Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **4**, Number **2**, May **2019**

Aperture Generation for Intensity-modulated Radiotherapy Scheme based on Gradient Information

Jie Yang[1,2], Zhi-Guo Gui [1,3], Li-Yuan Zhang[1], Peng-Cheng Zhang[1*]

[1]National Key Laboratory for Electronic Measurement Technology,
North University of China, Taiyuan Shanxi 030051, China
[2]School of Medicine Management,
Shanxi University of TCM, Taiyuan 030619, China
[3]Key Laboratory of Instrumentation Science and Dynamic Measurement of Ministry of Education,
North University of China, Taiyuan Shanxi 030051, China
*Corresponding Author: zhangpc198456@163.com

Received March 2018; revised December 2018

Abstract. *We propose an algorithm for aperture generation (AG) for step and shoot delivery of intensity-modulated radiotherapy. The method is an approach to direct aperture optimization (DAO) that exploits gradient information to initialize the aperture shapes. Because the relationship between dose distribution and leaf positions of multileaf collimators is not linear, new apertures are generated by using a region growing segmentation algorithm, rather than solving a shortest-path problem. This yields a well-behaved optimization problem for the aperture shapes, which can be solved efficiently. In this paper, the AG algorithm is embedded into a DAO approach. After a new aperture is generated based on the region growing segmentation algorithm and added to the treatment plan, we optimize aperture weights and leaf positions for the new set of apertures. We tested the proposed method on four prostate cases using the Computational Environment for Radiotherapy Research algorithm for our dose calculation. The computational results indicate that the proposed AG algorithm can produce highly conformal treatment plans and decrease optimization time.*
Keywords: Intensity-modulated radiotherapy; Direct aperture optimization; Gradient; Region growing segmentation.

1. **Introduction.** Treatment planning for intensity-modulated radiotherapy (IMRT) has conventionally employed a two-step process. The first stage is called the fluence map optimization (FMO) problem, which involves finding an optimal intensity profile for each beam. In the second stage, the fluence maps are decomposed into a set of apertures that are deliverable using a multileaf collimator (MLC). The main disadvantage of the two-step process is that a loss in treatment quality can occur. To address this drawback, direct aperture optimization (DAO)—which directly optimizes the aperture shapes and weights—was developed [1-14]. There are basically two types of DAO approaches, and each one usually starts with different initial aperture shapes. One type optimizes the shape and weight of a predetermined number of apertures using a deterministic or stochastic search (also called aperture modulation). In the works of Shepard et al. [1], Earl et al. [2], and Cao et al. [5], the apertures are initialized with beam-eye-view (BEV) projections of the target shapes. The original methods of Li et al. [3] and Cotrutz and Xing [4] are initialized with leaf sequencing after FMO. Methods based on deterministic or stochastic search methods rely on a good initial set of apertures, which can enable them

to escape local minima. Nonetheless, they often cannot guarantee that all deliverable apertures are considered. The second approach involves column generation, which starts with an empty set of apertures and iteratively adds apertures to the treatment plan [6-14]. In the generic column-generation approach described by Romeijn et al. [7], the treatment plan iteratively adds apertures by solving a shortest-path problem without modifying the existing apertures. However, Carlsson [13] proved that the plan improvement rate is much higher with the adjustable aperture approach than with the fixed aperture approach early in the solution process. Meanwhile, Cassioli and Unkelbach [12] proposed an algorithm that uses a trust-region-based local search approach combined with column generation to optimize the aperture shapes. However, because the relationship between dose distribution and the leaf positions of multileaf collimators is not linear, the shortest-path problem is complex, with a variable corresponding to each possible aperture shape. Consequently, this approach tends to generate large apertures, and is not sufficiently fast. On the basis of the above comparison and considerations, to improve the DAO algorithm and reduce computation time, an algorithm for aperture generation (AG) based on gradient information is herein proposed. The proposed method is different from those in previous studies insofar as new apertures are generated using a region growing segmentation algorithm, rather than solving a shortest-path problem. The AG algorithm is embedded into a DAO approach; the optimization scheme starts with an empty set of apertures, and then generates a new aperture based on the AG algorithm. After the aperture is added to the treatment plan, we optimize aperture weights and leaf positions for the new set of apertures. The above procedure is then iteratively executed until the termination condition is reached. In addition, the constraints imposed by the MLC are incorporated during optimization. In this paper, results for four prostate cases are presented. The remainder of this paper is organized as follows. In Section 2, the mathematical model and solution for the proposed AG method are introduced. In Section 3, the computational results are presented. Finally, the conclusions are summarized in Section 4.

2. **Material and Methods.**

2.1. **Cost Function.** Traditionally, some criteria are used to construct the cost function, whereas others constitute constraints. In the proposed approach, on the other hand, the weighted sum of a sub-objective function is used. The cost function can then be formulated:

Traditionally, some criteria are used to construct the cost function, whereas others constitute constraints. In the proposed approach, on the other hand, the weighted sum of a sub-objective function is used. The cost function can then be formulated:

$$minimize \quad F = \sum_{i=1}^{I} \delta_i f_i(D) \tag{1}$$

where I is the number of sub-objective functions, weight δ_i reflects the importance of the ith sub-objective function, and $f_i(D)$ is the ith sub-objective function. In addition, D is the dose distribution vector defined on the patient voxels. There are many different types of sub-objectives [15-23]. They correspond to specifications for a variety of factors ranging from the dose to the region of interest (ROI). In this work, four different specification types are used. The sub-objective function for the specification $D_{min} \geq yG_y$, representing the stipulation that no ROI voxel should receive a dose less than yGy, is given by

$$f_{\min}(D) = \frac{1}{V} \sum_{v=1}^{V} H(y - d_v) \cdot (y - d_v)^2 \tag{2}$$

$$H(y-d_v)=\begin{cases}1 & d_v<y\\ 0 & d_v\geq y\end{cases} \tag{3}$$

where Vspecifies the voxels included in the ROI; and d_v is the dose in the voxel v. For the specification $D_{min}\leq yG_y,$, representing the stipulation that the uniform dose to the ROI should not exceed yGy, the sub-objective function is given by

$$f_{\text{mean}}(D)=\frac{1}{V}\sum_{v=1}^{V}(d_v-y)^2 \tag{4}$$

For the specification$V_{d_1}\leq y\%$, representing the stipulation that the volume of the ROI receiving doses greater than d_1 should not exceed $y\%$, the sub-objective function is given by

$$f_{\text{DVH}}(D)=\frac{1}{V}\sum_{v=1}^{V}H(d_v-d_1)\cdot H(d_2-d_v)\cdot(d_v-d_1)^2 \tag{5}$$

$$H(d_v-d_1)=\begin{cases}1 & d_v>d_1\\ 0 & d_v\leq d_1\end{cases} \tag{6}$$

$$H(d_2-d_v)=\begin{cases}1 & d_2>d_v\\ 0 & d_2\leq d_v\end{cases} \tag{7}$$

where d_2 is in the current dose volume histogram (DVH) and $V_{(d2)}=y\%$. For the specification $NTCP\leq y\%$, representing the stipulation that the normal tissue complication probability (NTCP) does not exceed $y\%$, the sub-objective function is given by

$$f_{\text{NTCP}}(D)=H(\ln(1-y))-\ln(1-NTCP(D))\cdot(\ln(1-y))-\ln(1-NTCP(D)) \tag{8}$$

$$H(\ln(1-y))-\ln(1-NTCP(D))=\begin{cases}1 & NTCP(D)>y\\ 0 & NTCP(D)\leq y\end{cases} \tag{9}$$

where

$$NTCP(D)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{gEUD(D)-D_{50}}{mD_{50}}}e^{\frac{-t^2}{2}}\,\mathrm{dt} \tag{10}$$

in which

$$gEUD\,(D)=\left(\frac{1}{V}\sum_{v\in V}d_v^a\right)^{1/a} \tag{11}$$

D_{50} is the whole organ dose for which the NTCP is 50%. In addition, m is obtained by fitting the tolerance doses for the uniform whole and partial organ irradiation to the function [24], and a is a control parameter [25]. Choi and Deasy [26] proved that a convex function will be minimized when minimizing gEUD(di) to normal tissues ($a\geq 1$). Alternatively, when maximizing the $gEUD(d_i)$ of the target ($a<1$), a concave function will be maximized.

In FMO, the sub-objective function can usually be formulated by the beamlet intensities. By contrast, for DAO, the objective function cannot be easily formulated in the closed form as a function of the leaf positions. For the proposed AG method, a mathematical model was developed to formulate dose D as a linear function of aperture shapes and aperture weights on the basis of the dose-influence matrix. An advantage of this mathematical model is that the highly nonlinear and non-convex problem of aperture optimization is decomposed into a sequence of well-behaved optimization problems that can be efficiently solved using standard algorithms. For the sake of convenience, the positions of leaves contained within the MLC replace the aperture shapes. We assume that the aperture set is denoted as K, the set of weights is denoted as W, and each aperture $k\in K$has a weight $w_k\in W(w_k>0)$. The MLC consists of M identical leaf pairs, which

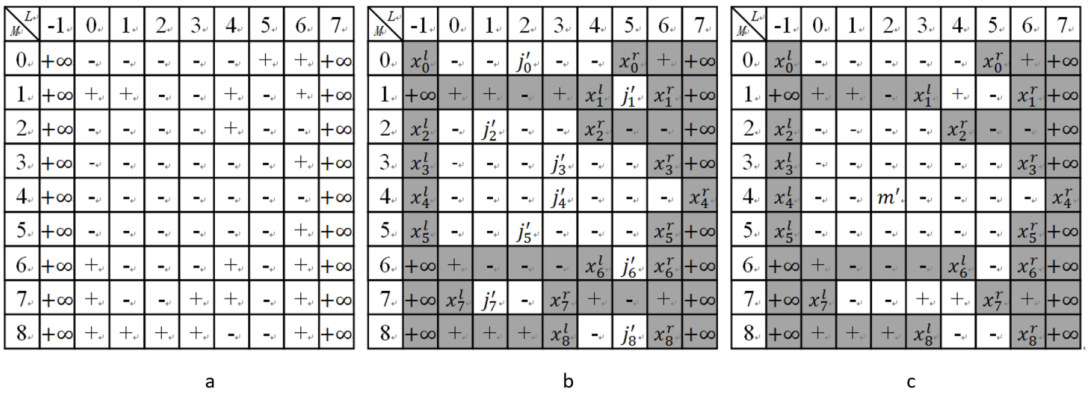


FIGURE 1. Illustration of aperture generation (M = 9, L = 8)

we also refer to as rows. Each row has length L, and the positions of the left and right leaves, respectively, of the m^{th} row in the k^{th} aperture are denoted by $x^{\mathrm{l}}_{mk}, x^{\mathrm{r}}_{mk} \in [-1, L]$. Dose D can be mathematically expressed as

$$D = P \sum_{k \in K} B_k w_k \tag{12}$$

$$B_k = \bigcup_{m \in M, j \in L} \left(b^j_{mk}\right) \tag{13}$$

$$b^j_{mk} = \Psi_{mj}(x^{\mathrm{l}}_{mk}, x^{\mathrm{r}}_{mk}) = \begin{cases} 1, & x^{\mathrm{l}}_{mk} < j < x^{\mathrm{r}}_{mk} \\ 0, & j \le x^{\mathrm{l}}_{mk} \text{ or } x^{\mathrm{r}}_{mk} \le j \end{cases} \tag{14}$$

where P is the dose-influence matrix that contains the dose to each patient voxel from each pencil beam, and B_k is the set of binary vectors specifying the beamlets that are exposed by aperture k. Meanwhile, b^j_{mk} denotes the binary variable to bixel j from row m in aperture k as a step function of the leaf position. To improve the quality of the deliverable apertures, we consider the following two common sets of hardware constraints. *(i)* The consecutiveness constraint simply corresponds to the fact that apertures are shaped by pairs of leaves; i.e., in each given row m, each voxel should be exposed consecutively. *(ii)* The interdigitation constraint holds that, in addition to *(i)*, the left leaf of a row cannot overlap with the right leaf of an adjacent row.

2.2. **AG method.** In the generic column-generation approach, the shape of the new aperture is determined by solving a shortest-path problem as described by Carlsson [13]. However, the relationship between the dose distribution and the leaf positions of a multi-leaf collimator is not linear, making the shortest-path problem overly complex. Unlike the generic column-generation approach, the proposed AG method generates new apertures using a simple method—namely, the region growing segmentation method. In the aperture generation module, one aperture per beam can be generated with the AG method; i.e., the leaf is assigned to a bixel. This is conducted in three serial steps to obtain the aperture shape. The first step is to calculate the gradient matrix of the cost function. The second step is to directly assign the leaf of each row to a bixel using the region growing segmentation method. The third step is to use the matrix's minimum gradient as a starting point and impose constraints on the leaf positions. A detailed description of the procedure is summarized in what follows.

First, the gradient matrix of the cost function is computed. Let $g_{mk}(j)$ denote the gradient of cost function F with respect to the beamlet j in row m in aperture k, which

is evaluated for the current aperture set $(\bar{K}, \bar{W})$. Meanwhile, $g_{mk}(j)$ is given by

$$g_{mk}(j) = \frac{\partial F}{\partial D} P_{mkj} w_k \tag{15}$$

Second, to solve the problem of assigning the leaf of each row to a bixel, we apply region growing to open or close the bixel. We define some seed points and then calculate the connectivity between each seed bixel and its neighborhoods. The main steps for aperture generation based on region growing are as follows:

Step 1: Selection of initial seed points. The matrix's minimum gradient and each row's minimum gradient is found quickly. Let $g_{minmk} = \min_{j \in L} g_{mk}(j)$ denote the minimum gradient of row m in aperture k, and let the position of g_{minmk} be(m,j'_m). We select the initial seed points for region growing according to the minimum gradient of row m in aperture k. The set S of initial seed points is defined as

$$S = \{g_{minmk} | m \in M\} \tag{16}$$

Step 2: Region growing. If any bixel belonging to the seed point's 2-connected neighborhood (those to the left and right of the seed points) satisfies the similarity criterion $g_{mk}(j) < 0$, we add it to the seed point set.

Step 3: Termination. The left leaf is assigned to bixel $j + 1$ when the termination condition is met (i.e., if moving to the left $g_{mk}(j)$ is positive for the first time). Similarly, the right leaf is assigned to bixel j–1 (i.e., if moving to the right $g_{mk}(j)$ is positive for the first time). Note that we set $g_{mk}(-1)$ and $g_{mk}(L)$ to large values.

In the third stage, we must ensure that the new aperture remains deliverable. Thus, we have to ensure that the selected positions x^{l}_{mk}, x^{r}_{mk} for all $m \in M$ will lead to deliverable solutions. We therefore use the G_{mink} as a starting point, and going line-by-line, up or down, have satisfy

$$\begin{array}{ll} x^{\mathrm{l}}_{mk} < x^{\mathrm{r}}_{mk} & \forall k \in K, m \in M, x^{\mathrm{l}}_{mk} \in [-1, L], x^{\mathrm{r}}_{mk} \in [-1, L] \\ x^{\mathrm{l}}_{mk} < x^{\mathrm{r}}_{m-1k} & \forall k \in K, m \in M, x^{\mathrm{l}}_{mk} \in [-1, L], x^{\mathrm{r}}_{mk} \in [-1, L] \\ x^{\mathrm{r}}_{mk} > x^{\mathrm{l}}_{m-1k} & \forall k \in K, m \in M, x^{\mathrm{l}}_{mk} \in [-1, L], x^{\mathrm{r}}_{mk} \in [-1, L] \end{array} \tag{17}$$

where $G_{mink} = \min_{m \in M, j \in L} g_{mk}(j)$ denotes the minimum gradient of aperture k, and the position of G_{mink} is(m',j'_m). Finally, constraints on the minimum aperture size are imposed. If the number of non-zero b_{mj} values for the new aperture $(m \in M, j \in L)$ exceeds the lower limit, it is added to $\bar{K}$.

Algorithm 1

Input: M,L,k,j'_m,m'
Output: B_k
1 For $m = 0{:}M$ // the region growing
2 For $j = j'_m : L$ // position the right leaf
3 if $g_{mk}(j) > 0$ then
4 $x^{r}_{mk} \leftarrow j$
5 break;
6 end
7 $j \leftarrow j + 1$
8 next j
9 For $j = j'_m : 0$ // position the left leaf
10 if $g_{mk}(j) > 0$ then
11 $x^{l}_{mk} \leftarrow j$
12 break;
13 end
14 $j \leftarrow j - 1$
15 next j
16 $m \leftarrow m + 1$
17 next m
18 For $m = m' - 1 : 0$ // The upper part of B_k
19 if $x^{l}_{mk} \geq x^{r}_{m+1k}$ then
20 $x^{l}_{mk} \leftarrow x^{r}_{m+1k} - 1$
21 end
22 if $x^{r}_{mk} \leq x^{l}_{m+1k}$ then
23 $x^{r}_{mk} \leftarrow x^{l}_{m+1k} + 1$
24 end
25 next m
26 For $m = m' + 1{:}M - 1$ // The lower part of B_k
27 if $x^{l}_{mk} \geq x^{r}_{m-1k}$ then
28 $x^{l}_{mk} \leftarrow x^{r}_{m-1k} - 1$
29 end
30 if $x^{r}_{mk} \leq x^{l}_{m-1k}$ then
31 $x^{r}_{mk} \leftarrow x^{l}_{m-1k} + 1$
32 end
33 next m

2.3. **Integration into A DAO Approach.** The proposed AG algorithm is next embedded into a DAO approach. This approach is similar to the one used by Cassioli et al. [12] and Wang et al. [14], but our approach replaces the shortest-path problem with the AG method. Simultaneous optimization of the aperture shape and weight is indeed possible, but because there is a coupling, optimization is difficult. Thus, it is preferable to handle them separately. To optimize the weights $\bar{W}$ of the current aperture set, the weight optimization module is entered after a new aperture is added to the treatment plan. This procedure is equivalent to Eq. (12) with bk fixed. The extension of the L-BFGS algorithm for bound-constrained problems (L-BFGS-B) [27] is used to optimize all weights for the set of apertures. This is because this restricted problem is much easier to solve owing to the number of variables being smaller and D being linear in $\bar{W}$. This is in contrast to the non-convexity introduced by the leaf position variables, and it ensures that the weights are optimal or close to optimal so the plan can be saved. The aperture

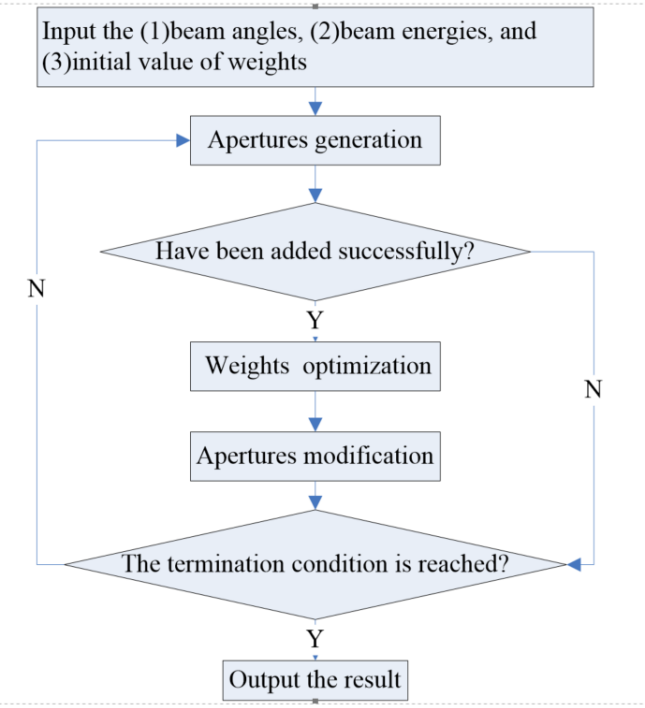


FIGURE 2. Flow chart of the AG method

weights optimization problem is given by

$$minimize_W \quad F = \sum_{i=1}^{I} \delta_i f_i(D) \tag{18}$$

subject to $w_k > 0, B_k$ fixed. Then, the aperture modification module is entered. In the approach described by Carlsson [13], apertures with weights at the lower limit will be removed in the aperture removal module to improve the plan, but this makes computation extremely complex. Cassioli and Unkelbach [12] exploited gradient information to locally optimize the positions of the leaves of a multileaf collimator. Thus, we exploit the aperture shape optimization (ASO) algorithm to modify the existing apertures $\bar{K}$. We repeat the procedure until the relative improvement is less than 10^{-1} between two consecutive iterations. The overall scheme for the complete DAO algorithm is shown in Fig. 2.

3. Experiments and Analysis.

3.1. **Algorithm Details.** In this section, we investigate the feasibility and performance of the proposed method in four clinical prostate tumor cases. We investigate the integration of the AG algorithm into a DAO approach. Specifically, we compare the following two algorithms: $ag - our$ DAO algorithm in which a new aperture is generated with the AG algorithm (Algorithm 1); $sp - the$ DAO proposed by Cassioli et al. [12] in which a new aperture is generated with shortest-path problem algorithm.

In these cases, the planning target volume (PTV) was the prostate, and the organs at risk (OARs) were the bladder and rectum. Normal tissue was 5 cm outside the PTV. For all four cases, we designed plans using five $6 - MV$ photon beams. The five beams were distributed around the patient with angles of $36^o, 100^o, 180^o, 260^o$, and 324^o, respectively. Figure 3 shows an example of the target delineation. The matrix of the beamlet dose deposition coefficients was computed using the pencil-beam dose calculation algorithm through the Computational Environment for Radiotherapy Research (CERR) interface [28]. The proposed method was implemented in a Microsoft Visual $C++$ environment on

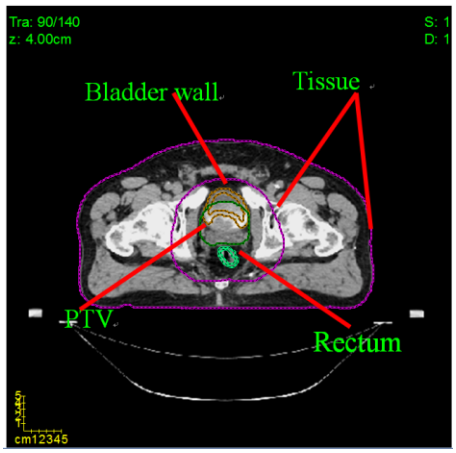


FIGURE 3. Typical CT slice illustrating the target and the critical structure delineation

TABLE 1. Total volumes (cubic cm) for the four prostate tumor cases

Case	Bladder	Rectum	PTV	Tissue
1	27.81	141.10	186.83	21893.97
2	29.34	84.79	103.72	26032.31
3	30.15	54.76	136.86	12212.74
4	34.34	66.81	164.56	14727.72

TABLE 2. Weights of cost function for the four cases

CASE	δ_1	δ_2	δ_3	δ_4	δ_5
1	30	120	30	30	5
2	45	60	30	30	5
3	30	120	30	30	5
4	60	50	30	30	5

a personal computer equipped with an Intel Core i3 CPU@ 3.50 GHz and 4 GB memory. The termination criterion was defined as a relative improvement of less than 10^{-1} between two consecutive iterations. Table 1 summarizes the model dimensions for the four prostate tumor cases.

We considered a biological indices (NTCP-based) cost function:

$$\begin{aligned}&\delta_1 f_{\mathrm{NTCP}}(D^{\mathrm{Bladder}}) + \delta_2 f_{\mathrm{NTCP}}(D^{\mathrm{Rectum}})\\ &+\delta_3 f_{\mathrm{mean}}(D^{\mathrm{PTV}}) + \delta_4 f_{\mathrm{min}}(D^{\mathrm{PTV}}) + \delta_5 f_{\mathrm{DVH}}(D^{\mathrm{Normal}})\end{aligned} \tag{19}$$

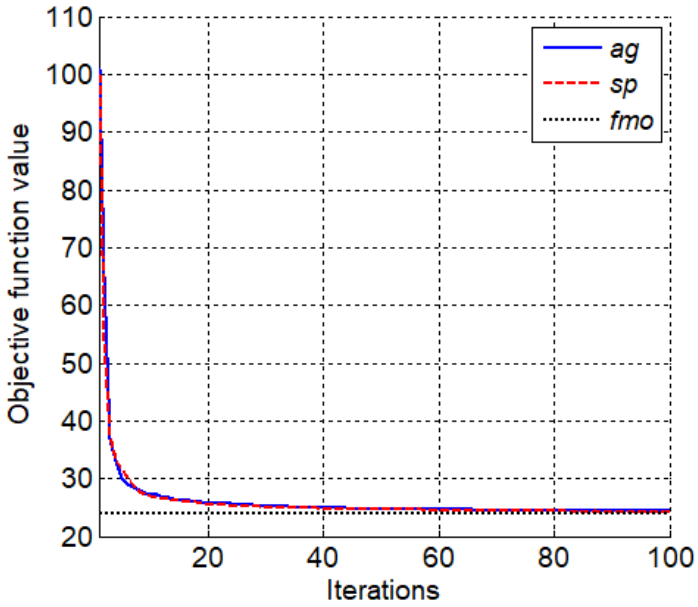
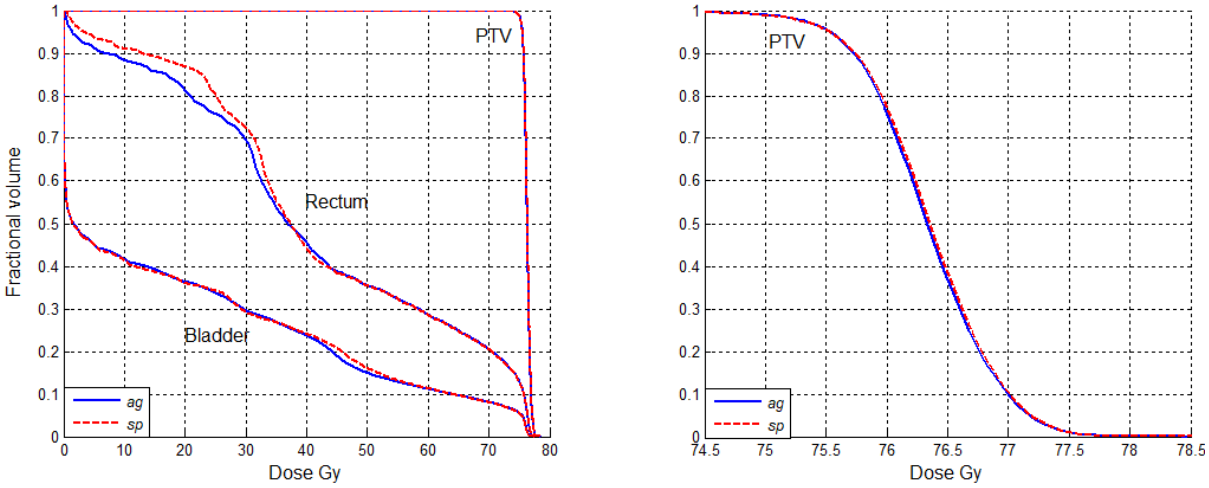


FIGURE 4. Objective function values for case 1 for NTCP-based cost function

TABLE 3. Parameters for the NTCP

Organ at Risk	a	m	D_{50}(Gy)
Bladder	8	0.11	62
Rectum	8	0.14	80.7



(A) a (B) b

FIGURE 5. DVHs of case 1: (a) all DVHs, and (b) DVHs of PTV

Table 2 shows the weights of the NTCP-based cost function (Eq. (19)) for the four cases. The parameters for NTCP [29, 30] are tabulated in Table 3. The mean dose for the PTV was 78 Gy, and the minimum dose for the PTV was 74 Gy.

TABLE 4. Dose for case 1

	ag			*sp*		
	Bladder	Rectum	PTV	Bladder	Rectum	PTV
Mean dose/Gy	19.21	41.57	76.36	19.32	42.74	76.37
Max. dose/Gy	77.48	77.52	78.48	77.63	77.57	78.52
Min. dose/Gy	0.03	0.13	71.53	0.03	0.58	71.36

TABLE 5. Final value of cost function, number of apertures, and runtime

CASE	*ag*			*sp*		
	Final value	No. of apertures	Runtime (min.)	Final value	No. of apertures	Runtime (min)
1	24.37	35	29.38	24.30	38	37.78
2	14.34	32	20.04	15.50	32	19.03
3	33.81	29	15.80	33.09	35	27.57
4	23.58	30	15.20	23.67	30	15.83

TABLE 6. ag: DVH criteria

C A S E	PTV			Bladder		Rectum	
	@93%	@100%	@110%	@65Gy	@80Gy	@50Gy	@75Gy
1	100.0	99.0	0.0	9.6	0.0	35.5	13.1
2	100.0	43.5	0.0	13.3	0.0	39.5	11.2
3	100.0	98.7	0.0	37.7	0.0	42.7	16.2
4	100.0	98.8	0.0	32.3	0.0	32.4	11.5

TABLE 7. sp: DVH criteria

C A S E	PTV			Bladder		Rectum	
	@93%	@100%	@110%	@65Gy	@80Gy	@50Gy	@75Gy
1	100.0	98.9	0.0	9.7	0.0	35.4	18.2
2	99.9	48.0	0.0	13.3	0.0	42.8	11.7
3	100.0	98.6	0.0	36.4	0.0	41.2	16.3
4	100.0	98.7	0.0	30.7	0.0	34.6	11.9

3.2. **Evaluate Criteria.** We used clinical DVH criteria to objectively verify the ability of our models to create clinically acceptable treatment plans. The criteria employed are based on the current clinical guidelines formulated by Marks et al. [31]:

PTV

– At least 99% of the PTV should receive 93% of the prescribed dose ($0.93 \times 75Gy$).

– At least 95% should receive the prescribed dose (75 Gy).
– No more than 10% should be overdosed by more than 10% of the prescribed dose (1.1 × 75 Gy).
– No more than 1% should be overdosed by more than 20% of the prescribed dose (1.2 × 75 Gy).

Rectum

– No more than 50% should be overdosed by more than 50 Gy.
– No more than 35% should be overdosed by more than 60 Gy.
– No more than 25% should be overdosed by more than 65 Gy.
– No more than 20% should be overdosed by more than 70 Gy.
– No more than 15% should be overdosed by more than 75 Gy.

Bladder

– No more than 50% should be overdosed by more than 65 Gy.
– No more than 35% should be overdosed by more than 70 Gy.
– No more than 25% should be overdosed by more than 75 Gy.
– No more than 15% should be overdosed by more than 80 Gy.

3.3. **Convergence Properties.** In Fig. 4, we plot the objective function value as a function of the iteration. The ag and sp algorithms stop after at most 100 iterations. The dotted line indicates the objective function obtained for f_{mo} (the FMO algorithm as proposed by Zhang et al. [32], used as an ideal solution benchmark. The details of the progress of FMO have been published elsewhere [32-34].). Note that this objective function refers to the ideal FMO solution without a sequencing step.

The test results show that both *ag* (solid line) and *sp* (dashed line) come substantially close to the f_{mo} solution (dotted line). They also show that the *ag* algorithm is convergent and stable. This indicates that, with an increasing number of iterations, at some point the incremental improvement in the treatment plan quality may no longer be clinically significant. Hence, rather than allowing the ag algorithm to formally converge, we propose terminating the algorithm when it is observed that the treatment plan quality with respect to a particular criterion has not markedly improved in recent iterations, i.e., when the relative improvement between successive iterations is less than 10^{-1}.

3.4. **Comparison of Plans.** Figure 5 compares the DVHs of the final plan for the two algorithms. The solid lines indicate the DVHs obtained for *ag*; those for sp are indicated by dashed lines. From the DVH comparison in Fig. 6, it can be confirmed that ag yields treatment plans that are close to the *sp* solution based on the NTCP cost function. The proposed method is thus efficient and effective. We observe the same dose coverage on the PTV for these two optimization results. The mean doses for PTV were *76.36 Gy* and *76.38 Gy* for *ag* and *sp*, respectively. For the OARs, both results met the current clinical guidelines. However, ag's overall protection performance for OARs, shown in Table 4, was slightly better than sp's. Table 4 shows the mean dose, maximum dose, and minimum dose for the OARs and PTV.

We computed the final cost function value, number of apertures, and runtime (in minutes), all of which can be found in Table 5. Tables 6 and 7 show the values of the DVH criteria for the targets and main critical structures for all four cases of *ag* and *sp*, respectively. For example, the data in the column labeled PTV@93% represent the fraction of volume (in %) of PTV receiving 0.93 × 75 Gy. The labels of the other columns follow a similar format.

In comparing Tables 5, 6, and 7, it is apparent that treatment plans of a very similar quality can be obtained. We observe in the results that there is the same dose coverage for

the PTV using the *sg* method, the *ag* method, and objective function values. However, the *ag* approach can reduce the programs' execution time by over 15.90% on average.

4. **Conclusions.** We demonstrated the feasibility of a direct aperture optimization method based on gradient information for IMRT. The proposed AG algorithm generates new apertures using the region growing segmentation method, leading to the fewest possible apertures and the simplest possible aperture shapes. All of the cases investigated demonstrate that the proposed AG algorithm provides simple, effective IMRT solutions, and results that approximate the ideal FMO solution. In particular, for the examples considered here, the AG algorithm obtained slightly better treatment plan quality and less optimizing time compared to the column-generation algorithm. On average, the proposed algorithm reduced the programs' execution time by 15.90%. The proposed approach facilitates a tradeoff between treatment plan quality and delivery efficiency. Future research can extend this work by explicitly incorporating a treatment plan efficiency measure, such as reducing the number of apertures per beam.

Acknowledgment. This work was supported in part by the National Nature Science Foundation of China under Grant 61271357, in part by the National Nature Science Foundation of China under Grant 11605160, in part by the National Key Scientific Instrument and Equipment Development Project under Grant 2014YQ24044508, in part by the opening project of State Key Laboratory of Explosion Science and Technology (Beijing Institute of Technology) under Grant KFJJ13-11M, and in part by Research Project Supported by Shanxi Scholarship Council of China under Grant 2016-089.

REFERENCES

[1] D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. X. Yu, Direct aperture optimization: A turnkey solution for step-and-shoot IMRT, *Med. Phys.*, vol. 29, pp. 1007-1018, 2002.

[2] M. A. Earl, M. K. N. Afghan, C. X. Yu, Z. Jiang, and D. M. Shepard, Jaws-only IMRT using direct aperture optimization, *Med. Phys.*, vol. 34, pp. 307-314, 2007.

[3] Y. Li, J. Yao, and D. Yao, Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy, Phys. Med. Biol., vol. 48, pp. 3353-3374, 2003.

[4] C. Cotrutz and L. Xing, Segment-based dose optimization using a genetic algorithm, Phys. *Med. Biol.,* vol. 48, p. 2987, 2003.

[5] R. Cao, X. Pei, H. Zheng, L. Hu and Y. Wu, Direct aperture optimization based on genetic algorithm and conjugate gradient in intensity modulated radiation therapy, *Chin. Med. J.,* vol. 127, pp. 4152-4153, 2014.

[6] F. Preciado-Walters, M. P. Langer, R. L. Rardin, and V. Thai, Column generation for IMRT cancer therapy optimization with implementable segments, *Ann. Operat. Res.*, vol. 148, pp. 65-79, 2006.

[7] 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.,* vol. 15, pp. 838-862, 2005.

[8] C. Men, H. E. Romeijn, Z. C. Taskin, and J. F. Dempsey, An exact approach to direct aperture optimization in IMRT treatment planning, *Phys. Med. Biol.*, vol. 52, pp. 7333-7352, 2007.

[9] E. Salari, C. Men, and H. E. Romeijn, Accounting for the tongue-and-groove effect using a robust direct aperture optimization approach, *Med. Phys.,* vol. 38, pp. 1266-1279, 2011.

[10] G. Bednarz, D. Michalski, C. Houser, M. S. Huq, Y. Xiao, P. R. Anne, and J. M. Galvin, The use of mixed-integer programming for inverse treatment planning with pre-defined field segments, *Phys. Med. Biol.*, vol. 47, pp. 2235-2245, 2002.

[11] E. Salari and J. Unkelbach, A column-generation-based method for multi-criteria direct aperture optimization, *Phys. Med. Biol.,* vol. 58, pp. 621-639, 2013.

[12] A. Cassioli and J. Unkelbach, Aperture shape optimization for IMRT treatment planning, *Phys. Med. Biol.,* vol. 58, pp. 301-318, 2012.

[13] F. Carlsson, Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy, *Med. Phys.,* vol. 35, pp. 3828-3838, 2008.

[14] J. Wang, X. Pei, R. Cao, L. Hu, and Y. Wu, A multiphase direct aperture optimization for inverse planning in radiotherapy, *Nucl. Sci. Tech.*, vol. 26, pp. 48-53, 2015.
[15] J. F. Fowler, The linear-quadratic formula and progress in fractionated radiotherapy, *Brit. J. Radiol.*, vol. 62, pp. 679-694, 1989.
[16] A. Brahme, Treatment Optimization Using Physical and Radiobiological Objective Functions in Smith (ed.), Radiation Therapy Physics, Springer, Berlin, pp. 209-246, 1995.
[17] G. J. Kutcher and C. Burman, Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method, *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 16, pp. 1623-1630, 1989.
[18] M. Alber and R. Reemtsen, Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method, *Optim. Method. Softw.,* vol. 22, pp. 391-411, 2007.
[19] J. O. Deasy, Multiple local minima in radiotherapy optimization problems with dose–volume constraints. *Med. Phys.*, vol. 24, pp. 1157-1161, 1997.
[20] J. Llacer, J. O. Deasy, T. R. Portfeld, T. D. Solberg, and C. Promberger, Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints, *Phys. Med. Biol.*, vol. 48, pp. 183-210, 2003.
[21] H. E. Romeijn, J. F. Dempsey, and J. G. Li, A unifying framework for multi-criteria fluence map optimization models, *Phys. Med. Biol.*, vol. 49, pp. 1991-2013, 2004.
[22] I. E. Naqa, G. Suneja, P. E. Lindsay, A. J. Hope, J. R. Alaly, and V. Vicic, Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose–volume outcome relationships, *Phys. Med. Biol.*, vol. 51, pp. 5719-5735, 2006.
[23] J. Zhu, B. S. Li, H. Z. Shu, T. Bai,Y. Yin, and L. M. Luo, Review of normal tissues complication probability models, Chin. J. Biomed. Eng., vol. 33, pp. 233-240, 2014.
[24] J. T. Lyman, "Complication probability as assessed from dose-volume histograms," *Radiat. Res. Suppl.,* vol. 8, pp. S13-S19, 1985.
[25] A. Niemierko, A generalized concept of equivalent uniform dose (EUD), *Med. Phys.,* vol. 26, p. 1100, 1999.
[26] B. Choi and J. O. Deasy, The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning, *Phys. Med. Biol.*, vol. 47, pp. 3579-3589, 2002.
[27] R. H. Byrd, P. Lu, and J. Nocedal, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comp.*, vol. 16, pp. 1190-1208, 1995.
[28] J. O. Deasy, A. I. Blanco, and V. H. Clark, CERR: a computational environment for radiotherapy research, *Med. Phys.*, vol. 30, pp. 979-985, 2003.
[29] E. Dale, T. P. Hellebust, A. Skjønsberg, T. Høgberg, and D. R. Olsen, Modeling normal tissue complication probability from repetitive computed tomography scans during fractionated high-dose-rate brachytherapy and external beam radiotherapy of the uterine cervix, *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, pp. 963-971, 2000.
[30] S. T. H. Peeters, M. S. Hoogeman, W. D. Heemsbergen, A. A. M. Hart, P. C. M. Koper, and J. V. Lebesque, Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: Normal tissue complication probability modeling. *Int. J. Radiat. Oncol. Biol. Phys.,* vol. 66, pp. 11-19, 2006.
[31] L. B. Marks, E. D. Yorke, and A. Jackson, Use of normal tissue complication probability models in the clinic, *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 76, pp. 10-19, 2010.
[32] L. Y. Zhang, P. C. Zhang, Z. G. Gui, H. Z. Shu, and J. Yang, Plan optimization method of IMRT based on biological indices, *Appl. Res. of Comp.,* vol. 34, no. 5, pp. 1303-1307, 2017.
[33] J. Yang, and Z. G. Gui, "Base on Study Of L-BFGS and NSGA-II Hybrid Calculation to IMRT Inverse Planning Optimization, *J. North Uni. Chin.*, vol. 6, pp. 706-712, 2015.
[34] J. Yang, P. C. Zhang, L. Y. Zhang, H. Z. Shu, B. S Li, and Z. G. Gui, Particle swarm optimizer for weighting factor selection in intensity-modulated radiation therapy optimization algorithms, *Phys. Med.*, vol. 33, pp. 136-145, 2017.