Medical Physics Letter

Ultrafast treatment plan optimization for volumetric modulated arc therapy (VMAT)

Chunhua Men
Center for Advanced Radiotherapy Technologies and Department of Radiation Oncology, University of California San Diego, La Jolla, California 92037-0843

H. Edwin Romeijn
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109-2117

Xun Jia and Steve B. Jiang[a)]
Center for Advanced Radiotherapy Technologies and Department of Radiation Oncology, University of California San Diego, La Jolla, California 92037-0843

(Received 26 May 2010; revised 26 July 2010; accepted for publication 30 August 2010; published 14 October 2010)

Purpose: To develop a novel aperture-based algorithm for volumetric modulated arc therapy (VMAT) treatment plan optimization with high quality and high efficiency.
Methods: The VMAT optimization problem is formulated as a large-scale convex programming problem solved by a column generation approach. The authors consider a cost function consisting two terms, the first enforcing a desired dose distribution and the second guaranteeing a smooth dose rate variation between successive gantry angles. A gantry rotation is discretized into 180 beam angles and for each beam angle, only one MLC aperture is allowed. The apertures are generated one by one in a sequential way. At each iteration of the column generation method, a deliverable MLC aperture is generated for one of the unoccupied beam angles by solving a subproblem with the consideration of MLC mechanic constraints. A subsequent master problem is then solved to determine the dose rate at all currently generated apertures by minimizing the cost function. When all 180 beam angles are occupied, the optimization completes, yielding a set of deliverable apertures and associated dose rates that produce a high quality plan.
Results: The algorithm was preliminarily tested on five prostate and five head-and-neck clinical cases, each with one full gantry rotation without any couch/collimator rotations. High quality VMAT plans have been generated for all ten cases with extremely high efficiency. It takes only 5–8 min on CPU (MATLAB code on an Intel Xeon 2.27 GHz CPU) and 18–31 s on GPU (CUDA code on an NVIDIA Tesla C1060 GPU card) to generate such plans.
Conclusions: The authors have developed an aperture-based VMAT optimization algorithm which can generate clinically deliverable high quality treatment plans at very high efficiency. [DOI: 10.1118/1.3491675]

Key words: VMAT, plan optimization, column generation method, GPU

I. INTRODUCTION

Volumetric modulated arc therapy (VMAT) is considered as one of the most promising radiotherapy technologies with great potential to improve the treatment quality. Additionally, a shorter delivery time indicates a reduced probability of treatment errors caused by patient motion during the treatment.

In a VMAT treatment process, a treatment gantry rotates around the patient while a radiation beam dynamically changes its aperture shape and associated intensity. By optimizing the beam aperture shape formed by a multileaf collimator (MLC) and the beam intensity at each gantry angle, a precisely sculpted desirable 3D dose distribution can be attained. This optimization problem is extremely complicated due to the very large scale of the problem and hardware constraints imposed on neighboring beam apertures and intensities. It can hardly be mathematically modeled in a concise and clean manner. Currently, most available algorithms to solve this problem are heuristic, which usually take up to hours to find a solution and cannot guarantee its optimality.[1–16] The use of such algorithms may limit the exploitation of VMAT's great potentials. In this letter, we present a novel aperture-based algorithm for VMAT treatment plan optimization with high plan quality and computational efficiency.

II. METHODS AND MATERIALS

II.A. Optimization model

We denote the number of beams by N and these beams are sorted based on the beam angles from 0° to 359°. Note that a beam aperture is a snapshot of the MLC leaf positions at a time point during the radiation dose delivery. Let us decompose each beam aperture into a set of beamlets and denote the set of beamlets exposed in beam k at angle θ_k by A_k. With beam k, we associate a decision variable y_k that indicates the intensity of that aperture. The set of voxels that represents the patient's CT image is denoted by V. In addition, we denote the dose to a voxel j by $z_j (j \in V)$ and it is calculated using a linear function of the intensities of the apertures through the so-called dose deposition coefficients D_{ij}: $z_j = \Sigma_{k=1}^{N} y_k \Sigma_{i \in A_k} D_{ij}$. D_{ij} is the dose received by the voxel $j \in V$ from the beamlet $i \in A_k$ at unit intensity. We calculate D_{ij}'s using our in-house dose calculation engine implemented on a general purpose graphics processing unit (GPU).[17]

Our VMAT optimization model employs a cost function with quadratic one-sided voxel-based penalties. Specifically, we write the cost function for a voxel $j \in V$ as

$$F(\mathbf{z}) = \alpha_j(\max\{0, T_j - z_j\})^2 + \beta_j(\max\{0, z_j - T_j\})^2, \quad (1)$$

where α_j and β_j represent the weights for underdosing and overdosing penalty, respectively. For target voxels, we set $\alpha_j > 0$ and $\beta_j > 0$ to penalize any deviation from the prescription dose T_j. As for critical structures, $\alpha_j = 0$ and $\beta_j > 0$ are chosen to add penalty for only those voxels received dose exceeding a threshold T_j.

In a VMAT system, the dose rate variation between neighboring angles is constrained within a certain range. To ensure the plan deliverability regarding this constraint, we add a smoothing term in the cost function to minimize difference between beam intensities at two neighboring beam directions, which is formulated as

$$G(\mathbf{y}) = \sum_{k=1}^{N-1} (y_{k+1} - y_k)^2/(\theta_{k+1} - \theta_k). \quad (2)$$

Our VMAT optimization model then can be written as

$$\min_{y, A_k} F(\mathbf{z}) + \gamma G(\mathbf{y}),$$

$$\text{subject to} \quad z_j = \sum_{k=1}^{N} y_k \sum_{i \in A_k} D_{ij},$$

$$y_k \geq 0 \quad k = 1, 2, \ldots, N, \quad (3)$$

where $\gamma > 0$ is a factor adjusting the relative weights between the two terms. Note that A_k, the set of beamlets in the aperture of beam k, is also a decision variable to be optimized.

II.B. Optimization algorithm

The extremely large dimensionality of the VMAT optimization problem poses a computational difficulty. In this letter, we solve this problem using a column generation method. This method has been successfully used to solve direct aperture optimization (DAO) problem for IMRT treatment planning in our previous studies.[18–20] In VMAT optimization problem, in addition to nonnegative beam intensity constraints and MLC hardware deliverability constraints, as in the DAO problem, there are two additional VMAT specific constraints: (1) The maximum leaf motion speed and (2) the maximum dose rate variation. Here, the first constraint is handled as a hard constraint while the second constraint is formulated as a penalty-based soft constraint in the objective function.

In our VMAT treatment plan optimization, a single 360° gantry rotation is discretized into uniformly spaced 180 beam angles. In the column generation method, we generate one deliverable aperture at each beam angle and find out the associated beam intensity. Apertures are generated one by one in a sequential way. We start our algorithm without any initial apertures. At each iteration, by solving a so-called subproblem (also called a *pricing* problem), a new aperture is generated for a beam angle that is not occupied by another aperture generated in previous iterations. Note that this beam angle is automatically selected out of all unoccupied angles while solving the subproblem. Moreover, generating a new aperture in this subproblem has to account for all the deliverability constraints imposed by the MLC system. After the new aperture is generated, a master problem is then followed to find the optimal intensities associated with those currently already generated apertures. This iteration terminates when every beam angle attains one aperture. Specifically, the column generation method is conducted as following.

For the subproblem, by checking the KKT conditions,[21] we first obtain the "price" w_i for each beamlet i.[18–20] The KKT conditions are necessary and sufficient for a solution to be optimal in nonlinear convex programming, provided that some regularity conditions are satisfied. The price of a beamlet characterizes the degree we could lower the objective value if we include this beamlet in the aperture. It is our goal at this step to find a set of beamlets to form a deliverable aperture at one of the unoccupied beam angles which has the lowest total price, i.e., which potentially decreases the objective function most effectively. Let K_U denote the set of all deliverable apertures in those unoccupied beam angles, then the subproblem becomes

$$\min_{A_k \in K_U} \sum_{i \in A_k} w_i. \quad (4)$$

We seek the solution to Eq. (4) through the following three steps. First, the deliverability of the solution aperture A_k is ensured by checking the left and right MLC leaf positions in each MLC row at the neighboring occupied beam angles. Second, the minimization in Eq. (4) can be realized row by row: For each row of MLC, finding a consecutive set of beamlets for which the sum of their w_i values is minimized. This can be achieved by passing through all beamlets in each row from left to right only once.[19] The above two steps result in a set of consecutive beamlets in each MLC row at all unoccupied beam angles. We group those rows at each angle as an aperture and pick the aperture at one of the

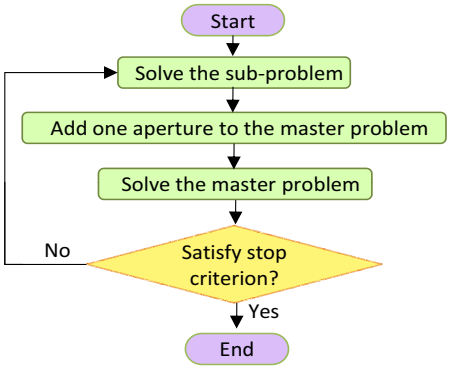

FIG. 1. A flowchart of our algorithm for solving the VMAT plan optimization problem.

unoccupied beam angles which gives the smallest objective values as the final solution to this subproblem in the last step.

As for the master problem, the objective function (3) is simply of a quadratic form given those already determined apertures A_k in previous steps. We therefore utilize a gradient projection method to solve this master problem. The flowchart of column generation method for VMAT treatment plan optimization algorithm is summarized in Fig. 1.

II.C. GPU implementation

GPU offers a potentially powerful computational platform for convenient and affordable high-performance computing and researchers have been starting to use GPU in solving heavy duty problems in a clinical context.[17,20,22–25] To speed up the VMAT optimization algorithm, we implemented the column generation method on GPU under the Compute Unified Device Architecture (CUDA). The GPU implementation is very similar to our DAO implementation[20] and therefore not described in this letter.

III. RESULTS AND DISCUSSION

Five clinical prostate cases (P1–P5) and five clinical head-and-neck cases (H1–H5) were used to evaluate our new algorithm in terms of treatment plan quality and planning efficiency. For prostate cases, the prescription dose to planning target volume (PTV) was 73.8 Gy and for the head-and-neck cases, the prescription dose was 73.8 Gy to PTV1 and 54 Gy to PTV2. PTV1 consists of the gross tumor volume expanded to account for both subclinical disease as well as daily setup errors and internal organ motion; PTV2 is a larger target that also contains high-risk nodal regions and is again expanded for same reasons. For all cases, we used a beamlet size of 10×10 mm^2 and voxel size of $2.5\times2.5\times2.5$ mm^3 for target and organs at risk (OARs). For unspecified tissue (i.e., tissues outside the target and OARs), we increased the voxel size in each dimension by a factor of 2 to reduce the optimization problem size. The full resolution was used when evaluating the treatment quality via dose volume histograms (DVHs), dose color wash, isodose curves, etc. The case dimensions are showed in Table I.

Figure 2 shows two typical VMAT plans for a prostate case and a head-and-neck case. For the head-and-neck case, there are many more critical structures used in the optimization, such as brain stem, optic nerve, spinal cord, etc., whose doses are very low and thus DVH curves are not shown in Fig. 2 for clarity purpose.

As we can see in Table I, our new algorithm has very high planning efficiency. To generate a high quality plan, it only takes 5–8 min with the MATLAB implementation on an Intel Xeon 2.27 GHz CPU and 18–31 s with the CUDA implementation on an NVIDIA Tesla C1060 GPU card, as shown in Table I, which makes VMAT a possible treatment delivery technique for online adaptive radiation therapy.

The shape of an aperture does not change once it is generated, while the intensity of this aperture is still adjustable at each iteration by solving the master problem. We would like to point out that the shapes of initially generated apertures may not be optimal since they were generated in a sequential manner without taking into account the shapes of the downstream apertures. Fortunately, this theoretical deficiency does not translate into a practical problem. We found that after generating one aperture for each beam angle using our method, adding more apertures to a beam angle does not significantly improve the quality of a VMAT plan. This is because a large number of beam angles are used here (180 angles in this work) and the contribution of each individual

TABLE I. Case dimensions and CPU/GPU running time on an Intel Xeon 2.27 GHz CPU and an NVIDIA Tesla C1060 GPU for our VMAT plan optimization implementations.

Case	No. of beamlets	No. of voxels	No. of nonzero D_{ij}'s ($\times10^7$)	CPU time (s)	GPU time (s)
P1	40 620	45 912	2.3	340	22
P2	59 400	48 642	3.2	265	18
P3	38 880	28 931	1.8	276	20
P4	43 360	39 822	2.6	410	26
P5	51 840	49 210	3.0	348	23
H1	51 709	33 252	2.5	290	21
H2	78 874	59 615	5.0	468	27
H3	90 978	74 438	5.5	342	25
H4	71 280	31 563	2.6	363	25
H5	53 776	42 330	3.5	512	31

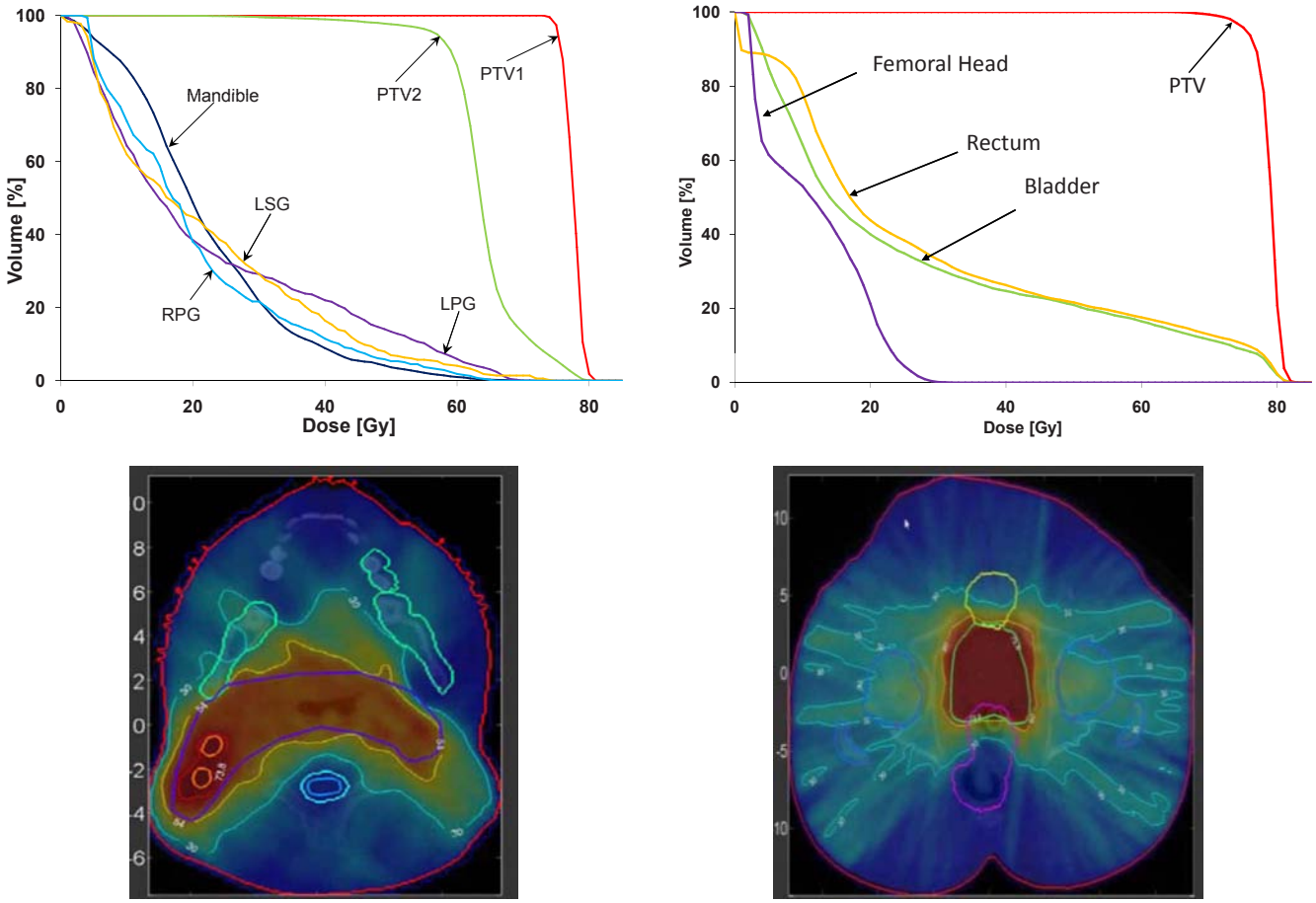

FIG. 2. VMAT plans for a head-and-neck case (left) and a prostate case (right) (LSG: Left submandibular gland; LPG: Left parotid gland; RPG: Right parotid gland).

beam angle is relatively small. In fact, when the optimization is finished, the "not-so-optimal" apertures are found to have small intensities and thus minimal contributions to the overall cost function.

Although the dose rate constraints are formulated as penalty-based soft constraints in the cost function, we still require them to be satisfied at the end of the optimization. Note that we only have to satisfy these dose rate constraints at the last iteration after generating the last aperture. If these constraints are not satisfied at the last iteration, the coefficient γ in the objective function will be automatically increased and then the intensity for each aperture will be reoptimized until the dose rate constraints are satisfied. A similar approach can also be used to handle the maximum and minimum dose rate constraints, which are not considered in this preliminary work.

ACKNOWLEDGMENTS

The authors would like to thank NVIDIA for providing GPU cards. This work is supported in part by the University of California Laboratory Fees Research Program.

[a]Electronic mail: sbjiang@ucsd.edu

[1]S. Luan, C. Wang, D. L. Cao, D. Z. Chen, D. M. Shepard, and C. X. Yu, "Leaf-sequencing for intensity-modulated arc therapy using graph algorithms," Med. Phys. **35**, 61–69 (2008).

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

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

[4]P. Zhang, L. Happersett, M. Hunt, A. Jackson, M. Zelefsky, and G. Mageras, "Volumetric modulated arc therapy: Planning and evaluation for prostate cancer cases," Int. J. Radiat. Oncol., Biol., Phys. **76**, 1456–1462 (2010).

[5]J. L. Bedford, "Treatment planning for volumetric modulated arc therapy," Med. Phys. **36**, 5128–5138 (2009).

[6]K. Bzdusek, H. Friberger, K. Eriksson, B. Hardemark, D. Robinson, and M. Kaus, "Development and evaluation of an efficient approach to volumetric arc therapy planning," Med. Phys. **36**, 2328–2339 (2009).

[7]G. Tang, M. Earl, and C. Yu, "Variable dose rate single-arc IMAT delivered with a constant dose rate and variable angular spacing," Phys. Med. Biol. **54**, 6439–6456 (2009).

[8]A. Fogliata, A. Clivio, G. Nicolini, E. Vanetti, and L. Cozzi, "Intensity modulation with photons for benign intracranial tumours: A planning comparison of volumetric single arc, helical arc and fixed gantry techniques," Radiother. Oncol. **89**, 254–262 (2008).

[9]D. Palma, E. Vollans, K. James, S. Nakano, V. Moiseenko, R. Shaffer, M. McKenzie, J. Morris, and K. Otto, "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 (2008).

[10]F. Kjær-Kristoffersen, L. Ohlhues, J. Medin, and S. Korreman, "RapidArc volumetric modulated therapy planning for prostate cancer patients," Acta Oncol. **48**, 227–232 (2009).

[11]L. Masi, F. Casamassima, R. Doro, A. Colantuoni, I. Bonucci, and C. Menichelli, "VMAT plans solution with single and double gantry rotation

compared with IMRT step & shoot for different treatment sites," Int. J. Radiat. Oncol., Biol., Phys. **75**, S706–S707 (2009).
[12]R. Shaffer, W. J. Morris, V. Moiseenko, M. Welsh, C. Crumley, S. Nakano, M. Schmuland, T. Pickles, and K. Otto, "Volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for simultaneous maximal intraprostatic boost: A planning comparison study," Clin. Oncol. **21**, 401–407 (2009).
[13]T. Bortfeld, "The number of beams in IMRT-theoretical investigations and implications for single-arc IMRT," Phys. Med. Biol. **55**, 83–97 (2010).
[14]C. Popescu, I. Olivotto, W. Beckham, W. Ansbacher, S. Zavgorodni, R. Shaffer, E. Wai, and K. Otto, "Volumetric modulated arc therapy improves dosimetry and reduces treatment time compared to conventional intensity-modulated radiotherapy for locoregional radiotherapy of left-sided breast cancer and internal mammary nodes," Int. J. Radiat. Oncol., Biol., Phys. **76**, 287–295 (2010).
[15]R. Shaffer, A. M. Nichol, E. Vollans, M. Fong, S. Nakano, V. Moiseenko, M. Schmuland, R. Ma, M. McKenzie, and K. Otto, "A comparison of volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for frontal and temporal high-grade gliomas," Int. J. Radiat. Oncol., Biol., Phys. **76**, 1177–1184 (2010).
[16]G. Tang, M. A. Earl, S. Luan, C. Wang, M. M. Mohiuddin, and C. X. Yu, "Comparing radiation treatments using intensity modulated beams, multiple arcs, and single arcs," Int. J. Radiat. Oncol., Biol., Phys. **76**, 1554–1562 (2010).
[17]X. Gu, D. J. Choi, C. Men, H. Pan, A. Majumdar, and S. B. Jiang, "GPU-based ultra-fast dose calculation using a finite size pencil beam model," Phys. Med. Biol. **54**, 6287–6297 (2009).
[18]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).
[19]C. Men, H. E. Romeijn, Z. C. Tas, and J. F. Dempsey, "An exact approach to direct aperture optimization in IMRT treatment planning," Phys. Med. Biol. **52**, 7333–7352 (2007).
[20]C. Men, X. Jia, and S. B. Jiang, "GPU-based ultra-fast direct aperture optimization for online adaptive radiation therapy," Phys. Med. Biol. **55**, 4309–4319 (2010).
[21]M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. (Wiley, Hoboken, 2006).
[22]R. Li, X. Jia, J. Lewis, X. Gu, M. Folkert, C. Men, and S. B. Jiang, "Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy," Med. Phys. **37**, 2822–2826 (2010).
[23]X. Gu, H. Pan, Y. Liang, R. Castillo, D. S. Yang, D. J. Choi, E. Castillo, A. Majumdar, T. Guerrero, and S. B. Jiang, "Implementation and evaluation of various demons deformable image registration algorithms on a GPU," Phys. Med. Biol. **55**, 207–219 (2010).
[24]C. Men, X. Gu, D. J. Choi, A. Majumdar, Z. Zheng, K. Mueller, and S. B. Jiang, "GPU-based ultrafast IMRT plan optimization," Phys. Med. Biol. **54**, 6565–6573 (2009).
[25]X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, "Cone beam CT reconstruction from undersampled and noisy projection data via total variation," Med. Phys. **37**, 1757–1760 (2010).