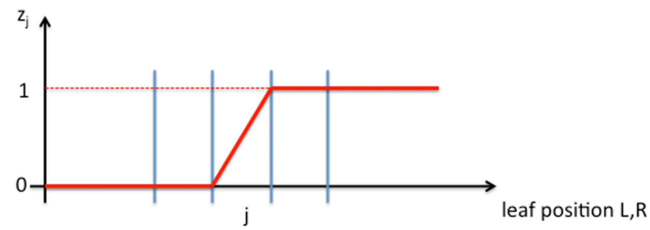

FIG. 3. Piecewise linear approximation of the intensity of beamlet j as a function of leaf position. The function z_j is zero if the leaf is to the left of beamlet j, one if the leaf is to the right of beamlet j, and a linear function while the leaf edge is positioned within the beamlet. The intensity of beamlet j is then given by $x_{jn}^\varphi = y^\varphi(z_j(R_n^\varphi) - z_j(L_n^\varphi))$.

effective fluence x_{jn}^φ is restricted and can be approximated as a piecewise linear function: The fluence is given by the aperture weight for those beamlets that are fully exposed by the leaf pair, and zero for those beamlets that are covered completely. When a leaf edge is located in between the boundaries of a beamlet, the effective fluence of the beamlet is often approximated linearly. This is illustrated in Fig. 3.

In this view, VMAT planning aims at optimizing the intensities and shapes of all apertures. Thus, VMAT planning is essentially reduced to a DAO problem of the form

$$\underset{y,L,R}{\text{minimize}}\ f(d) \tag{5}$$

$$\text{subject to } d_i = \sum_\varphi \sum_n \sum_j D_{ijn}^\varphi x_{jn}^\varphi \tag{6}$$

$$x_{jn}^\varphi = y^\varphi\left[z_j(R_n^\varphi) - z_j(L_n^\varphi)\right] \tag{7}$$

$$y^\varphi \geq 0 \tag{8}$$

$$L_n^\varphi \leq R_n^\varphi \tag{9}$$

where z is the function defined in Fig. 3 and we neglect additional MLC constraints such as interdigitation. This formulation of the DAO problem is nonconvex due to the shape of the function z and cannot be solved to optimality in practice. The representation of leaf trajectories and the calculation of the effective fluence are illustrated in Fig. 4. Different solutions approaches to the DAO problem have been pursued, including local gradient based optimization,[16–18] stochastic search,[19,20] and column generation inspired methods.[21,22]

An approximate solution to the DAO problem yields a sequence of control points, which, in principle, represents a VMAT plan. Depending on machine capabilities it may also be deliverable, in particular if gantry speed and dose rate may vary continuously. The trajectories for MLC leaves, dose rate, and gantry speed are determined by the machine controller. In particular, the aperture weights y^φ are converted to gantry speeds and dose rates according to

$$y^\varphi = \frac{\delta^\varphi}{s^\varphi}\Delta\varphi, \tag{10}$$

where $\Delta\varphi$ is the angular distance between two control points in degrees, s^φ is the angular gantry speed in degrees per second, and δ^φ is the dose rate in MU per second. This shows that gantry speed and dose rate are not uniquely determined by the aperture weight. A large aperture weight can be achieved

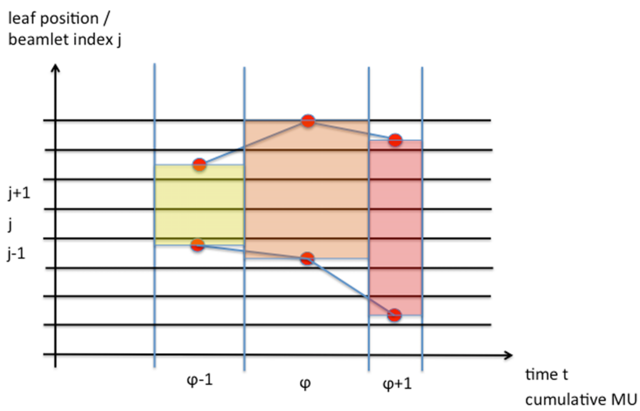

Fig. 4. Schematic illustration of the parameterization of leaf trajectories in a DAO approach to VMAT planning. The leaf positions are specified at a discrete control point for each arc sector (red dots). The beamlet fluence is approximated based on the leaf position at the control point. The leaves are expected to move linearly in between two control points.

by a large dose rate and/or a slow gantry speed. To minimize delivery time, the combination with the largest possible gantry speed should be taken.[23]

The above DAO formulation does not account for efficient delivery yet. If the leaf positions change substantially between two adjacent control points, the gantry speed has to be low, which in turn increases the treatment time. To improve efficiency, aperture shapes between adjacent control points should be similar. DAO algorithms adapted to VMAT planning may approach this by requesting that the plan is deliverable given a minimum gantry speed. This yields restrictions on maximum leaf travel between control points, which represent linear constraints for leaf positions of adjacent control points,

$$|L_n^{\varphi} - L_n^{\varphi+1}| \leq \frac{v_{\max}}{s_{\min}} \Delta\varphi \tag{11}$$

and similarly for the right leaves. Here, $v_{\max}$ denotes the maximum leaf speed, and $s_{\min}$ is the desired minimum gantry speed.

3. OVERVIEW OF VMAT PLANNING APPROACHES

In this section, we summarize VMAT algorithms suggested in the literature, as well as commercial implementations as far as the algorithm is disclosed. We start by reviewing arc sequencing methods in Sec. 3.A. Subsequently, the adaptation of DAO methods in VMAT planning is discussed. This includes the most widely used commercial implementations in ECLIPSE (Varian),[4] PINNACLE (Philips),[24] RayStation (RaySearch), and MONACO (Elekta). Finally, an approach to directly optimize leaf trajectories is discussed in Sec. 3.E.

3.A. Arc sequencing methods

The overall success and widespread use of FMO plus sequencing in IMRT planning suggests to develop a similar two-step approach for VMAT, i.e., FMO plus arc sequencing. As in IMRT, the goal of the sequencer is to convert a fluence map into a sequence of aperture shapes and intensities. The main difference to step&shoot IMRT is that the concatenation of apertures has to form a single or multiple deliverable arcs. Therefore, neighboring apertures should be similar to allow for efficient delivery and promote dose calculation accuracy. In this section, we consider arc sequencing methods that attempt to reproduce the fluence maps faithfully in order to obtain a final (or near final) treatment plan. Multiple approaches have been published, the majority of which share the idea of a sliding window (SW) type conversion. One of the advantages of SW sequencing is that the apertures are naturally ordered by leaf position.[25]

3.A.1. *Delivery through multiple arcs*

An early arc sequencing method has been proposed in Ref. 26. In this method, FMO is performed at (typically) 10° resolution. Each fluence map is sequenced into k apertures in a sliding window fashion. The apertures are subsequently connected to form k arcs. The sliding window type sequencing has the advantage that the apertures can be sorted according to leaf position, which facilitates an efficient formation of arcs with limited leaf travel. The main disadvantage is the need for multiple arcs, especially for complicated geometries, which leads to long treatment times.

3.A.2. *Arc sequencing using graph algorithms*

To overcome the need for multiple arcs, FMO can be performed at a coarser angular resolution. Each fluence map is then sequenced into k apertures, which are distributed over the corresponding arc sector instead of forming multiple arcs. This makes the arc sequencing problem more challenging compared to the setting in Ref. 26. To that end, Ref. 5 suggests a method using shortest path algorithms adopted from graph theory. Algorithmic details of the approach are described in Ref. 27. The arc sequencing problem is represented as a directed acyclic graph consisting of layers of vertices as illustrated in Fig. 5. Each layer represents an arc sector, and each vertex represents a possible leaf configuration at the beginning of the arc sector. An edge corresponds to a leaf trajectory that starts at a particular leaf configuration at the beginning of arc sector η_i and ends at a particular leaf configuration at

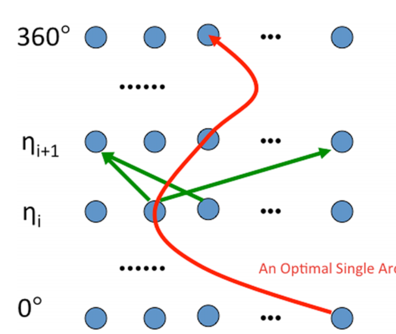

Fig. 5. Representation of the arc sequencing problem as a directed acyclic graph. The vertices represent possible leaf configurations at the beginning and end of an arc sector η. The cost associated with an edge (green arrow) corresponds to the minimum error in the delivered fluence. The optimal leaf trajectory corresponds to the shortest path through the graph (red line).

the beginning of arc sector η_{i+1}. The cost associated with an edge is given by the minimum error between the arc sector's ideal fluence map and the effective fluence realized by the best deliverable leaf trajectory. Given this as input, the arc sequencing problem corresponds to a shortest path problem, i.e., finding a path through the graph that yields the smallest cumulative error in all fluence maps.

In the classical sliding window delivery, the leaves are positioned at the edge of the field at the beginning/end of an arc sector. If the time allowed for leaf travel is large, the graph includes this solution (which yields an exact reproduction of the fluence map with no error). If the treatment time is constrained, the algorithm will determine an optimal way to approximate the fluence map by either forcing the leaves to move across the field faster, or by moving the terminal leaf positions to inside the field.

Determining the cost associated with an edge represents a hard problem if arbitrary leaf motion is allowed. To that end, Ref. 27 makes the restriction that only unidirectional leaf motion within the arc sector is considered.

3.A.3. Merging of fluence maps

Unlike the arc sequencing methods above that perform FMO at sparse angular resolution, the algorithm described by Craft *et al.*[28] first determines an ideal benchmark treatment plan consisting of 180 IMRT fields at 2° angular resolution. In principle, each of the 180 IMRT fields can be delivered using a sliding window conversion, which recreates the ideal benchmark solution. However, this would be very time consuming. In order to reduce the treatment time, adjacent fluence maps are combined by adding their fluence values beamlet by beamlet. The resulting combined fluence map is delivered over the enlarged arc sector that corresponds to both fluence maps. In the original publication,[28] merging of two adjacent fluence maps is based on a similarity measure, i.e., in each step the most similar neighboring fluence maps are combined. The merging strategy has been improved by Salari *et al.*[29] In this work, optimal merging patterns are determined via solving a bicriteria quadratic integer programming problem that aims at minimizing the treatment time as well as the dose deviation from the benchmark solution. However, this work also showed that the simplified strategy of successively merging neighboring fluence maps was similarly effective.

3.B. Local DAO with FMO-informed segment initialization

The DAO problem in Sec. 2.C is a linearly constrained nonlinear program[30] for which standard solvers exists, see, e.g., Gill *et al.*[31] A solver can, however, guarantee at most local optimality because of the formulation's nonconvexity, meaning that the quality of the optimized plan depends on the quality of the initial apertures. One possibility to obtain good initial apertures is to sequence optimized fluence maps, which results in the following three-step scheme:

1. *FMO:* The fluence of static beams spaced η degrees apart is optimized.
2. *Arc sequencing:* Each fluence map is sequenced into apertures that are distributed equidistantly over the associated η-degree arc sector.
3. *DAO:* The apertures are refined by gradient based optimization where the leaf positions and aperture intensities are treated as variables.

The algorithm above has been suggested by Bzdusek *et al.*[24] and has been commercialized as SmartArc (Philips), Oncentra VMAT (Nucletron), and RayArc (Raysearch Laboratories). In addition, the VMAT implementation in MONACO (Elekta), although developed independently, is built around the same general three steps. Likewise, Bedford[32] as well as Wild *et al.*[33] suggested algorithms that use a similar three-step framework. VMAT planning approaches in this category do not *per se* aim to reproduce the fluence profiles of the FMO step faithfully by sequencing. Instead, these methods rely primarily on the DAO step to reproduce the dose quality of an FMO plan. However, the sequencing in step 2 has to provide a good enough starting position for the DAO step to succeed.

In step 3, a local optimum to the DAO problem (5)–(9) is obtained. This can, for example, be performed using gradient based methods such as quasi-Newton methods. This requires evaluation of the gradient of the objective function with respect to the machine parameters, which can be obtained by invocation of the chain rule. For example, using Eqs. (6) and (7), differentiation with respect to the position R_n^φ of the nth right leaf at angle φ yields

$$\begin{aligned}\frac{\partial f}{\partial R_n^\varphi} &= \sum_i \frac{\partial f}{\partial d_i}\frac{\partial d_i}{\partial R_n^\varphi}\\ &= \sum_i \frac{\partial f}{\partial d_i}\sum_j \frac{\partial d_i}{\partial x_{jn}^\varphi}\frac{\partial x_{jn}^\varphi}{\partial R_n^\varphi}\\ &= \sum_i \frac{\partial f}{\partial d_i} y^\varphi \sum_j D_{ijn}^\varphi \frac{\partial Z_j}{\partial R_n^\varphi},\end{aligned} \tag{12}$$

see also Hårdemark *et al.*[16] Note that, using the approximation in Fig. 3, only one term in the sum over the beamlet index j will be nonzero, corresponding to the beamlet that the leaf end is currently positioned in the expression for the left leaves and the aperture weights is analogous.[34] The early DAO work by de Gersem *et al.*[18] does not explicitly calculate gradients according to Eq. (12), but can be considered a local leaf position refinement method using a finite difference approximation of the gradient. An alternative algorithm for solving the DAO problem (5)–(9), which uses a trust region like method, has been published in Ref. 17.

Various realizations of this three-step approach differ in their exact implementation of the individual steps.

- *RayArc and SmartArc:* FMO in step 1 is typically performed for 15 equispaced beam angles. This choice of angular distance η is influenced by the aperture adjustments in step 2. Bzdusek *et al.* observed that more narrowly spaced beams than $\eta = 24°$ lead to large leaf adjustments and therefore have an overall detrimental effect, even though the angular displacements become smaller when the apertures are repositioned. Regularization to obtain smooth fluence maps is achieved by early stopping of the FMO optimizer. During sequencing, the apertures are adjusted to comply with motion constraints such as leaf speed limitations. Linearly interpolated apertures are finally inserted up to the desired angular resolution. Finally, sequential quadratic programming is used to find a local optimum of the DAO problem (5)–(9).

- *Monaco:* FMO is performed for a small number of directions between 7 and 36. The resulting fluence maps are initially sequenced in a sliding window fashion, and are manipulated to yield a more time efficient initial VMAT plan. In addition, if more than one gantry rotation is desired, the fluence profiles are split and distributed into the rotations so as to minimize delivery time. The latter is further minimized by adjusting the gantry angles of control points according to leaf travel and dose rate constraints. All of these measures aim to overcome the inherent delivery time disadvantage of sliding window sequencing. During the DAO stage, the doses of each dynamic arc segment are computed without approximation with a Monte Carlo dose algorithm, in order to avoid dose discrepancies by discretization of leaf and gantry movements into static positions. Hence, at any stage of the DAO, the dose distribution is final and deliverable. The means to achieve equivalence with the FMO dose is by defining constraints for all treatment goals of FMO. The DAO step uses gradient information to select apertures and drive changes of leaf and jaw positions.

3.C. Global DAO with geometry-based segment initialization

The methods in Sec. 3.B use gradient based optimization of aperture shapes, and therefore require a good starting point for obtaining a local optimum of high quality. Other works suggest to use stochastic search methods instead to optimize leaf positions. The approach by Otto[4] uses simulated annealing to optimize leaf positions, similar to the DAO approach introduced by Shepard *et al.*[19] and others.[18] Other authors suggest to use Tabu search[6] to circumvent the problem of local minima in gradient based aperture shape optimization. Both Refs. 4 and 6 use a geometry-based initialization of apertures in which the initial aperture shape corresponds to the beam's eye view projection of the target volume [possibly excluding the projection of organs at risk (OARs)].

In Ref. 4, DAO using simulated annealing is combined with an approach for iteratively adding apertures to the treatment plan. DAO starts with one aperture at a small number of equispaced gantry angles. During the optimization additional apertures are added for intermediate angles until the desired final angular resolution is reached. The approach in Ref. 4 has been commercialized as RapidArc (Varian). Recent implementations of VMAT planning in ECLIPSE may deviate from the original algorithm,[35] however, the use of simulated annealing remains the underlying method to optimize leaf positions.

3.D. Successive generation of apertures

Another approach to DAO uses an iterative generation of apertures. The approach consists of two steps that are repeated until a stopping criterion is fulfilled.

1. A new aperture is identified which promises a large improvement to the current plan quality.
2. The intensities of all apertures are optimized jointly.

One possibility to identify a promising aperture is based on the gradient of the objective function with respect to a candidate aperture, evaluated at the current treatment plan where the existing apertures have optimal intensity. To identify a promising aperture, the partial derivatives $\partial f/\partial x_{jn}^{\varphi}$ are considered. If $\partial f/\partial x_{jn}^{\varphi}$ is negative, adding this beamlet to the treatment plan with positive intensity lowers the objective function value, i.e., improves plan quality. To identify a promising aperture A_φ at angle φ, one can therefore seek for an aperture that minimizes

$$\sum_{(j,n)\in A_\varphi} \frac{\partial f}{\partial x_{jn}^{\varphi}}, \tag{13}$$

where the sum is taken over the beamlets contained in the aperture. This problem can be solved efficiently as described in the original publication[21] for every gantry angle. Subsequently, the aperture with the best score is added. The method was originally proposed for step&shoot IMRT and referred to as column generation approach to DAO due to its resemblance to the large scale linear programming technique of the same name.

The application of this approach to VMAT planning has been described in Refs. 36 and 37. In step&shoot IMRT, several apertures per beam direction are typically generated, while there is no restriction on the similarity of different apertures. For the VMAT application, Ref. 37 suggests the following two modifications of the original method in Ref. 21:

1. Only one aperture per gantry angle is generated. Once an aperture is determined, the corresponding beam angle is removed from the set of candidate apertures.
2. Neighboring apertures should be similar in the sense of maximum leaf travel. This can be accounted for in step one. When a new aperture is generated, the pool of candidate apertures is restricted and depends on the angular distance to apertures that are already fixed.

3.E. Direct leaf trajectory optimization for sliding window delivery

The work in Ref. 38 considers a sliding window delivery of VMAT and directly optimizes leaf trajectories based on the dosimetric objective function f. In the DAO formulation, a VMAT plan is described by specifying leaf positions at particular time points. In case of a sliding window delivery of VMAT, the leaf trajectories can instead be defined by specifying the times at which a leaf edge arrives and departs from a given beamlet. This is illustrated in Fig. 6.

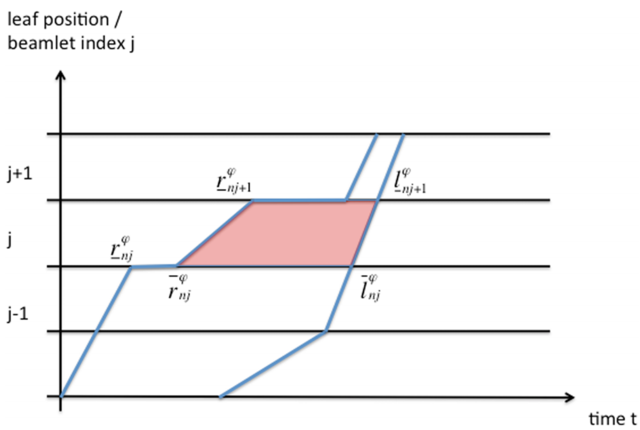

FIG. 6. Piecewise linear approximation of the leaf trajectories, specified via arrival and departure times. $\overline{l}^{\varphi}_{nj}$ ($\overline{r}^{\varphi}_{nj}$) denotes the time when the left (right) leaf departs from the boundary of beamlet j, and $\underline{l}^{\varphi}_{nj+1}$ ($\underline{r}^{\varphi}_{nj+1}$) denotes the time when the left (right) leaf arrives at the boundary of beamlet $j+1$. The red area illustrates the exposure time of beamlet j, which is a linear function of the arrival/departure times [Eq. (14)].

Let us assume that, while the gantry sweeps over the arc sector φ, the MLC leaves move unidirectional from left to right. With the notation introduced in Fig. 6, the fluence of beamlet (n,j) can be written as a linear function of the arrival and departure times,

$$x^{\varphi}_{jn} = \frac{\delta\varphi}{2}\left[(\overline{l}^{\varphi}_{nj}-\overline{r}^{\varphi}_{nj})+(\underline{l}^{\varphi}_{nj+1}-\underline{r}^{\varphi}_{nj+1})\right]. \qquad (14)$$

Thus, we can introduce the arrival and departure times as new optimization variables, and thereby directly optimize a sliding window trajectory over each arc sector φ. In order to obtain a valid and deliverable leaf trajectory, the arrival/departure times have to satisfy a set of linear constraints. For example, to account for maximum leaf speed, the constraint $\overline{l}^{\varphi}_{nj} + \Delta t \leq \underline{l}^{\varphi}_{nj+1}$ has to be satisfied, where Δt is the time required to traverse a beamlet at maximum leaf speed. Furthermore, the maximum treatment time can be controlled by the constraints $0 \leq \underline{l}^{\varphi}_{nj} \leq \overline{l}^{\varphi}_{nj} < T_{\varphi}$, where T_{φ} is the maximum time allowed for arc segment φ. Since (14) as well as all other leaf constraints are linear, the resulting optimization problem remains convex if the objective function $f(d)$ is convex. The result of the optimization is a piecewise linear approximation of deliverable leaf trajectories.[39]

The treatment time for this approach depends on the number of arc segments. For 2° arc sectors, the treatment time would be long. In Ref. 38, approximately 20 arc sectors are suggested. For typical leaf speed constraints, and depending on the treatment site, a 20 field FMO plan can be closely reproduced within 3–5 min. For large arc sectors such as 18°, the dose-influence matrix will not be constant over the arc sector, requiring a modification of the above method. To that end, Ref. 38 suggests an iterative reassignment of dose-influence matrices to beamlets during the optimization.

3.F. Extensions

Adding control points: Several methods have been suggested to improve on an existing VMAT plan by inserting additional control points. The intuition behind this approach is that some gantry arc sectors benefit from additional fluence modulation that is not realized by the current VMAT plan with one aperture per gantry angle. In other words, VMAT may benefit from a nonuniform distribution of control points. One such method has been termed FusionArc and is presented in Ref. 40. The method uses a variation of the column generation method discussed in Sec. 3.D to identify beam angles which potentially benefit from adding an intensity-modulated field. The method named Station Parameter Optimized Radiation Therapy (SPORT)[41] pursues the same goal but introduces a different demand metric, the modulation index (MI). An initial single arc VMAT plan is used to calculate the MI. Additional segments are then added in the vicinity of angles with high MI. Subsequently, the original plan and the added segments are jointly reoptimized to provide the final treatment plan.

4. DISCUSSION

4.A. Advantages and limitations

The different approaches to VMAT planning outlined in Sec. 3 have advantages and limitations. It can be hypothesized that the main improvements to VMAT planning will come from combining the different methods to capitalize on each method's strengths.

DAO using gradient based methods represents a powerful tool. One advantage is that the DAO formulation reflects exactly the DICOM specification of a treatment plan that is communicated to the treatment machine. Gradient based DAO is a module that can be added as a final refinement step to any other VMAT algorithm. Gradient based leaf refinement could in particular improve the column generation approach to VMAT discussed in Sec. 3.D. Improvements of gradient based DAO to the pure column generation approach have been demonstrated for step&shoot IMRT,[17,22] and it can be expected that these improvements will also be observed for VMAT.

Current implementations that heavily rely on gradient based DAO (PINNACLE, RayStation, MONACO) can potentially benefit from improved arc sequencing methods to reach better local optima. For example, a combination of the method described in Sec. 3.B with the graph based arc sequencing method of Sec. 3.A.2 could be investigated.

The main motivation behind the arc sequencing as discussed in Sec. 3.A is as follows: The fluence map optimization problem is well studied and yields globally optimal solutions for objective functions that are convex functions of dose. If it is possible to closely reproduce the optimal fluence through a VMAT plan, one can expect near-optimal plan quality from the VMAT plan. On the other hand, closely reproducing a large number of fluence maps requires time. In a sliding window type conversion, the leaves have to traverse the field once for each fluence map to be reproduced. Even if adequate results can be achieved with acceptable treatment times, reproducing fluence maps faithfully is expected to yield suboptimal delivery efficiency. In addition, arc sequencing of fluence maps typically leads to dose degradation due to distributing apertures over an arc sector (over which the dose-influence matrix varies). A final gradient based DAO step may therefore be beneficial for most arc sequencing methods.

The current results on sliding window VMAT plans such as direct leaf trajectory optimization (Sec. 3.E) or the work by Craft (Sec. 3.A.3) suggest that very high plan quality can be reached for acceptable treatment times in the order of 3–5 min, depending on the field size and machine parameters. Such approaches may represent a reliable way to obtain near optimal quality treatment plans, e.g., due to the convexity properties of the formulation in Sec. 3.E (i.e., generate the green dot in Fig. 2), but are not applicable to very short delivery times. It should be noted that these works currently lack an experimental verification of dosimetric accuracy and deliverability. These methods do not enforce a finite leaf gap for moving leaves, and do not address the final conversion of sliding window leaf trajectories into a sequence of control points that is communicated to the treatment machine in DICOM standard. Sliding window approaches may tend to produce narrow field openings, potentially leading to dosimetric inaccuracy. Thus, an experimental validation is warranted.

Directly comparing the performance of VMAT planning algorithms is hampered by the lack of shared datasets. A reliable comparison requires that algorithms are tested on common patient data sets, the same objective/constraint functions, and the same assumptions regarding delivery time estimates and dose calculation. Such studies have not been performed and require a collaborative effort. For that reason, no reliable conclusions can be drawn regarding the relative performance of algorithms at this stage.

4.B. Disconnect between TPS and machine control

In current practice, treatment planning systems have to specify a VMAT plan in DICOM format, i.e., leaf positions and gantry angle are specified for discrete values of cumulative MU, not time. The Linac control system subsequently performs the final conversion of the DICOM plan to leaf trajectories, dose rate, and gantry speed. VMAT planning and delivery could potentially benefit from a tighter integration of planning system and Linac controller. Currently, the TPS has to make assumptions on the treatment parameters in order to perform dose calculation and estimate the delivery time. However, at the moment there is no guarantee that the actual plan delivery matches the model that the TPS uses during planning.

4.C. Delivery time savings

The reduction of treatment time is typically thought of as one of the main motivations for VMAT. This is intuitive given that it is the defining feature of VMAT that dose is delivered while the gantry rotates. Thus, less treatment time is wasted by turning the beam off during treatment. However, it is instructive to take a critical look at the true time savings that can be attributed to the fact that the treatment beam is on during gantry rotation. To that end, we make the following argument: Let us consider a VMAT plan with 180 equally spaced control points. Let us further assume that the 360° arc is divided into 20 sectors. Each sector is 18° long and is associated with a leaf trajectory corresponding to nine control points. We can now assume that this plan is delivered using the dMLC technique while the gantry stays fixed in the middle of each arc sector. With minor adjustments of the trajectory to account for differences in the dose calculation, we can expect to obtain an equally good plan. In this case, it becomes apparent that the increase in delivery time is given by 1 min (the time for a full rotation) plus the time to accelerate and decelerate the gantry at each beam angle. Assuming 20 gantry angles and 3 s acceleration/deceleration time per angle, the latter would add another minute to the treatment time. This suggests that the delivery time for a 20 field IMRT plan is only 2 min longer compared to a similar VMAT plan. Publications on IMRT versus VMAT comparisons may report larger time savings for VMAT over step&shoot IMRT. It should be recognized that this is the empirically observed difference in delivery time for current commercial implementations of these technologies. The value does not reflect the time saving that can truly be attributed to allowing the treatment beam to be on during gantry rotation.[42] This also suggests that the treatment planning techniques that are now being developed for VMAT, could also be adapted to generate efficiently deliverable IMRT plans for dMLC delivery.

4.D. Multicriteria capabilities

In order to assess trade-offs between different planning goals, multicriteria optimization (MCO) methods have been developed. A MCO framework involves interactive navigation of the Pareto surface spanned by different treatment objectives. Continuous navigation of the Pareto surface involves forming convex combinations of precomputed data base plans. This is traditionally done by averaging fluence maps. Combining two treatment plans via averaging their fluence maps has the advantage that the corresponding dose distribution is exactly the average of the two plans dose distributions. This type of plan averaging is not generally applicable to VMAT plans. One approach to VMAT MCO planning consists of an initial navigation in fluence map space, followed by the generation of a VMAT plan that approximates the dose distribution that the planner navigated to Refs. 28 and 43. This approach is independent of the VMAT algorithm used, but bears the risk of discrepancies between the navigated and final dose distribution. In Ref. 44, it is shown that sliding window type delivery allows for VMAT plan averaging, which represents a potential advantage for MCO planning.

4.E. Noncoplanar generalizations of VMAT

VMAT treatments usually facilitate coplanar irradiation trajectories where the gantry rotates around the patient while the couch is fixed at 0°. Technically, however, it is already possible to realize arbitrary noncoplanar irradiation trajectories through simultaneous rotation of couch and gantry with a conventional linear accelerator (which may require access to a research mode of operation).[45]

For step&shoot IMRT, it has been established that such noncoplanar irradiation fields may allow for substantial dose reductions in OARs, especially if the OARs are close to irregularly shaped target volumes that are located asymmetrically within the body.[46–49] Leaving potential issues regarding patient immobilization on a robotic couch aside, the first treatment planning studies investigating noncoplanar VMAT have reconfirmed a benefit of noncoplanar irradiation with regard to OAR sparing.[33,45,47,50–53]

From the mathematical and technical side, the optimization of an irradiation trajectory shares the known issues of beam angle optimization (BAO) for step&shoot IMRT. When viewed as a continuous optimization problem, BAO is highly nonconvex.[54] In practice, BAO is typically formulated as a large combinatorial optimization problem, i.e., selecting a small number of beam angles from a larger pool of candidate beams.[55] Many BAO methods amount to solving a large number of plan optimization problems for different beam ensembles. This includes stochastic search methods[55] and integer programming methods.[56] These inherent challenges of BAO persist in noncoplanar VMAT planning.

Current approaches to noncoplanar VMAT planning pursue a two-step approach, which first determines the trajectory of the incident beam (i.e., a sequence of gantry/couch angle pairs), and then optimizes a VMAT plan along the predetermined trajectory. For the second step, all VMAT methods discussed in this paper can be applied with minor adjustments. The trajectories used for the above treatment planning studies were either found through inspection by a human planner,[50–52] using geometric heuristics,[45,53] or exhaustive search considering nine different couch tilts.[47] Wild *et al.*[33] use a noncoplanar beam ensemble which was optimized with a genetic algorithm[49] for step&shoot IMRT to construct a noncoplanar VMAT trajectory. By connecting the optimized discrete beam orientations to a rotation trajectory, they could reproduce the dosimetric quality of the noncoplanar step&shoot IMRT treatment plan. Their results also indicate that a simple couch tilt is not sufficient to deliver a meaningful dosimetric benefit compared to an arbitrary noncoplanar trajectory which involves simultaneous couch and gantry rotations. In the future, it will be interesting to see further developments of the underlying optimization strategies and if the results of the preliminary treatment planning studies generalize to larger patient cohorts.

5. CONCLUSION

This paper discusses the state of VMAT planning from an algorithmic perspective. The main approaches published in the literature as well as commercial implementations are reviewed. VMAT algorithms use the concepts of FMO, arc sequencing, and DAO. Different algorithms distinguish themselves through the component that is emphasized, as well as the exact implementation of each step. A rigorous comparison between different VMAT algorithms requires that all approaches are tested on identical data and the same objectives and constraints. This requires a collaborative effort, which has not been performed yet. In the absence of direct comparative results, it appears that improvements to VMAT planning may come from combining methods to benefit from each method's strengths. DAO using gradient information may represent a powerful tool in combination with most other algorithms, e.g., as final refinement step. Sliding window based VMAT shows promise in obtaining high quality treatment plans reliably for acceptable treatment times, but is not expected to yield the shortest possible delivery times. In addition, an experimental validation of sliding window approaches regarding dose accuracy and deliverability is warranted.

[a]Electronic mail: junkelbach@mgh.harvard.edu

[1]T. R. Mackie, "History of tomotherapy," Phys. Med. Biol. **51**, R427–R453 (2006).

[2]T. R. Mackie, T. Holmes, S. Swerdloff, P. Reckwerdt, J. O. Deasy, J. Yang, B. Paliwal, and T. Kinsella, "Tomotherapy: A new concept for the delivery of dynamic conformal radiotherapy," Med. Phys. **20**, 1709–1719 (1993).

[3]C. X. Yu, "Intensity-modulated arc therapy with dynamic multileaf collimation: An alternative to tomotheraphy," Phys. Med. Biol. **40**(9), 1435–1449 (1995).

[4]K. Otto, "Volumetric modulated arc therapy: IMRT in a single gantry arc," Med. Phys. **35**, 310–317 (2008).

[5]C. Wang, S. Luan, G. Tang, D. Z. Chen, M. A. Earl, and X. Y. Cedric, "Arc-modulated radiation therapy (amrt): A single-arc form of intensity-modulated arc therapy," Phys. Med. Biol. **53**, 6291–6303 (2008).

[6]S. Ulrich, S. Nill, and U. Oelfke, "Development of an optimization concept for arc-modulated cone beam therapy," Phys. Med. Biol. **52**, 4099–4119 (2007).

[7]S. M. Crooks, X. Wu, C. Takita, M. Watzich, and L. Xing, "Aperture modulated arc therapy," Phys. Med. Biol. **48**, 1333–1344 (2003).

[8]C. Cameron, "Sweeping-window arc therapy: An implementation of rotational IMRT with automatic beam-weight calculation," Phys. Med. Biol. **50**, 4317–4336 (2005).

[9]X. Y. Cedric and G. Tang, "Intensity-modulated arc therapy: Principles, technologies and clinical implementation," Phys. Med. Biol. **56**, R31–R54 (2011).

[10]For clarity: By VMAT planning, we refer to treatment planning for conventional hardware, i.e. Linacs equipped with MLCs. We exclude specialized hardware such as tomotherapy.

[11]An exception are dose-volume constraints, which are nonconvex. However, the nonconvexity does typically not coarse difficulties in finding high quality IMRT plans.

[12]T. Bortfeld, "The number of beams in IMRT: Theoretical investigations and implications for single-arc IMRT," Phys. Med. Biol. **55**, 83–97 (2010).

[13]J. D. Fenwick and J. Pardo-Montero, "Numbers of beam angles required for near-optimal IMRT: Theoretical limits and numerical studies," Med. Phys. **38**, 4518–4530 (2011).

[14]Unless the gantry stops.

[15]Unfortunately, current VMAT implementations do not always realize this potential.

[16]B. Hårdemark, A. Liander, H. Rehbinder, and J. Löf, "Direct machine parameter optimization with RayMachine in Pinnacle[3]," *White paper* (RaySearch Laboratories, Stockholm, Sweden, 2003).

[17]A. Cassioli and J. Unkelbach, "Aperture shape optimization for IMRT treatment planning," Phys. Med. Biol. **58**, 301–318 (2013).

[18]W. De Gersem, F. Claus, C. De Wagter, B. Van Duyse, and W. De Neve, "Leaf position optimization for step-and-shoot IMRT," Int. J. Radiat. Oncol., Biol., Phys. **51**, 1371–1388 (2001).

[19]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," Med. Phys. **29**, 1007–1018 (2002).

[20]Y. Li, J. Yao, and D. Yao, "Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy," Phys. Med. Biol. **48**, 3353–3374 (2003).

[21]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 J. Optim. **15**, 838–862 (2005).

[22]F. Carlsson, "Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy," Med. Phys. **35**, 3828–3838 (2008).

[23]The conversion of aperture weights to gantry speed and dose rate is performed by the machine controller outside of the TPS. Therefore, the TPS relies on approximate assumptions on the controller to estimate delivery time (see also discussion in Sec. 4.B)
[24]K. Bzdusek, H. Friberger, K. Eriksson, B. Hårdemark, D. Robinson, and M. Kaus, "Development and evaluation of an efficient approach to volumetric arc therapy planning," Med. Phys. **36**, 2328–2339 (2009).
[25]In this paper, we do not review FMO methods as this component does not require VMAT specific modifications. However, regularization of the FMO problem to favor smooth fluence maps is typically recommended for arc sequencing.
[26]S. Luan, C. Wang, D. Cao, D. Z. Chen, D. M. Shepard, and X. Y. Cedric, "Leaf-sequencing for intensity-modulated arc therapy using graph algorithms," Med. Phys. **35**, 61–69 (2008).
[27]D. Z. Chen, S. Luan, and C. Wang, "Coupled path planning, region optimization, and applications in intensity-modulated radiation therapy," Algorithmica **60**, 152–174 (2011).
[28]D. Craft, D. McQuaid, J. Wala, W. Chen, E. Salari, and T. Bortfeld, "Multicriteria VMAT optimization," Med. Phys. **39**, 686–696 (2012).
[29]E. Salari, J. Wala, and D. Craft, "Exploring trade-offs between VMAT dose quality and delivery efficiency using a network optimization approach," Phys. Med. Biol. **57**, 5587–5600 (2012).
[30]The nonlinear relations (6) and (7) define the objective function, but are not formal constraints themselves.
[31]P. E. Gill, W. Murray, and M. A. Saunders, "SNOPT: An SQP algorithm for large-scale constrained optimization," SIAM Rev. **47**, 99–131 (2005).
[32]J. L. Bedford, "Treatment planning for volumetric modulated arc therapy," Med. Phys. **36**, 5128–5138 (2009).
[33]E. Wild, M. Bangert, S. Nill, and U. Oelfke, "Non-coplanar VMAT for nasopharyngeal tumors: Plan quality versus treatment time" (2015).
[34]The gradient (12) exists everywhere if the objective function is one-time continuously differentiable in dose and z_j is smooth relative to the leaf positions. The piecewise linear definition of z_j according to Fig. 3 does not possess the necessary smoothness, but this issue can be resolved through the definition of a piecewise smooth gradient or the use of a more sophisticated physical fluence model.
[35]E. Vanetti, G. Nicolini, J. Nord, J. Peltola, A. Clivio, A. Fogliata, and L. Cozzi, "On the role of the optimization algorithm of rapidarc for volumetric modulated arc therapy on plan quality and efficiency," Med. Phys. **38**, 5844–5856 (2011).
[36]C. Men, H. E. Romeijn, X. Jia, and S. B. Jiang, "Ultrafast treatment plan optimization for volumetric modulated arc therapy (VMAT)," Med. Phys. **37**, 5787–5791 (2010).
[37]F. Peng, X. Jia, X. Gu, M. A. Epelman, H. E. Romeijn, and S. B. Jiang, "A new column-generation-based algorithm for VMAT treatment plan optimization," Phys. Med. Biol. **57**, 4569–4588 (2012).
[38]D. Papp and J. Unkelbach, "Direct leaf trajectory optimization for volumetric modulated arc therapy planning with sliding window delivery," Med. Phys. **41**, 011701 (10pp.) (2014).
[39]Note that, in order to communicate the VMAT plan through DICOM, the arrival/departure times have to be converted into a different piecewise linear leaf trajectory, i.e. control points.
[40]M. M. Matuszak, J. M. Steers, T. Long, D. L. McShan, B. A. Fraass, H. E. Romeijn, and R. K. Ten Haken, "FusionArc optimization: A hybrid volumetric modulated arc therapy (VMAT) and intensity modulated radiation therapy (IMRT) planning strategy," Med. Phys. **40**, 071713 (10pp.) (2013).
[41]R. Li and L. Xing, "An adaptive planning strategy for station parameter optimized radiation therapy (SPORT): Segmentally boosted VMAT," Med. Phys. **40**, 050701 (9pp.) (2013).
[42]In addition, planning comparisons between IMRT and VMAT do not always enforce the same planning criteria for both plans rigorously. For example, achiving target dose homogeneity and conformity in complex geometries tends to require strongly modulated fluence maps (requiring larger treatment times). If homogeneity is sacrificed in a VMAT plan more than in an IMRT plan, observed delivery time savings cannot be truly assigned to the delivery technique.
[43]R. Bokrantz, "Multicriteria optimization for volumetric-modulated arc therapy by decomposition into a fluence-based relaxation and a segment weight-based restriction," Med. Phys. **39**, 6712–6725 (2012).
[44]D. Craft, D. Papp, and J. Unkelbach, "Plan averaging for multicriteria navigation of sliding window IMRT and VMAT," Med. Phys. **41**, 021709 (5pp.) (2014).
[45]Y. Yang, P. Zhang, L. Happersett, J. Xiong, J. Yang, M. Chan, K. Beal, G. Mageras, and M. Hunt, "Choreographing couch and collimator in volumetric modulated arc therapy," Int. J. Radiat. Oncol., Biol., Phys. **80**, 1238–1247 (2011).
[46]M. Bangert and U. Oelfke, "Spherical cluster analysis for beam angle optimization in intensity-modulated radiation therapy treatment planning," Phys. Med. Biol. **55**, 6023 (2010).
[47]P. W. J. Voet, S. Breedveld, M. L. P. Dirkx, P. C. Levendag, and B. J. M. Heijmen, "Integrated multicriterial optimization of beam angles and intensity profiles for coplanar and noncoplanar head and neck IMRT and implications for VMAT," Med. Phys. **39**, 4858 (2012).
[48]S. Breedveld, P. R. M. Storchi, P. W. J. Voet, and B. J. M. Heijmen, "iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans," Med. Phys. **39**, 951–963 (2012).
[49]M. Bangert, P. Ziegenhein, and U. Oelfke, "Comparison of beam angle selection strategies for intracranial IMRT," Med. Phys. **40**, 011716 (11pp.) (2013).
[50]J. Krayenbuehl, J. B. Davis, and I. F. Ciernik, "Dynamic intensity-modulated non-coplanar arc radiotherapy (INCA) for head and neck cancer," Radiother. Oncol. **81**, 151–157 (2006).
[51]S. F. Shaitelman, L. H. Kim, D. Yan, A. A. Martinez, F. A. Vicini, and I. S. Grills, "Continuous arc rotation of the couch therapy for the delivery of accelerated partial breast irradiation: A treatment planning analysis," Int. J. Radiat. Oncol., Biol., Phys. **80**, 771–778 (2011).
[52]C. C. Popescu, W. A. Beckham, V. V. Patenaude, I. A. Olivotto, and M. T. Vlachaki, "Simultaneous couch and gantry dynamic arc rotation (CG-Darc) in the treatment of breast cancer with accelerated partial breast irradiation (APBI): A feasibility study," J. Appl. Clin. Med. Phys. **14**, 4035–4044 (2013).
[53]G. Smyth, J. C. Bamber, P. M. Evans, and J. L. Bedford, "Trajectory optimization for dynamic couch rotation during volumetric modulated arc radiotherapy," Phys. Med. Biol. **58**, 8163–8177 (2013).
[54]D. Craft, "Local beam angle optimization with linear programming and gradient search," Phys. Med. Biol. **52**, N127–N135 (2007).
[55]M. Bangert, P. Ziegenhein, and U. Oelfke, "Characterizing the combinatorial beam angle selection problem," Phys. Med. Biol. **57**, 6707–6723 (2012).
[56]E. Lee, T. Fox, and I. Crocker, "Integer programming applied to intensity-modulated radiation treatment planning," Ann. Oper. Res., Optim. Med. **119**, 165–181 (2003).