IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 5, 2019, accepted January 25, 2019, date of publication January 31, 2019, date of current version March 25, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2896175*

# A Column Generation Approach Based on Region Growth

**LIYUAN ZHANG[1], ZHIGUO GUI[1], JIE YANG[1,2], AND PENGCHENG ZHANG[1]**

[1]Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan 030051, China
[2]School of Medicine Management, Shanxi University of Chinese Medicine, Taiyuan 030619, China

Corresponding author: Pengcheng Zhang (zhangpc198456@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11605160, in part by the Research Project supported by the Shanxi Scholarship Council of China under Grant 2016-089, in part by the National Key Scientific Instrument and Equipment Development Project of China under Grant 2014YQ24044508, in part by the National Natural Science Foundation of China under Grant 61671413, and in part by the National Key Research and Development Program of China under Grant 2016YFC0101605.

**ABSTRACT** In intensity-modulated radiation therapy (IMRT), a network flow is adopted to solve the pricing problem of the generic column generation approach in order to obtain a deliverable aperture. However, excessive computation results from the direct use of a network flow. In addition, a decline in plan quality may result from the direct determination of the leaf position using the gradient information. To overcome these problems, a column generation approach based on region growth is proposed. The proposed method is designed to reduce the computational cost of solving the pricing problem and improve the IMRT plan quality. First, the gradients of the beamlets are obtained by an objective function constructed under the constraint conditions of the organs. Second, the gradients are transformed nonlinearly. Third, the positions of the continuous negative gradient regions in each row of the aperture are determined and stored. Fourth, these gradients are taken as a whole and added to the aperture network flow, which is solved as a shortest-path problem. Finally, the deliverable aperture is obtained and added to the treatment plan. To verify the effectiveness of the proposed method, experiments involving five five-field prostate cancer cases and five nine-field head and neck cancer cases were conducted. Compared with the generic column generation method, the dose distribution of the target is ensured by the proposed method, which also effectively protects organs at risk and reduces the running time. Specifically, in ten groups of comparative experiments, the normal tissue complication probability of the proposed method is reduced by up to 3.37%, and the maximum acceleration rate is 20.44%. According to the experimental results, the proposed method is more consistent with clinical requirements compared with the generic column generation method.

**INDEX TERMS** Column generation, direct aperture optimization, image processing, intensity-modulated radiation therapy, region growth.

## I. INTRODUCTION

Intensity-modulated radiation therapy (IMRT) is an advanced three-dimensional conformal radiotherapy technology. In IMRT, the two known methods for realizing the static intensity-modulated mode are the two-step approach [1]–[4] and direct aperture optimization (DAO) [5]–[8]. In the two-step approach, a fluence map that satisfies the clinical requirements is initially generated. Then, a sequencing step is performed to obtain the deliverable apertures and corresponding monitor units (MUs). However, the dose distribution of the fluence map is degraded after the sequencing step [9]. The sequencing step is avoided in DAO by direct optimization of the aperture shape and intensity [9]. It involves two steps: aperture shape optimization (ASO) and aperture weight optimization (AWO) [10]. In ASO, the deliverable aperture that provides the largest potential improvement in the objective function is identified and added to the treatment plan. In AWO, gradient-based optimization methods are typically used to optimize the aperture weights [9]. In DAO, the physical constraints of the multileaf collimator (MLC) (e.g., the leaf interdigitation limitation) [10] are incorporated into the optimization [7]. Thus, compared to the two-step approach, the IMRT plan quality is considerably improved with DAO [11].

In general, to achieve DAO, each beam is modeled as a collection of hundreds of small beamlets. The intensity of each beamlet is considered to be independently controllable.

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Zhang.



2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

The gradients of all beamlets, calculated by the objective function, constitute a gradient map, which is used to search for a deliverable aperture shape. A change in leaf position that satisfies any of the constraints imposed by the MLC is accepted if the objective function decreases. Such a change indicates the opening or closing of the rays irradiated to the beamlets. When the summation of the gradients of the beamlets involved in such a change is negative, the objective function decreases when these beamlets are added to the aperture. Hence, the corresponding rays should be opened. On the other hand, when the summation of the gradients is positive, the rays corresponding to the beamlets should be closed. Thus, the gradients of the beamlets are important parameters for generating the aperture shape. Existing DAO methods [12]–[18] are categorized into the following types.

1) Stochastic search methods: In these methods, small changes in the leaf position are stochastic. A change in the leaf position that improves the objective function is accepted. Otherwise, the current modification is accepted with a probability of skipping the local optimum. This method was first proposed by Shepard *et al.* [12], who used the simulated annealing method to solve the optimization problem. Subsequently, this method was improved by Earl *et al.* [19]. The optimization problem was solved by Li *et al.* [20] and Cotrutz and Xing [21] using a genetic algorithm.

2) Local gradient-based methods: In these methods, the leaf position is used as the optimization variable. The relationship between the objective function and the leaf position is established, and the first derivative is given. Such algorithms have been applied to various commercial therapeutic systems, including the direct machine parameter optimization model used in Pinnacle and RayStation systems [22].

3) Column generation methods: In these methods, the initial apertures are not set at the beginning of an iteration; instead, deliverable apertures with iteration are produced. The deliverable apertures are individually added to the treatment plan. The iteration process involves two steps. First, the pricing problem is solved to generate the deliverable aperture that can provide the largest improvement in the objective function, and this aperture is added to the treatment plan. Then, the weights of the new set of apertures are reoptimized in the master problem. Several researchers have studied this method [23]–[29]. The column generation method applied to mixed photon–electron planning can efficiently generate treatment plans, as demonstrated by Renaud *et al.* [30].

Stochastic search methods and gradient-based methods start with an initial set of apertures. By contrast, new suitable apertures are added to the treatment plan by the column generation method instead of modifying existing apertures. However, because a network flow is directly adopted by the generic column generation algorithm to solve the pricing problem in order to obtain the deliverable apertures, its computational cost is extremely high.

Furthermore, for the generic column generation algorithm, there is a complex relationship between the aperture shape and the gradients of the beamlets. Direct determination of the leaf position using the gradients may result in a suboptimal aperture shape [27]. For example, a positive value in the gradient map may be canceled by the surrounding negative values. Therefore, a beamlet with this positive gradient value may be grouped into the newly generated aperture shape. The addition of this aperture shape into the treatment plan may degrade the plan quality. Thus, in order to overcome this problem, beamlets with large positive gradient values should be excluded from the aperture shape, and beamlets with small negative gradient values should be included in the aperture shape as much as possible.

Therefore, to reduce the excessive computation due to the direct use of the network flow in the column generation algorithm and to improve the quality of the plan, a novel column generation approach based on region growth is proposed herein. In an iteration of the proposed method, the gradient map of the aperture is first obtained. Then, the gradients are nonlinearly transformed to improve the plan quality. Next, for each row of gradients in the gradient map, the positions of the continuous negative gradients are obtained and preserved by region growth. Further, the adjacent negative gradients of each row are integrated into a single element in the gradient map, with the corresponding gradient value being the sum of the gradients of the beamlets in that element. The excessive computation is effectively reduced by these two steps. Then, the network flow is formulated by a gradient map after region growth and merging, and the pricing problem is solved to obtain the deliverable aperture. Finally, the master problem is solved, and this iteration is completed.

The proposed method aims to reduce the number of computations required to solve the pricing problem of column generation and improve the quality of the plan; that is, this work focuses on solving the pricing problem and evaluating the plan quality. The remainder of this paper is organized as follows. The theory of the proposed method is introduced in Section II. In Section III, the experimental settings and experimental results are presented. In Section IV, the experimental results are analyzed and discussed. Finally, the conclusions are summarized in Section V.

## II. EXPERIMENTAL THEORY AND METHODS

In this section, a new mathematical model for solving the pricing problem is described. The dose is formulated as a function of the leaf positions in Section II.A. In Section II.B, the solution process of the algorithm is described, and the new column generation algorithm based on region growth is discussed Section II.C.

### A. DOSE CALCULATIONS

The structures of a phantom include both targets and critical structures. They are irradiated using a predetermined set of beams. Each beam is decomposed into a rectangular grid of beamlets $B$ comprising $m$ rows and $n$ columns. In this study, the size of a typical beamlet is 0.5 cm $\times$ 0.5 cm. The set of deliverable apertures is denoted by $K$, and the weights corresponding to these apertures are denoted by $y_k (k \in K)$.

Let $A_k$ denote the set of beamlets that are exposed in the aperture $k$, and let $S$denote the total number of structures. Each structure is discretized into a finite number of voxels $v_s$. The dose received by voxel $j$ ($j = 1, \cdots, v_s$) in structure $s$ ($s = 1, \cdots, S$) from beamlet $i$ of aperture $A_k$at unit intensity is denoted by $W_{ijs}$, i.e., the deposition coefficient. Therefore, the dose $D_{js}$ received by voxel $j$ in structure $s$ is given by

$$D_{js} = \sum_{k \in K} \left( \sum_{i \in A_k} W_{ijs} \right) y_k, \quad j = 1, \cdots, v_s, \ s = 1, \cdots, S. \tag{1}$$

### B. SOLUTION PROCESS

For step-and-shoot IMRT, the plan optimization problem is formulated as

$$minmize \ F(D_s) = minmize \sum_{s=1}^{S} \sum_{j=1}^{N_s} F_{js}(D_s), \tag{2}$$

where $F(D_s)$ is the objective function, $F_{js}(D_s)$ is the $j$th subobjective function applied to structure $s$, and $D_s$ is the dose distribution of structure $s$. To control the dose distribution delivered to a structure, $N_s$subobjective functions are applied to this structure. A column generation algorithm is implemented to solve this optimization problem and generate the deliverable treatment plan. The column generation algorithm includes the solution of the pricing problem and master problem. Fig. 1 shows the solution process of the generic column generation algorithm.

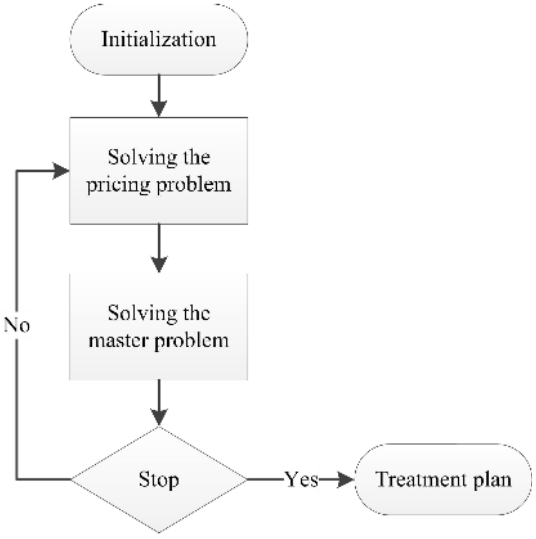


**FIGURE 1.** Flowchart of the generic column generation algorithm.

As shown in Fig. 1, the solution process consists of two main steps. In the first stage of the loop, a new deliverable aperture shape is generated and added to the treatment plan by solving the pricing problem. Next, the weights of the new set of apertures are optimized using the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm for bound constrained optimization (L-BFGS-B) [31]–[33] to solve the master problem. L-BFGS-B is a quasi-Newton algorithm. L-BFGS-B is an advanced version of the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS). If the planner is satisfied with the plan or if the maximum number of loops is reached, the solution process is terminated. Otherwise, the next loop is entered.

#### 1) PRICING PROBLEM

In the process of solving the pricing problem with the generic column generation algorithm, the beam is first decomposed into a rectangular grid of beamlets. Then, the gradient information of each beamlet at the current dose is calculated, and the negative gradient map is created according to the locations of the beamlets. Finally, because the new aperture is required to minimize the objective function, the sum of the gradient components corresponding to the beamlets contained in the new aperture should be as small as possible. The new aperture that is added to the plan is the smallest sum of the gradient components. Thus, the pricing problem is formulated as a network flow, which is solved as a shortest-path problem. Furthermore, when generating a new aperture, the mechanical constraints of the MLC, such as the connectivity, interdigitation, and minimum gap, are considered.

As IMRT planning optimization is a large-scale optimization problem with constraints, it can be solved using the Karush–Kuhn–Tucker (KKT)-conditions for optimality. The Lagrange function is constructed using the objective function and constraint conditions of the optimization problem as follows:

$$\begin{aligned} &L(D_{js}, y_k, \pi_{js}, \rho_k) \\ &\quad = \sum_{s=1}^{S} \sum_{j=1}^{v_s} F_{js}(D_{js}) + \sum_{k=1}^{K} \rho_k(-y_k) \\ &\qquad + \sum_{s=1}^{S} \sum_{j=1}^{v_s} \pi_{js} \left( \sum_{k \in K} \left( \sum_{i \in A_k} W_{ijs} \right) y_k - D_{js} \right), \end{aligned} \tag{3}$$

where $\rho_k$ and $\pi_{js}$ are the KKT multipliers. The following five conditions must be satisfied:

$$\nabla_{D_{js}, y_k} L\left(D_{js}, y_k, \pi_{js}, \rho_k\right) = 0, \tag{4}$$

$$-y_k \le 0, \quad k = 1, \cdots, K, \tag{5}$$

$$\sum_{k \in K} \left( \sum_{i \in A_k} W_{ijs} \right) y_k - D_{js} = 0, \quad j = 1, \cdots, v_s, \ s = 1, \cdots, S, \tag{6}$$

$$\rho_k \ge 0, \quad k = 1, \cdots, K, \tag{7}$$

$$-\rho_k y_k = 0, \quad k = 1, \cdots, K. \tag{8}$$

Thus, the following results can be obtained:

$$\pi_{js} = \frac{\partial F_{js}(D_{js})}{\partial D_{js}}, \tag{9}$$

$$\rho_k = \sum_{s=1}^{S} \sum_{j=1}^{v_s} \left( \sum_{i \in A_k} W_{ijs} \right) \pi_{js}. \tag{10}$$

From (7) and (10),

$$\sum_{s=1}^{S}\sum_{j=1}^{v_s}\left(\sum_{i\in A_k} W_{ijs}\right)\pi_{js} \geq 0, \tag{11}$$

which is equivalent to

$$\min_{k\in K}\sum_{i\in A_k}\left(\sum_{s=1}^{S}\sum_{j=1}^{v_s} W_{ijs}\pi_{js}\right) \geq 0. \tag{12}$$

When (12) is satisfied, the current solution is already the optimal solution, and the generated aperture should not be added to the plan. Therefore, in the process of generating a new aperture, the pricing of this aperture should be calculated. The pricing problem of column generation is

$$\min_{k\in K}\sum_{i\in A_k}\left(\sum_{s=1}^{S}\sum_{j=1}^{v_s} W_{ijs}\pi_{js}\right). \tag{13}$$

#### 2) NETWORK FLOW

Because the mechanical constraints of the MLC system involved in this study do not allow interdigitation, a network flow is used to solve the pricing problem, i.e., to generate a new aperture conforming to (12). As mentioned earlier, the beam is decomposed into a rectangular grid of beamlets comprising $m$ rows and $n$ columns. For a given beam $l \in B$, the beamlet row is denoted by $r(r = 1, \cdots, m)$, $c_1(c_1 = 0, \cdots, n)$ denotes the last beamlet blocked by the left leaf in row $r$, and $c_2(c_2 = 1, \cdots, n+1)$ denotes the first beamlet blocked by the right leaf in row $r$. According to (13), the price of leaf position combination $(r, c_1, c_2)$ is given by

$$\sum_{c=c_1+1}^{c_2-1}\sum_{s=1}^{S}\sum_{j=1}^{v_s} W_{(l,\,r,\,c)js}\pi_{js}. \tag{14}$$

By substituting (1), (2), and (9) into (14), the following is obtained:

$$\sum_{c=c_1+1}^{c_2-1}\sum_{s=1}^{S}\sum_{j=1}^{v_s} W_{(l,\,r,\,c)js}\pi_{js} = \sum_{c=c_1+1}^{c_2-1}\sum_{s=1}^{S}\sum_{j=1}^{v_s}\frac{\partial F_{js}(y_k)}{\partial y_k}. \tag{15}$$

Thus, the price of each leaf position combination $(r, c_1, c_2)$ is a linear sum of the gradient components of the nonoccluded beamlets in the gradient map.

A network flow is used to optimize the aperture shape. First, the current aperture shape network must be established by setting the source and root nodes. Each leaf position combination of each line represents the nodes from $(r_1, c_{11}, c_{12})$ to $(r_2, c_{21}, c_{22})$; if this combination satisfies the mechanical constraints of the MLC, a path is formed. The path length is the price of node $(r_1, c_{11}, c_{12})$. From the source node to the root node, all possible combinations of leaf positions in the aperture comprise a network; that is, all combinations that satisfy the constraints are enumerated. Then, the shortest-path algorithm is used to solve the pricing problem and obtain the deliverable aperture that can minimize the objective function and satisfy the MLC mechanical constraints of the leaf position combinations. Finally, according to the nodes in the shortest path, the leaf positions are set up, the nonobscured beamlets are opened to form a new deliverable aperture, and this aperture shape is added to the treatment plan. The enumeration of all possible combinations in the aperture involves a massive number of computations, and direct determination of the leaf position by the gradients may result in a decline in the plan quality. Therefore, the number of enumerations must be reasonably reduced in order to reduce the computational cost of column generation. In addition, beamlets with large positive gradient values should be excluded from the aperture shape, and beamlets with small negative gradient values should be included in the aperture shape as much as possible to improve the plan quality.

### C. COLUMN GENERATION APPROACH BASED ON REGION GROWTH

In this section, to improve the plan quality and reduce the computational cost, the gradients of the beamlets are first transformed by nonlinear transformation; then, the continuous negative gradient regions in the gradient map are segmented. Generally, in the pricing problem, the gradient map of a aperture is considered as an image. In image processing, gray level transformation can selectively highlight the characteristics of interest or suppress the unnecessary features in the image. In order to achieve this goal, a logarithmic function is often used in gray level transformation; that is, nonlinear gray scale transformation. Furthermore, region growth is the process of grouping pixels or regions into larger regions. Starting from the collection of seed points, region growth from these points involves incorporation into the region of adjacent pixels having similar properties for each seed point, such as intensity, gray level, texture, and color. This concept is incorporated into column generation in order to easily obtain the shape of a deliverable aperture.

#### 1) NONLINEAR TRANSFORMATION

The gradients corresponding to a row of beamlets are shown in Fig. 2(a). Direct determination of the leaf position by the gradients may cause the positive values (shown in red and green in Fig. 2(a)) to be canceled by the surrounding negative values, and the beamlets corresponding to these two positive gradients are then added to the aperture shape.

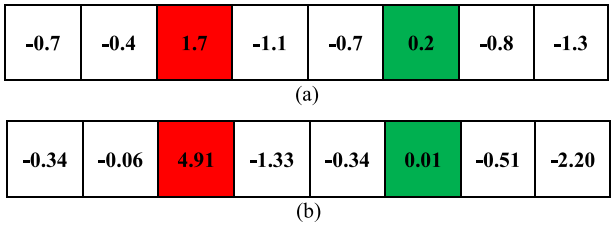


**FIGURE 2.** Nonlinear transformation of gradients: (a) gradients before transformation and (b) results after transformation.

The beamlet with the large positive gradient value marked in red is added to the aperture, which may affect the plan quality because a beamlet with a large gradient value may play a more prominent role in calculating the leaf positions than the sum of the surrounding beamlets with small gradient values. In image processing, in order to selectively enhance or suppress the gray scale, nonlinear transformation is used. Therefore, a nonlinear transformation of the gradients of the gradient map can also be performed so that the gradients with larger values are not easily canceled during the calculation of the leaf positions. The following form is generally used for nonlinear transformation:

$$\tilde{g} = \alpha \cdot g^{\beta}, \tag{16}$$

where $\alpha$ and $\beta$ are constants and can be chosen according to actual experimental conditions, $g$ is the original gradient, and $\tilde{g}$ is the gradient after nonlinear transformation. Equation (16) is only a form of nonlinear transformation. For example, let $\alpha = 1$ and $\beta = 3$ for the nonlinear transformation of the gradients shown in Fig. 2(a). The result after transformation is shown in Fig. 2(b).

According to Fig. 2(b), after nonlinear transformation, the positive gradient value marked in red increases, and the positive gradient value marked in green decreases. Thus, it is ensured by the nonlinear transformation that the larger value of the positive gradient cannot be easily canceled by the surrounding negative gradients, which greatly reduces the possibility that the corresponding beamlet is added to the aperture shape. Therefore, the quality of the plan is improved.

### 2) OBTAINING THE INITIAL SHAPE BY REGION GROWTH

After nonlinear transformation, for each row of the generated aperture, the gradients of the beamlets from left to right in the row of the corresponding gradient map can be calculated. In a row, the beamlet position of the first negative gradient is used as the seed for region growth, and all beamlets of negative gradients associated with this seed are added to the region of this seed. During the region growth process, when a beamlet with a positive gradient is encountered, the growth process of the seed is terminated. In order to start another region growth process, the search for another seed is continued with the remaining beamlets in this row. If a new seed is present, then region growth is performed using this seed. Otherwise, we move to the next step. In addition, all start and stop positions of the regions are determined by the seeds in the region growth process. It is known that the gradient map has $m$ rows and $n$ columns. The gradient values of the gradient map are stored in the array $grad\,[m]\,[n]$. The $i$th $(0 \le i < m)$ row data are taken for segmentation by region growth. Here, *count* represents the number of regions segregated by region growth. The start and end positions of the continuous negative gradient regions are stored in $Begin\,[i]\,[count]$ and $End\,[i]\,[count]$, respectively, and the sums of the negative gradients in each region are stored in $Sum\,[i]\,[count]$. The process described above is summarized in Algorithm 1.

**Algorithm 1** Segmentation by Region Growth

**Input**: grad $[i]$ $[n]$
**Output**: $Begin\,[i]\,[count]$, $End\,[i]\,[count]$, $Sum\,[i]\,[count]$, $count$

$flag \leftarrow 0$, $len \leftarrow 0$, $begin \leftarrow 0$
$end \leftarrow 0$, $count \leftarrow 0$, $sum \leftarrow 0$
For $j = 0 : n - 1$
  if $grad\,[i]\,[j] \le 0$ then
    $flag \leftarrow 1$, $len \leftarrow len + 1$
    // Find a seed and perform region growth
    if $flag = len$ then
    $begin \leftarrow j$, $end \leftarrow j$
    End
    $end \leftarrow j$, $count \leftarrow count + 1$,
    $sum \leftarrow sum + grad\,[i]\,[j]$
  else
    // End the region growth of the seed
    if $flag = 1$ then
      $Begin\,[i]\,[count] \leftarrow begin$,
      $End\,[i]\,[count] \leftarrow end$,
      $Sum\,[i]\,[count] \leftarrow sum$
      $count \leftarrow count + 1$, $flag \leftarrow 0$,
      $len \leftarrow 0$,$begin \leftarrow 0$,
      $end \leftarrow 0$, $sum \leftarrow 0$
    End
  End
next $j$
if $flag = 1$ then // End the region growth of the seed
  $Begin\,[i]\,[count] \leftarrow begin$,
  $End\,[i]\,[count] \leftarrow end$,
  $Sum\,[i]\,[count] \leftarrow sum$
  $count \leftarrow count + 1$
End

At this point, the region growth operation of the row has been completed. It is obvious that the region of negative gradients obtained by region growth will have two cases, as shown in Figs. 3(a) and 3(b). The former shows only one region, and the latter shows multiple regions. On this basis, a network flow is used to solve the pricing problem.

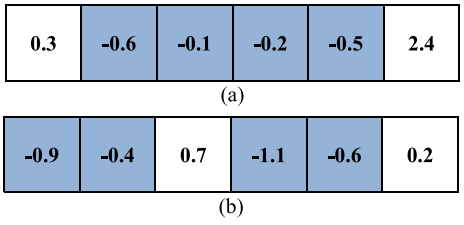


**FIGURE 3.** Results after region growth: (a) only one region and (b) multiple regions.

### 3) MERGING AND SOLVING THE PRICING PROBLEM

In this section, the $6 \times 6$ aperture after nonlinear transformation shown in Fig. 4(a) is used as an example to explain

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 0.3 | -0.6 | -0.1 | -0.2 | -0.5 | 2.4 |
| 1 | -0.2 | -0.9 | 0.2 | -0.3 | 0.4 | 0.1 |
| 2 | -0.5 | -0.1 | -3.1 | -0.5 | -7.4 | -0.6 |
| 3 | -0.9 | -0.4 | 0.7 | -1.1 | -0.6 | 0.2 |
| 4 | 0.1 | -0.6 | 0.8 | -0.2 | -0.5 | 1.3 |
| 5 | 0.4 | 0.1 | -0.5 | -0.9 | 0.1 | 0.7 |

(a)

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 0.3 | -0.6 | -0.1 | -0.2 | -0.5 | 2.4 |
| 1 | -0.2 | -0.9 | 0.2 | -0.3 | 0.4 | 0.1 |
| 2 | -0.5 | -0.1 | -3.1 | -0.5 | -7.4 | -0.6 |
| 3 | -0.9 | -0.4 | 0.7 | -1.1 | -0.6 | 0.2 |
| 4 | 0.1 | -0.6 | 0.8 | -0.2 | -0.5 | 1.3 |
| 5 | 0.4 | 0.1 | -0.5 | -0.9 | 0.1 | 0.7 |

(b)

**FIGURE 4.** Gradient maps used as an example: gradient maps (a) without any operation and (b) after region growth.

the proposed method. Region growth (Section II.C.2) is performed on the gradient map, and the processed gradient map shown in Fig. 4(b) is obtained. Next, a merging operation is performed on the processed gradient map. If there is only one continuous negative gradient region in the $i$th row, as shown in Fig. 3(a), the start and end positions of this region are $Begin\,[i]\,[0]$ and $End\,[i]\,[0]$, respectively. Then, all of the beamlets corresponding to this region are merged into a new beamlet, and the gradient value of this new beamlet is the sum of the continuous negative gradients in this segment. The position of this new beamlet in the gradient map is denoted by $(i, Begin\,[i]\,[0]\,, End\,[i]\,[0])$, as shown in Fig. 5(a). Conversely, if the $i$th row has $count$ continuously negative gradient regions, as shown in Fig. 3(b), the starting and ending positions of these regions are $Begin\,[i]\,[count]$ and $End\,[i]\,[count]$, respectively. Each region is successively merged, and finally, $count$ new beamlets are obtained. The gradients of these new beamlets are still the sums of the

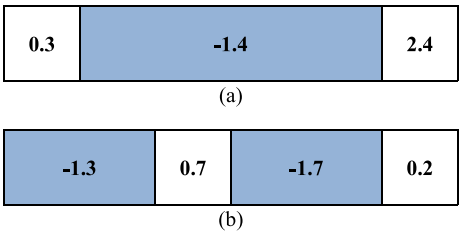


**FIGURE 5.** Performance of the merge operation: (a) only one region and (b) multiple regions.

corresponding regions, as shown in Fig. 5(b), and the position of the $j$th new beamlet in the gradient map is denoted by $(i, Begin\,[i]\,[j]\,, End\,[i]\,[j])$. After merging the aperture gradient map shown in Fig. 4(b), a new gradient map is obtained, as shown in Fig. 6.

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | 0.3 | -1.4 | | | | 2.4 |
| 1 | -1.1 | | 0.2 | -0.3 | 0.4 | 0.1 |
| 2 | -12.2 | | | | | |
| 3 | -1.3 | | 0.7 | -1.7 | | 0.2 |
| 4 | 0.1 | -0.6 | 0.8 | -0.7 | | 1.3 |
| 5 | 0.4 | 0.1 | -1.4 | | 0.1 | 0.7 |

**FIGURE 6.** Gradient map after the merging operation.

If the aperture gradient map shown in Fig. 4(b) is directly used to build the network flow, then the total number of nodes in this network is 168. However, as the aperture gradient map shown in Fig. 6 is used to build the network, the total number of nodes is 91. Thus, the number of nodes is reduced by 45.8%, which reduces the number of calculations for solving the pricing problem. In particular, a greater reduction in the number of computations is achieved when the shape of the region is simpler. The specific flow for solving the pricing problem is shown in Fig. 7.

As mentioned earlier, solving the pricing problem is a process of generating and adding new aperture to the treatment plan such that the improvement in the objective function is maximized. For the aperture to improve the objective function to the greatest extent, the maximum possible corresponding rays of beamlets with negative gradients need to be opened and the minimum possible beamlets with positive gradients to be introduced. If the region growth operation is performed on positive gradients in the gradient map, the resulting region will contain at least one positive gradient. After merging all

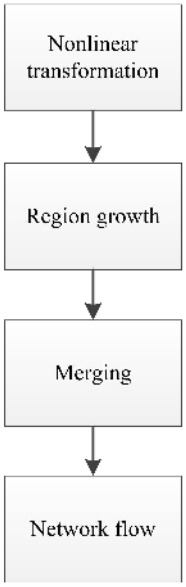


**FIGURE 7.** Specific flow for the solving the pricing problem.

positive gradients in a region, a new beamlet is no longer a single beamlet in the original gradient map. Then, when the merged gradient map is used to solve the pricing problem, if the beamlet with the positive gradient is introduced into the aperture, the introduced beamlet is equivalent to multiple beamlets with positive gradients of the corresponding region in the original gradient map. Thus, the quality of the plan will be affected. Therefore, region growth and merging operations are performed only on the negative gradients in a gradient map.

## III. EXPERIMENTS AND RESULTS

In this section, five cases of head and neck cancer and five cases of prostate cancer are considered to verify the effectiveness and feasibility of the proposed method. Two column generation algorithms are compared. The proposed column generation algorithm, which is based on region growth, is denoted by "RG." The generic column generation algorithm is denoted by "Original." The L-BFGS-B algorithm is used by both column generation algorithms to solve the restricted master problem. The maximum number of apertures in the prostate cancer cases is 60, and the maximum number of apertures in the head and neck cancer cases is 100.

In this study, the dose deposition matrix $W$ was calculated using the classical pencil beam algorithm [34] in the Computational Environment for Radiological Research (CERR) open-source software [35]. The total function was obtained by a weighted linear combination of the subobjective functions [36]:

$$f\left(D\left(x\right)\right) = \sum_{l=1}^{L} \xi_l f_l\left(D\left(x\right)\right), \tag{17}$$

where the dose distribution $D\left(x\right) = Wx$; that is, the dose distribution is a linear function of $W$ and the fluence matrix $x$. Further, $f_l\left(D\left(x\right)\right)$ is the $l$th subobjective function, $\xi_l$ is the weight coefficient representing the importance of the corresponding subobjective function, and $L$ is the number of subobjective functions.

The proposed method and generic column generation were implemented in VC++ (v. VS2012) on a computer equipped with an Intel® Core™i7-6700K CPU (3.40 GHz) running Windows 7 (64 bit).

### A. STUDY OF HEAD AND NECK CANCER CASES

The effectiveness and feasibility of the proposed method were verified with cases of head and neck cancer. Nine 6-MeV coirradiated photon fields were used to irradiate the target, and the radioactive sources were spaced 40° apart. For the head and neck cancer cases (as shown in Fig. 8), the parotids, spinal cord, and brain stem were selected as the organs at risk (OARs) according to [37]. Further, according to the guidelines for clinical practice [38], for head and neck cancer cases, the dose distributions of the spinal cord and brain stem were restricted by the maximum dose subobjective function, which required that the maximum doses for the spinal cord and brain stem should not exceed 50 and 54 Gy (the gray (Gy) is a derived unit in the International System of Units, which represents a standard unit of energy absorption dose of ionizing radiation), respectively. In addition, the dose-volume histogram (DVH) subobjective function [42] was used to constrain the dose for the parotid gland, requiring that the mean dose does not exceed 25 Gy. On this basis, three planning target volumes (PTVs) were considered (PTV 70 Gy, PTV 63 Gy, and PTV 56 Gy), and the dose was restricted using the minimum dose and mean dose subobjective functions. The "Tissue" organization was the tissue in the patient, except for the PTVs and OARs. The dose restriction for "Tissue" was controlled by the maximum dose subobjective function. The total volumes (cubic centimeters) of the organs for five cases of head and neck cancer are listed in Table 1.

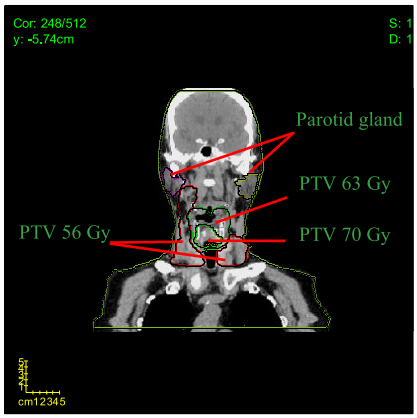


**FIGURE 8.** Structural distribution of various organs in the head and neck cancer case.

#### 1) EVALUATION CRITERIA

To evaluate the quality of the plan, the following criteria are considered. First, dose restriction and plan evaluation for the

**TABLE 1.** Total volumes (cubic centimeters) of organs for five cases of head and neck cancer.

| Case | Ipsilateral parotid gland | Contralateral parotid gland | Spinal cord | Brain stem | PTV 70 Gy | PTV 63 Gy | PTV 56 Gy | Tissue |
|---|---|---|---|---|---|---|---|---|
| 1 | 21.5742 | 29.9449 | 36.2685 | 33.6606 | 30.7134 | 105.7684 | 245.3428 | 1144.5943 |
| 2 | 18.2229 | 18.4386 | 45.3315 | 20.6956 | 90.0186 | 546.7944 | 38.2002 | 1649.2207 |
| 3 | 29.7185 | 28.223 | 11.9236 | 28.223 | 93.5865 | 300.8542 | 278.3335 | 3326.4023 |
| 4 | 12.1508 | 11.5535 | 9.3063 | 24.9865 | 109.5021 | 246.3392 | 113.342 | 2264.2493 |
| 5 | 14.5457 | 12.0295 | 44.1714 | 22.7259 | 265.1423 | 274.5555 | 326.4256 | 3062.6959 |

OARs and targets were performed according to [38] and [39] (see Table 2). These criteria are mainly used to evaluate the DVH of the result. $V_{xGy}$ is the volume of the organ receiving $\geq$ x Gy. $D_{max}$ is the maximum radiation dose, and the $D_{mean}$ is the mean radiation dose.

**TABLE 2.** Dose-Volume (DV) constraint conditions of organs.

| Structure | Dose-volume parameter | |
|---|---|---|
| Parotid gland | $D_{mean} \leq 25$ Gy | |
| Spinal cord | $D_{max} \leq 50$ Gy | |
| Brain stem | $D_{max} \leq 54$ Gy | |
| PTV 70 Gy | $V_{70\,Gy} > 95\%$ | $V_{77\,Gy} < 5\%$ |
| PTV 63 Gy | $V_{63\,Gy} > 95\%$ | $V_{70\,Gy} < 5\%$ |
| PTV 56 Gy | $V_{56\,Gy} > 95\%$ | $V_{62\,Gy} < 5\%$ |

Second, the generalized equivalent uniform dose (gEUD) and normal tissue complication probability (NTCP) were calculated to evaluate the performance of the algorithm for protecting the OARs. Since the evaluation standard of biological criteria such as the gEUD and NTCP have not been established internationally, we can only require that a smaller gEUD and NTCP results in better protection of the OARs. The conformity number (CN) proposed by van't Riet *et al.* [40] was used to measure the dose conformity of the target:

$$CN = \frac{TV_{ri}}{TV} \times \frac{TV_{ri}}{V_{ri}}, \tag{18}$$

where $TV$ represents the total volume of the target, $TV_{ri}$ represents the target volume enclosed within the 95% isodose line, and $V_{ri}$ represents the total volume of tissues enclosed within the 95% isodose line. In addition, the homogeneity index (HI) [41] was used to evaluate the target dose homogeneity:

$$HI = \frac{D_{5\%}}{D_{95\%}}, \tag{19}$$

where $D_{5\%}$ and $D_{95\%}$ denote the radiation doses of 5% and 95% of the PTV volume, respectively. Better conformability is obtained when the CN is closer to 1, and better uniformity is obtained when the HI is closer to 1. Then, the dose distribution of the plan was also evaluated. Finally, the running time of the plan and the number of apertures were considered to evaluate the effectiveness and feasibility of the method.

### 2) RESULTS OF THE HEAD AND NECK CANCER CASES

In the analysis of the optimization results of head and neck cancer cases, the dose distribution for the parotid gland was constrained by the DVH subobjective function, and the evaluation standard required that the mean dose for the parotid gland should not exceed 25 Gy. The mean dose should be an evaluation of the overall dose, and the mean dose for the parotid gland should be reduced as much as possible. However, because the maximum dose subobjective function was used to restrict the dose distribution of the spinal cord and brain stem, and the maximum dose subobjective function was used to penalize the dose exceeding the threshold. The thresholds of the maximum dose subobjective function of the spinal cord and brain stem were set at 50 and 54 Gy, respectively; thus, when the maximum dose for the spinal cord and brain stem was not beyond the corresponding threshold, the maximum dose subobjective function has no penalty effect. That is, the maximum doses for the spinal cord and brain stem were required to not exceed the thresholds, and there was no penalty when the maximum dose did not exceed the threshold. With this premise, the optimization results of the five head and neck cancer cases were analyzed.

The detailed optimization results of one head and neck cancer case are as follows. When analyzing the performance of the two methods, the curves of the target in the DVH obtained by the two methods should be consistent as much as possible. Of course, the curves of the targets obtained by the new method can be slightly better than the comparison method. On this basis, the curves of the OARs in the DVH were observed. Although it is better when the curves of the OARs in the DVH is lower, the optimized results of the OARs should also be analyzed in combination with clinical guidelines [38], [39]. As shown in Fig. 9(b), the curves of the three targets obtained by "RG" are slightly better than those obtained by "Original." On this basis, the curves of the OARs in the DVH are analyzed. For the ipsilateral parotid gland, although the curve of "RG" is slightly higher in the high-dose part than that of "Original," it is much lower than the "Original" result in the low-dose part. Similarly, for the contralateral parotid gland, the curve of "RG" is slightly higher in the low-dose part than the results obtained by "Original" but significantly lower in the high-dose part than that of "Original." On this basis, according to the dose-volume (DV) constraints in Table 2, the mean dose is used to evaluate the dose distribution on the parotid gland, and the DV constraints are satisfied by the mean dose of the

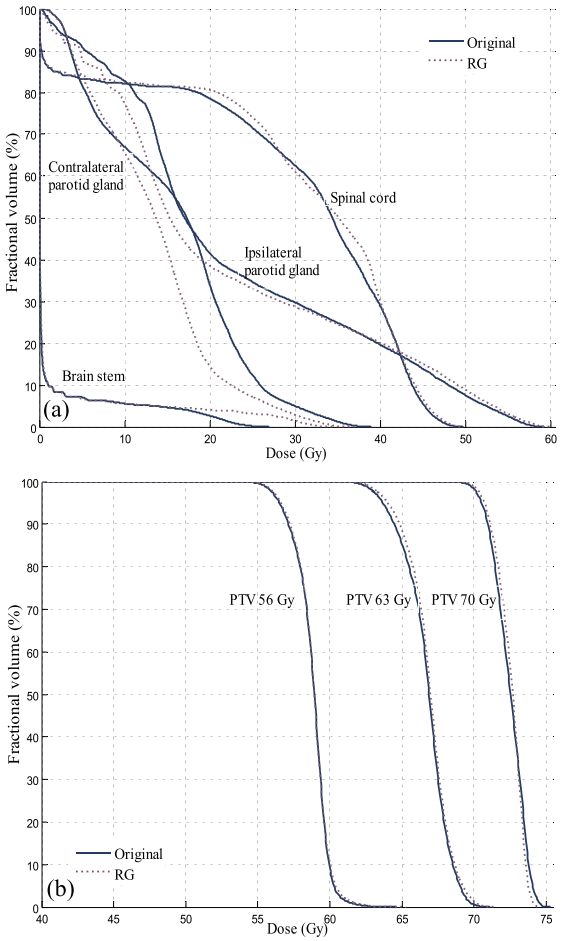


**FIGURE 9.** Contrast optimization results of generic column generation and column generation based on region growth: DVH for the (a) OARs and (b) targets.

two methods of the parotid gland. Furthermore, according to the data in Table 3, compared with ''Original,'' the mean dose of the ipsilateral parotid gland obtained by ''RG'' decreased by 1.01 Gy, and the mean dose of the contralateral parotid gland obtained by ''RG'' decreased by 2.16 Gy. Although the

**TABLE 3.** Optimized information of case 1.

| Case 1 | | Original | RG |
|---|---|---|---|
| Ipsilateral parotid gland | Mean dose | 22.96 Gy | 21.95 Gy |
| | gEUD | 22.96 Gy | 21.95 Gy |
| | NTCP | 33.37% | 31.13% |
| Contralateral parotid gland | Mean dose | 15.56 Gy | 13.40 Gy |
| | gEUD | 15.56 Gy | 13.40 Gy |
| | NTCP | 18.72% | 15.35% |
| Spinal cord | Max dose | 49.73 Gy | 49.68 Gy |
| Brain stem | Max dose | 26.93 Gy | 34.63 Gy |
| PTV 70 Gy | HI | 1.0501 | 1.0431 |
| | CN | 0.7824 | 0.9406 |
| PTV 63 Gy | HI | 1.0838 | 1.0810 |
| | CN | 0.2791 | 0.2109 |
| PTV 56 Gy | HI | 1.0667 | 1.0670 |
| | CN | 0.0021 | 0.0013 |

curves of the spinal cord in Fig. 9(a) are very close or even cross each other and the curve of the brain stem obtained by ''RG'' is significantly higher than that obtained by ''Original,'' the maximum dose is used to evaluate the dose distribution for the spinal cord and brain stem according to Table 2. Therefore, the requirements in Table 2 are satisfied by the maximum doses for the spinal cord and brain stem obtained by the two methods in Table 3. As shown in Fig. 10, the dose distribution obtained by the proposed method, especially the dose distribution of the contralateral parotid gland, is significantly better than that obtained by generic column generation. The other optimization information of these two methods is summarized in Table 3. According to the data of the CN and HI for the targets, the dose distributions for the targets obtained by the two methods are essentially identical; even the dose distributions of ''RG'' for the targets are slightly better. Furthermore, the gEUD of the ipsilateral parotid gland decreased by 1.01 Gy, and the NTCP decreased by 2.24%. The gEUD of the contralateral parotid gland decreased by 2.16 Gy, and the NTCP decreased by 3.37%.

The optimization results of the four other head and neck cancer cases are presented in Table 4 and Fig. 11. It can be seen from Table 4 that in the second case, although the

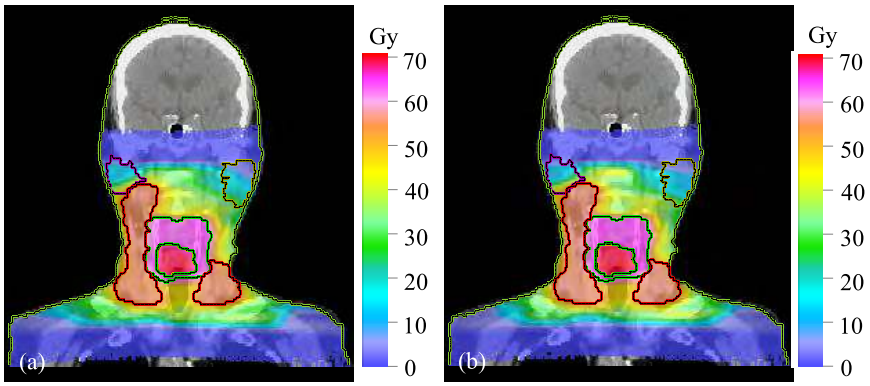


**FIGURE 10.** Dose distribution maps of the two methods for the head and neck cancer case: dose distribution maps obtained by (a) generic column generation and (b) column generation based on region growth.

**TABLE 4. Optimized information of the four other head and neck cancer cases.**

| Case | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Original | RG | Original | RG | Original | RG | Original | RG |
| Ipsilateral parotid gland | Mean dose | 25.91 Gy | 26.66 Gy | 32.98 Gy | 32.17 Gy | 41.69 Gy | 41.36 Gy | 44.72 Gy | 44.23 Gy |
| | gEUD | 25.91 Gy | 26.66 Gy | 32.98 Gy | 32.17 Gy | 41.69 Gy | 41.36 Gy | 44.72 Gy | 44.23 Gy |
| | NTCP | 40.25% | 42.05% | 57.56% | 55.58% | 76.74% | 76.12% | 78.78% | 78.08% |
| Contralateral parotid gland | Mean dose | 22.37 Gy | 22.35 Gy | 30.34 Gy | 29.21 Gy | 36.71 Gy | 36.30 Gy | 30.72 Gy | 30.78 Gy |
| | gEUD | 22.37 Gy | 22.35 Gy | 30.34 Gy | 29.21 Gy | 36.71 Gy | 36.30 Gy | 30.72 Gy | 30.78 Gy |
| | NTCP | 32.04% | 32.01% | 51.08% | 48.29% | 66.34% | 65.42% | 52.38% | 51.98% |
| Spinal cord | Max dose | 50.28 Gy | 50.03 Gy | 49.68 Gy | 49.38 Gy | 48.08 Gy | 49.88 Gy | 47.83 Gy | 47.33 Gy |
| Brain stem | Max dose | 49.23 Gy | 50.18 Gy | 50.23 Gy | 51.58 Gy | 49.18 Gy | 49.23 Gy | 31.38 Gy | 40.48 Gy |
| PTV 70 Gy | HI | 1.0414 | 1.0424 | 1.0507 | 1.0445 | 1.0574 | 1.0596 | 1.0543 | 1.0550 |
| | CN | 0.8996 | 0.7978 | 0.4055 | 0.4458 | 0.2733 | 0.1489 | 0.1709 | 0.2139 |
| PTV 63 Gy | HI | 1.0914 | 1.0983 | 1.0985 | 1.1002 | 1.1014 | 1.1007 | 1.0890 | 1.0956 |
| | CN | 0.0046 | 0.0073 | 0.0021 | 0.0006 | 0.0002 | $5.8\times10^{-5}$ | $2.9\times10^{-5}$ | 0.0017 |
| PTV 56 Gy | HI | 1.0920 | 1.0907 | 1.0975 | 1.1064 | 1.1015 | 1.1052 | 1.0842 | 1.0886 |
| | CN | $6.2\times10^{-5}$ | $2.3\times10^{-5}$ | $1.0\times10^{-7}$ | $4.6\times10^{-7}$ | $1.3\times10^{-5}$ | $2.7\times10^{-6}$ | $1.5\times10^{-7}$ | $1.9\times10^{-7}$ |

maximum dose of the spinal cord exceeds the threshold of 50 Gy, it is penalized by the maximum dose criterion. Nevertheless, the maximum dose obtained by the proposed method is less than that obtained by generic column generation. In addition, the maximum dose for the rest of the spinal cord and brain stem did not exceed the thresholds, and the maximum dose criterion has no penalty effect. Furthermore, according to the information related to the mean dose, the gEUD and NTCP for the parotid gland in Table 4, the parotid gland is better protected with the proposed method than generic column generation. However, it is noted that the requirements in some cases in Table 4 are not satisfied by the mean dose for the parotid gland because there is too much overlap between the targets and the OARs in these cases, as shown in the dose maps in Fig. 11. According to the CN and HI, there are no obvious differences between two methods.

On the basis of previous results, the running time and the number of apertures used to optimize the five cases for the two methods were studied (see Table 5). From Table 5, compared with generic column generation, the proposed method has a shorter running time and requires fewer apertures for optimization.

**TABLE 5. Running times of the plan and the numbers of apertures for the five head and neck cancer cases.**

| | Case | Running time | Number of apertures | Acceleration rate |
|---|---|---|---|---|
| 1 | Original | 819.83 s | 88 | 13.58% |
| | RG | 708.52 s | 90 | |
| 2 | Original | 768.61 s | 85 | 10.95% |
| | RG | 684.45 s | 82 | |
| 3 | Original | 763.70 s | 84 | 20.44% |
| | RG | 607.62 s | 79 | |
| 4 | Original | 799.12 s | 82 | 18.08% |
| | RG | 654.65 s | 79 | |
| 5 | Original | 833.19 s | 89 | 12.25% |
| | RG | 731.14 s | 88 | |

### B. STUDY OF PROSTATE CANCER CASES

The effectiveness and feasibility of the proposed method were verified for five cases of prostate cancer. First, the targets and OARs in a patient's computed tomography (CT) data were selected, as shown in Fig. 12. The main target was the prostate without regard for the pelvic lymph nodes. The PTV was expanded by 5 mm in the backward direction and by 10 mm in the other direction on the basis of the clinical target volume (CTV). The bladder and rectum were chosen as the OARs. The rectal and bladder walls were obtained by manual delineation, and the rectal and bladder contours were expanded by 5 mm. The ''Tissue'' organization was the tissue in the patient, except for the PTV expansion beyond 5 cm. The total volumes (cubic centimeters) of the organs for the five prostate cancer cases are listed in Table 6. Five 6-MeV coirradiated photon fields were used to irradiate the target. The angles of the frame were 36°, 100°, 180°, 260°, and 324°. The prescribed dose of the target was 78 Gy, and the prescription of complications of the bladder and rectum was set to zero to reduce the dose received by the normal tissues as much as possible. In the experiment, the objective functions for the two methods were the same as that of normal tissue via the subobjective function proposed by Wu and Mohan [42]. The dose distribution of the bladder was controlled by two DVH equivalent convex rules, and four weighted DVH subobjective functions were applied to the rectum. The target had one minimum dose objective subfunction and one mean dose subobjective function to control the dose distribution. Thus, the total objective function of the experiment was obtained.

**TABLE 6. Total volumes (cubic centimeters) of the organs for five cases of prostate cancer.**

| Case | Bladder | Rectum | PTV | Tissue |
|---|---|---|---|---|
| 1 | 141.1057 | 27.8146 | 186.8307 | 21893.9722 |
| 2 | 53.0338 | 36.2549 | 197.2458 | 14822.9324 |
| 3 | 92.98 | 20.7559 | 199.7946 | 13601.7499 |
| 4 | 77.7914 | 30.6134 | 201.6843 | 16312.3434 |
| 5 | 87.7917 | 29.3362 | 103.7247 | 26032.305 |

#### 1) EVALUATION CRITERIA

To evaluate the quality of the plan, the clinical practice guidelines proposed by Marks *et al.* [38] were first considered to

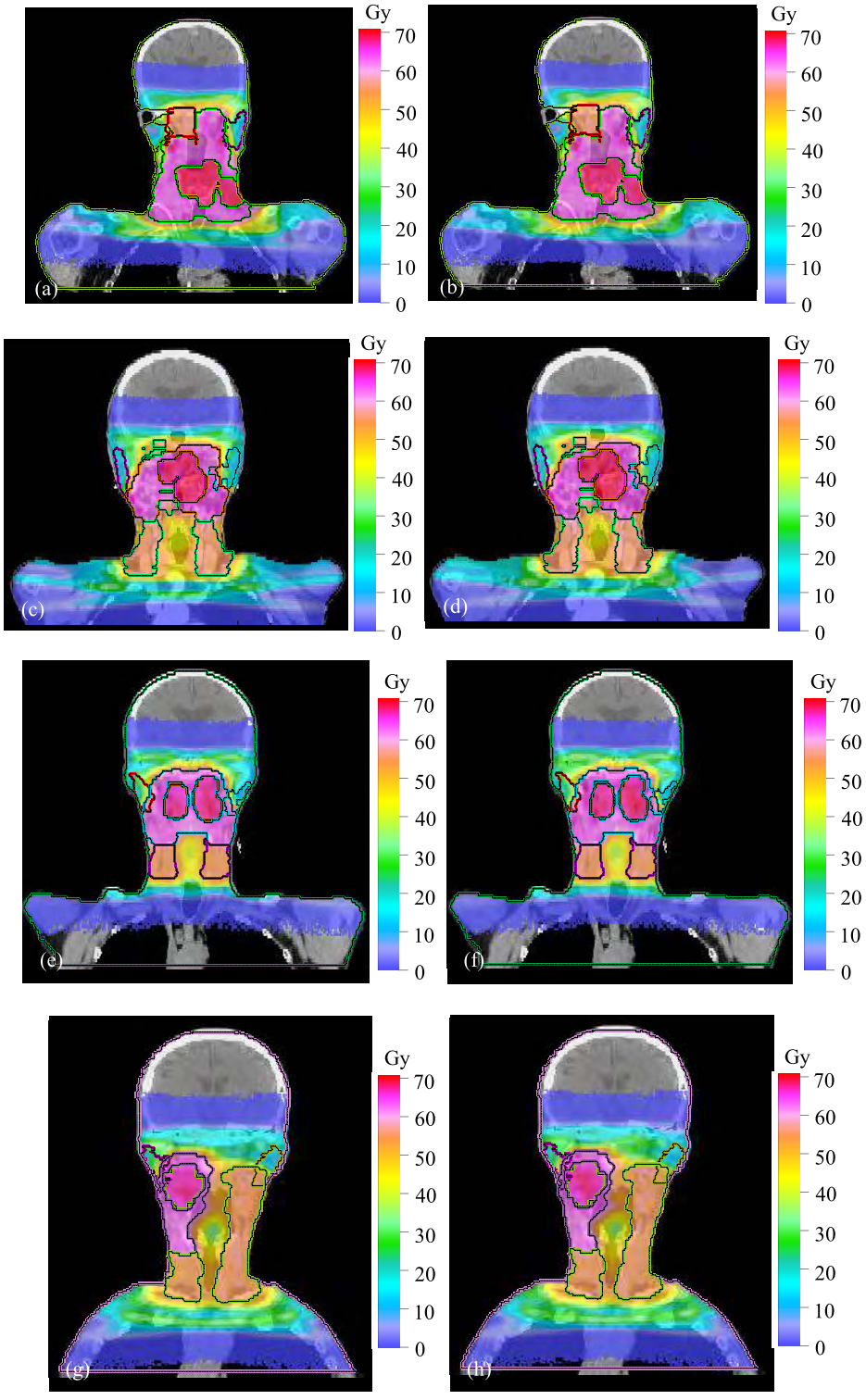


**FIGURE 11.** Comparison of the dose distributions of the two methods for the four other head and neck cancer cases. (a), (c), (e), and (g) show the dose distribution maps obtained by generic column generation for the four cases. (b), (d), (f), and (h) show the corresponding dose distribution maps obtained by column generation based on region growth for the four cases.

evaluate the DVH results of the optimized plan (see Table 7). Then, as with the evaluation of the head and neck cancer cases, the dose distribution, the HI, the CN, the gEUD, the NTCP, the running time, and the number of apertures were used to evaluate the effectiveness and feasibility of the proposed method.

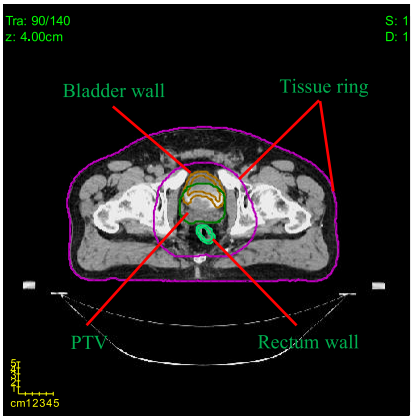


**FIGURE 12.** Structural distribution of various organs in a patient for the cases of prostate cancer.

**TABLE 7.** DV constraint conditions of the organs.

| | Bladder | Rectum |
|---|---|---|
| Dose-volume parameter | | $V_{50\,Gy} < 50\%$ |
| | | $V_{60\,Gy} < 35\%$ |
| | $V_{65\,Gy} < 50\%$ | $V_{65\,Gy} < 25\%$ |
| | $V_{70\,Gy} < 35\%$ | $V_{70\,Gy} < 20\%$ |
| | $V_{75\,Gy} < 25\%$ | $V_{75\,Gy} < 15\%$ |
| | $V_{80\,Gy} < 15\%$ | |

### 2) RESULTS OF THE PROSTATE CANCER CASES

First, a prostate cancer case is analyzed in detail. As shown in Fig. 13, the curves of the target basically coincide. In addition, the coverage of the prescribed dose (74 Gy) in the target obtained by both methods reached more than 95%. On this basis, the DV constraints in Table 7 are satisfied by all of the curves of the OARs obtained by the two methods. For the rectum, although the partial low-dose distribution (< 50 Gy) is increased by the proposed method, the effect of the low-dose distribution on the NTCP from radiotherapy is limited. In the clinical practice guidance given by Marks *et al.* [38], the DV evaluation standard with a dose lower than 50 Gy is not given. Although the curve obtained by the proposed method at 50 Gy is slightly higher than that of generic column generation, it still satisfies the DV constraints in Table 7. On this basis, since the curve in the DVH obtained by the proposed method is significantly lower than that obtained by generic column generation in the high-dose part of the rectum (> 56 Gy), the proposed method is more conducive for the protection of the rectum. For the bladder, in the clinical practice guidance given by Marks *et al.* [38], the DV evaluation standard with a dose lower than 65 Gy is not given. Therefore, it is evident from Fig. 13 that the curves of the OARs marked as "RG" are significantly lower compared to those of "Original." Furthermore, the dose distribution obtained by the proposed method is slightly better than that obtained by generic column generation, as shown in Fig. 14.

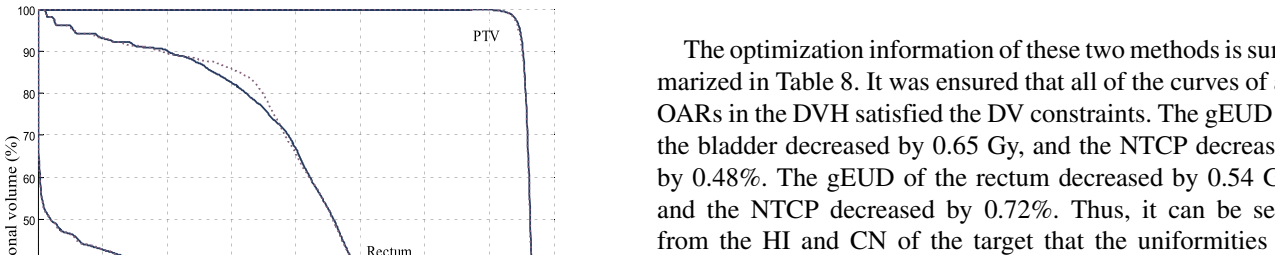


**FIGURE 13.** Contrast optimization results of generic column generation and column generation based on region growth.

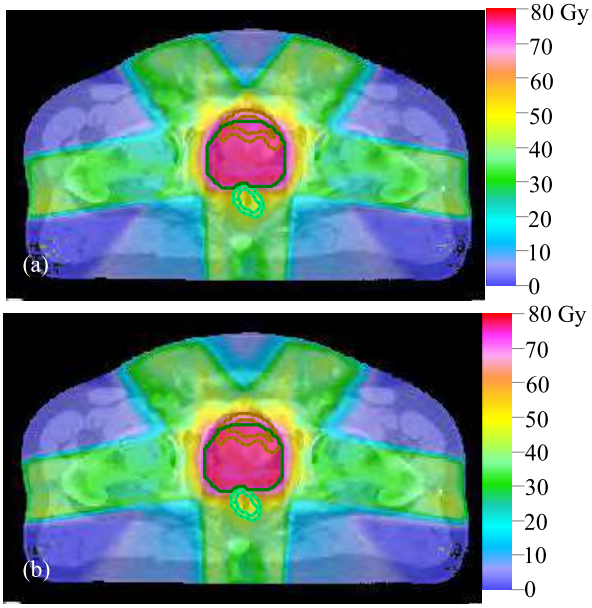


**FIGURE 14.** Dose distribution maps of the two methods for the prostate cancer cases obtained by (a) generic column generation and (b) column generation based on region growth.

The optimization information of these two methods is summarized in Table 8. It was ensured that all of the curves of all OARs in the DVH satisfied the DV constraints. The gEUD of the bladder decreased by 0.65 Gy, and the NTCP decreased by 0.48%. The gEUD of the rectum decreased by 0.54 Gy, and the NTCP decreased by 0.72%. Thus, it can be seen from the HI and CN of the target that the uniformities of

**TABLE 8.** Optimized information of prostate cancer case 1.

| | | Original | RG |
|---|---|---|---|
| Bladder | gEUD | 56.99 Gy | 56.34 Gy |
| | NTCP | 5.06% | 4.58% |
| Rectum | gEUD | 62.68 Gy | 62.14 Gy |
| | NTCP | 8.04% | 7.32% |
| PTV | HI | 1.0369 | 1.0390 |
| | CN | 0.7001 | 0.8159 |

**TABLE 9.** Optimized information of the four other prostate cancer cases.

| Case | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Original | RG | Original | RG | Original | RG | Original | RG |
| Bladder | gEUD | 66.75 Gy | 66.48 Gy | 61.36 Gy | 60.97 Gy | 72.21 Gy | 71.77 Gy | 58.95 Gy | 58.78 Gy |
| | NTCP | 18.09% | 17.55% | 9.45% | 8.97% | 25.44% | 24.34% | 6.77% | 6.61% |
| Rectum | gEUD | 61.48 Gy | 61.38 Gy | 64.92 Gy | 64.50 Gy | 67.48 Gy | 66.25 Gy | 61.27 Gy | 61.07 Gy |
| | NTCP | 6.51% | 6.40% | 11.56% | 10.82% | 15.61% | 14.09% | 6.28% | 6.05% |
| PTV | HI | 1.0294 | 1.0302 | 1.0474 | 1.0495 | 1.0326 | 1.0348 | 1.0213 | 1.0214 |
| | CN | 0.8907 | 0.9007 | 0.2860 | 0.3032 | 0.8579 | 0.8461 | 0.9043 | 0.9062 |

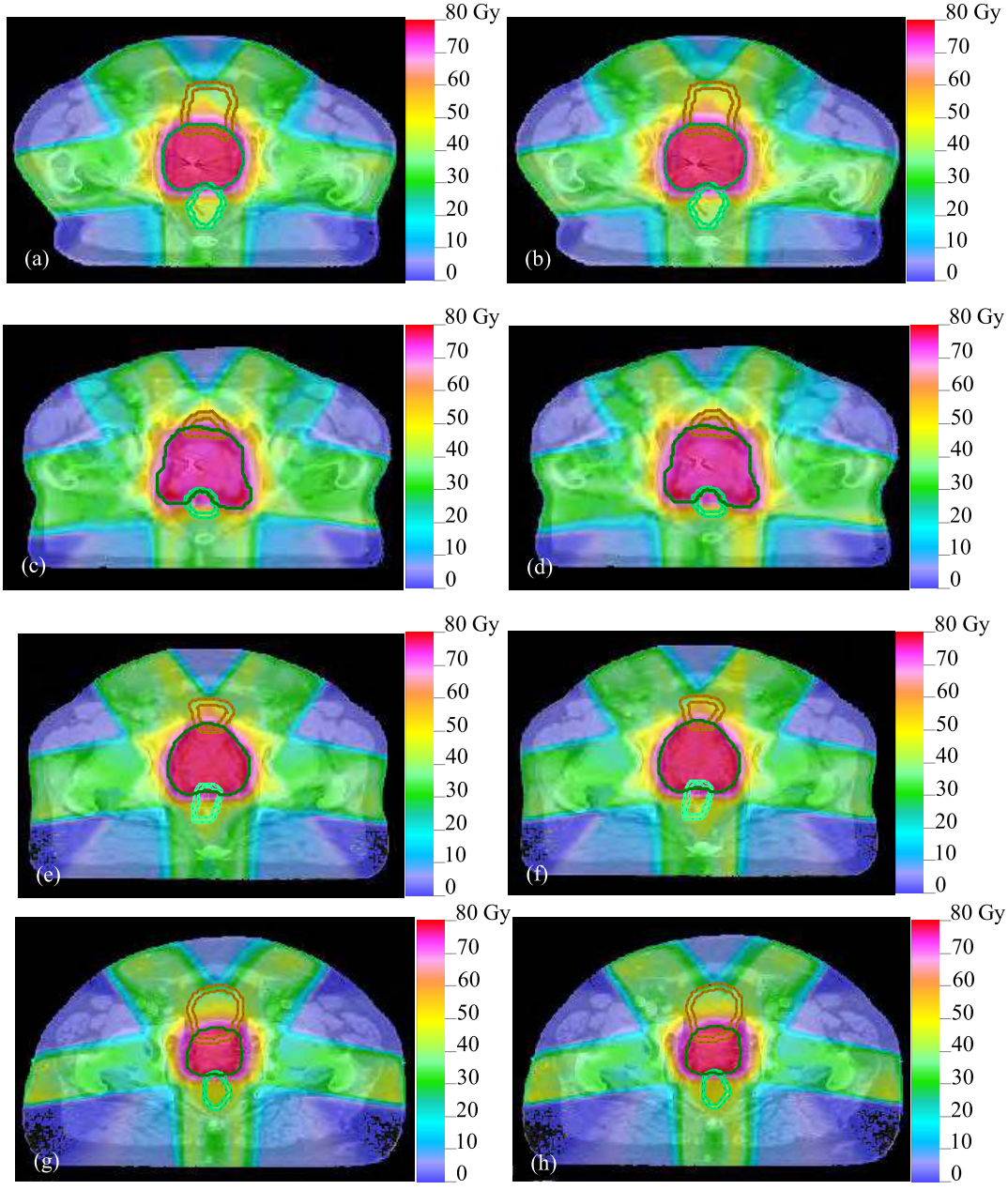


**FIGURE 15.** Comparison of the dose distributions of the two methods in the four other prostate cancer cases. (a), (c), (e), and (g) show the dose distribution maps obtained by generic column generation for the four cases. (b), (d), (f), and (h) show the corresponding dose distribution maps obtained by column generation based on region growth for the four cases.

the target for the two methods do not significantly differ, but the conformability of the target with the proposed method is significantly better than that with generic column generation.

The optimization results of the four other prostate cancer cases are presented in Table 9 and Fig. 15. According to Table 9, the gEUDs and NTCPs of the OARs obtained by the

proposed method in these four cases decreased to different values compared with those obtained by generic column generation. Since an evaluation standard for biological criteria such as the gEUD and NTCP has not been established internationally, we can only require that their values be as small as possible. Furthermore, the CN of the target obtained by the proposed method is better than that obtained by generic column generation. This conclusion can also be drawn from the results in Fig. 15.

The optimization time and the number of apertures of the two methods used to optimize the five prostate cancer cases are summarized in Table 10. From Table 10, the numbers of apertures used by the two methods are nearly the same, but the running time of the proposed method is less than that of generic column generation.

**TABLE 10. Running times of the plan and the numbers of apertures of the five prostate cancer cases.**

| | Case | Running time | Number of apertures | Acceleration rate |
|---|---|---|---|---|
| 1 | Original | 664.06 s | 57 | 4.66% |
| | RG | 633.11 s | 56 | |
| 2 | Original | 558.01 s | 55 | 6.72% |
| | RG | 520.52 s | 57 | |
| 3 | Original | 535.93 s | 57 | 7.70% |
| | RG | 494.69 s | 57 | |
| 4 | Original | 625.83 s | 58 | 2.55% |
| | RG | 609.86 s | 56 | |
| 5 | Original | 739.98 s | 58 | 2.00% |
| | RG | 725.19 s | 57 | |

### C. STATISTICAL ANALYSIS OF THE EXPERIMENTAL RESULTS

To compare the results of all of the above experiments, the optimization results of all ten cases will be statistically analyzed.

As shown in Table 11, with 10 cases, there are no significant differences in the dose distributions between ''Original'' and ''RG'' according to the P-values. On this basis, for the head and neck cancer cases, the running times for ''RG'' and ''Original'' are significantly different ($P = 0.002$). On the basis of no statistical difference in the quality of the plan, it is shown that the proposed method can effectively reduce the running time compared with the generic column generation. However, for prostate cancer cases, there is no significant difference in the running time between the two methods ($P = 0.626$). There are two reasons for this conclusion. First, the number of apertures required for prostate cancer cases is small, which is not sufficient to reflect a significant difference. Second, as shown in Figs. 14 and 15, with the five cases of prostate cancer selected in this study, the overlap between the OARs and the target is large; thus, the negative gradients in the aperture gradient map are more dispersed. This will affect the acceleration rate of the proposed method. Therefore, the acceleration rate of the proposed algorithm in prostate cancer cases is not as obvious as that in head and neck cancer cases.

## IV. DISCUSSION

Obviously, the relationship between the leaf positions and the gradients of the beamlets is not a simple linear relationship. The generic column generation algorithm directly determines the leaf positions by the gradients of the beamlets. The cancellation of positive gradients with large values, which play an important role in the process of calculating the leaf positions, by the surrounding negative values can easily occur. Thus, if beamlets with these positive gradient values are grouped into the new generated aperture shape, a decline in the plan quality is likely. Moreover, it is well-known that the aperture shape is determined by enumerating all possible combinations with the generic column generation method, which results in excessive computation. Column generation method needs to be studied and improved to reduce the number of computations and improve the plan quality. Therefore, column generation based on region growth is proposed in this study. A reduction in the number of computations and an improvement in the plan quality can be achieved by the proposed method compared to the generic column generation.

Furthermore, the number of calculations reduced by the proposed method depends on the number of negative gradients and the degree of dispersion in the distribution of negative gradients in the aperture gradient map. A larger number of negative gradients and a smaller degree of dispersion in the distribution of negative gradients will result in a larger reduction in the number of computations. For the two types of cases used in this study, the average acceleration rates are 4.72% and 15.06%, respectively, and the maximum acceleration rate is 20.44%.

Another advantage of the proposed method is that the OARs can be effectively protected while ensuring the conformability and uniformity of the target compared with generic column generation. This conclusion can be drawn from the figures and tables presented in Section III. Thus, the feasibility and applicability of the proposed method have been verified.

In this study, the nonlinear transformation results were obtained by adjusting $\alpha$ and $\beta$ according to actual experimental conditions. These parameters were manually obtained in a trial-and-error manner. For the same type of cancer case, the optimization results are not sensitive to these parameters, and they only need to be determined once for one type of cancer case [27]. $\alpha = 1$ and $\beta = 3$ for the nonlinear transformation of the gradients for 10 cases in this paper.

In particular, the effectiveness and feasibility of the proposed method were verified for head and neck cancer cases as well as prostate cancer cases. The following conclusion can be drawn: compared with generic column generation, the number of computations can be reduced, the conformability of the target can be ensured, and the OARs can be effectively protected when using the proposed method. In the future, additional image processing concepts will be introduced to column generation to further reduce the number of computations and improve the plan quality, thus improving the algorithm performance.

**TABLE 11. Statistical analysis of the optimization results.**

| Parameter | | Original | RG | *P*-value |
|---|---|---|---|---|
| Head and neck cancer cases | | | | |
| PTV 70 Gy | $V_{70\,Gy}$ (%) | 97.13 ± 1.77 | 97.22 ± 2.30 | 0.953 |
| | $V_{77\,Gy}$ (%) | 0 ± 0 | 0 ± 0 | 1.000 |
| | HI | 1.05 ± 0.01 | 1.05 ± 0.01 | 0.687 |
| | CN | 0.51 ± 0.29 | 0.51 ± 0.31 | 0.989 |
| PTV 63 Gy | $V_{63\,Gy}$ (%) | 96.08 ± 0.98 | 96.07 ± 1.19 | 0.985 |
| | $V_{70\,Gy}$ (%) | 1.84 ± 0.85 | 2.30 ± 0.69 | 0.418 |
| | HI | 1.09 ± 0.01 | 1.09 ± 0.01 | 0.642 |
| | CN | 0.057 ± 0.11 | 0.044 ± 0.08 | 0.856 |
| PTV 56 Gy | $V_{56\,Gy}$ (%) | 96.46 ± 0.89 | 96.48 ± 1.06 | 0.977 |
| | $V_{62\,Gy}$ (%) | 3.21 ± 2.00 | 3.67 ± 2.30 | 0.769 |
| | HI | 1.09 ± 0.01 | 1.09 ± 0.01 | 0.743 |
| | CN | 0.0004 ± 0.002 | 0.0003 ± 0.001 | 0.739 |
| Ipsilateral parotid gland | Mean dose (Gy) | 33.65 ± 8.51 | 33.27 ± 8.47 | 0.951 |
| | gEUD (Gy) | 33.65 ± 8.51 | 33.27 ± 8.47 | 0.951 |
| | NTCP | 0.57 ± 0.19 | 0.56 ± 0.18 | 0.956 |
| Contralateral parotid gland | Mean dose (Gy) | 27.14 ± 7.37 | 26.40 ± 7.88 | 0.895 |
| | gEUD (Gy) | 27.14 ± 7.37 | 26.40 ± 7.88 | 0.895 |
| | NTCP | 0.44 ± 0.17 | 0.42 ± 0.17 | 0.904 |
| Spinal cord | Max dose (Gy) | 49.12 ± 0.98 | 49.26 ± 0.99 | 0.845 |
| Brain stem | Max dose (Gy) | 41.39 ± 10.10 | 45.22 ± 6.57 | 0.545 |
| Prostate cancer cases | | | | |
| PTV | $V_{67.27\,Gy}$ (%) | 99.99 ± 0.02 | 99.99 ± 0.03 | 1.000 |
| | $V_{74\,Gy}$ (%) | 98.39 ± 1.24 | 98.27 ± 1.30 | 0.892 |
| | $V_{81.4\,Gy}$ (%) | 0 ± 0 | 0 ± 0 | 1.000 |
| | HI | 1.03 ± 0.01 | 1.03 ± 0.01 | 0.824 |
| | CN | 0.73 ± 0.23 | 0.75 ± 0.23 | 0.874 |
| Bladder | gEUD (Gy) | 63.25 ± 5.55 | 62.86 ± 5.57 | 0.925 |
| | NTCP | 0.13 ± 0.08 | 0.12 ± 0.07 | 0.920 |
| Rectum | gEUD (Gy) | 63.57 ± 2.35 | 63.07 ± 1.99 | 0.755 |
| | NTCP | 0.10 ± 0.04 | 0.09 ± 0.03 | 0.785 |

The number of generated apertures is not obviously reduced by the proposed method. In addition, the uniformity of the target has not been improved, especially in the case of multiple targets. In the future, the solution of these problems will be our focus.

## V. CONCLUSION

The column generation method based on region growth consists of four steps. First, a nonlinear transformation of the gradient of the beamlets was carried out. Second, the initial shape was obtained by region growth. Third, the adjacent negative gradients were integrated. Finally, the pricing problem and master problem were solved. From the experimental results, compared with generic column generation for IMRT treatment planning, the computation and planning times are reduced, the conformability of the target is improved, and the OARs are effectively protected by column generation based on region growth. Specifically, the NTCP of the proposed method is reduced by up to 3.37%, and the maximum acceleration rate is 20.44%. The effectiveness and feasibility of the proposed method were verified for different cancer cases. In conclusion, column generation based on region growth for IMRT treatment planning should be introduced into routine clinical practice for the radiotherapy of head and neck cancer as well as prostate cancer.

## REFERENCES


[1] D. Cao, M. K. N. Afghan, J. Ye, F. Chen, and D. M. Shepard, "A generalized inverse planning tool for volumetric-modulated arc therapy," *Phys. Med. Biol.*, vol. 54, no. 21, pp. 6725–6738, Nov. 2009. doi: 10.1088/0031-9155/54/21/018.

[2] S. Luan, C. Wang, D. Cao, D. Z. Chen, D. M. Shepard, and C. X. Yu, "Leaf-sequencing for intensity-modulated arc therapy using graph algorithms," *Med. Phys.*, vol. 35, no. 1, pp. 61–69, Jan. 2008. doi: 10.1118/1.2818731.

[3] D. M. Shepard, D. Cao, M. K. N. Afghan, and M. A. Earl, "An arc-sequencing algorithm for intensity modulated arc therapy," *Med. Phys.*, vol. 34, no. 2, pp. 464–470, Feb. 2007. doi: 10.1118/1.2409239.

[4] J. L. Bedford *et al.*, "Treatment of lung cancer using volumetric modulated arc therapy and image guidance: A case study," *Acta Oncol.*, vol. 47, no. 7, pp. 1438–1443, 2008. doi: 10.1080/02841860802282778.

[5] K. Otto, "Volumetric modulated arc therapy: IMRT in a single gantry arc," *Med. Phys.*, vol. 35, no. 1, pp. 310–317, Jan. 2008. doi: 10.1118/1.2818738.

[6] J. Unkelbach et al., "Optimization approaches to volumetric modulated arc therapy planning," Med. Phys., vol. 42, no. 3, pp. 1367–1377, Mar. 2015. doi: 10.1118/1.4908224.
[7] M. A. Earl, D. M. Shepard, S. Naqvi, X. A. Li, and C. X. Yu, "Inverse planning for intensity-modulated arc therapy using direct aperture optimization," Phys. Med. Biol., vol. 48, no. 8, pp. 1075–1089, Apr. 2003. doi: 10.1088/0031-9155/48/8/309.
[8] 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., vol. 36, no. 6, pp. 2328–2339, Jun. 2009. doi: 10.1118/1.3132234.
[9] A. Cassioli and J. Unkelbach, "Aperture shape optimization for IMRT treatment planning," Phys. Med. Biol., vol. 58, no. 2, pp. 301–318, Jan. 2013. doi: 10.1088/0031-9155/58/2/301.
[10] 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, no. 3, pp. 838–862, 2012. doi: .10.1137/040606612
[11] E. Ludlum and P. Xia, "Comparison of IMRT planning with two-step and one-step optimization: a way to simplify IMRT," Phys. Med. Biol., vol. 53, no. 3, pp. 807–821, Feb. 2008. doi: 10.1088/0031-9155/53/3/018.
[12] 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., vol. 29, no. 6, pp. 1007–1018, Jun. 2002. doi: 10.1118/1.1477415.
[13] B. van Asselen, M. Schwarz, C. van Vliet-Vroegindeweij, J. V. Lebesque, B. J. Mijnheer, and E. M. F. Damen, "Intensity-modulated radiotherapy of breast cancer using direct aperture optimization," Radiotherapy Oncol., vol. 79, no. 2, pp. 162–169, May 2006. doi: 10.1016/j.radonc.2006.04.010.
[14] J. L. Bedford and S. Webb, "Constrained segment shapes in direct-aperture optimization for step-and-shoot IMRT," Med. Phys., vol. 33, no. 4, pp. 944–958, Apr. 2006. doi: 10.1118/1.2163832.
[15] A. M. Bergman, K. Bush, M. P. Milette, I. A. Popescu, K. Otto, and C. Duzenli, "Direct aperture optimization for IMRT using Monte Carlo generated beamlets," Med. Phys., vol. 33, no. 10, pp. 3666–3679, Oct. 2006. doi: 10.1118/1.2336509.
[16] L. Zhu, L. Lee, Y. Ma, Y. Ye, R. Mazzeo, and L. Xing, "Using total-variation regularization for intensity modulated radiation therapy inverse planning with field-specific numbers of segments," Phys. Med. Biol., vol. 53, no. 23, pp. 6653–6672, Dec. 2008. doi: 10.1088/0031-9155/53/23/002.
[17] M. Broderick, M. Leech, and M. Coffey, "Direct aperture optimization as a means of reducing the complexity of intensity modulated radiation therapy plans," Radiation Oncol., vol. 4, no. 1, pp. 1–7, Feb. 2009. doi: 10.1186/1748-717X-4-8.
[18] R. A. C. Siochi, "Optimized removal of the tongue-and-groove underdose via constrained partial synchronization and variable depth recursion," Phys. Med. Biol., vol. 54, no. 5, pp. 1369–1381, Mar. 2009. doi: 10.1088/0031-9155/54/5/017.
[19] 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, no. 1, pp. 307–314, Jan. 2007. doi: 10.1118/1.2403966.
[20] Y. Li, J. Yao, and D. Yao, "Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy," Phys. Med. Biol., vol. 48, no. 20, pp. 3353–3374, Oct. 2003. doi: 10.1088/0031-9155/48/20/007.
[21] C. Cotrutz and L. Xing, "Segment-based dose optimization using a genetic algorithm," Phys. Med. Biol., vol. 48, no. 18, pp. 2987–2998, Sep. 2003. doi: 10.1088/0031-9155/48/18/303.
[22] B. Hårdemark, A. Liander, H. Rehbinder, and J. Löf, "Direct machine parameter optimization with ray machine in pinnacle," Raysearch Lab., Stockholm, Sweden, White Paper, 2003.
[23] F. Carlsson, "Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy," Med. Phys., vol. 35, no. 9, pp. 3828–3838, Sep. 2008. doi: 10.1118/1.2964096.
[24] F. Preciado-Walters, M. P. Langer, R. L. Rardin, and V. Thai, "Column generation for IMRT cancer therapy optimization with implementable segments," Ann. Oper. Res., vol. 148, no. 1, pp. 65–79, Nov. 2006. doi: 10.1007/s10479-006-0080-1.
[25] C. Men, H. E. Romeijn, Z. C. Ta kin, and J. F. Dempsey, "An exact approach to direct aperture optimization in IMRT treatment planning," Phys. Med. Biol., vol. 52, no. 24, pp. 7333–7352, Dec. 2007. doi: 10.1088/0031-9155/52/24/009.
[26] C. Men, X. Jia, and S. B. Jiang, "GPU-based ultra-fast direct aperture optimization for online adaptive radiation therapy," Phys. Med. Biol., vol. 55, no. 15, pp. 4309–4319, Aug. 2010. doi: 10.1088/0031-9155/55/15/008.
[27] P. Zhang, L. Zhang, J. Yang, and Z. Gui, "The aperture shape optimization based on fuzzy enhancement," IEEE Access, vol. 6, pp. 35979–35987, Jun. 2018. doi: 10.1109/ACCESS.2018.2849208.
[28] J. Yang, P. Zhang, L. Zhang, H. Shu, B. Li, and Z. Gui, "Particle swarm optimizer for weighting factor selection in intensity-modulated radiation therapy optimization algorithms," Phys. Med., vol. 33, pp. 136–145, Jan. 2017. doi: 10.1016/j.ejmp.2016.12.021.
[29] J. Yang, Z. Gui, L. Zhang, and P. Zhang, "Aperture generation based on threshold segmentation for intensity modulated radiotherapy treatment planning," Med. Phys., vol. 45, no. 4, pp. 1758–1770, Apr. 2018. doi: 10.1002/mp.12819.
[30] M. A. Renaud, M. Serban, and J. Seuntjens, "On mixed electron–photon radiation therapy optimization using the column generation approach," Med. Phys., vol. 44, no. 8, pp. 4287–4298, Aug. 2017. doi: 10.1002/mp.12338.
[31] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," SIAM J. Sci. Comput., vol. 16, no. 5, pp. 1190–1208, 1995. doi: 10.1137/0916069.
[32] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," ACM Trans. Math. Softw., vol. 23, no. 4, pp. 550–560, 1997. doi: 10.1145/279232.279236.
[33] J. L. Morales and J. Nocedal, "Remark on 'Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization,'" ACM Trans. Math. Softw., vol. 38, no. 1, pp. 1–4, Nov. 2011. doi: 10.1145/2049662.2049669.
[34] A. Ahnesjö, M. Saxner, and A. Trepp, "A pencil beam model for photon dose calculation," Med. Phys., vol. 19, no. 2, pp. 263–273, Apr. 1992. doi: 10.1118/1.596856.
[35] J. O. Deasy, A. I. Blanco, and V. H. Clark, "CERR: A computational environment for radiotherapy research," Med. Phys., vol. 30, no. 5, pp. 979–985, May 2003. doi: 10.1118/1.1568978.
[36] M. L. Kessler et al., "Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans," Optim. Eng., vol. 6, no. 4, pp. 421–448, Dec. 2005. doi: 10.1007/s11081-005-2066-2.
[37] G. Mu, E. Ludlum, and P. Xia, "Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer," Phys. Med. Biol., vol. 53, no. 1, pp. 77–88, Jan. 2008. doi: 10.1088/0031-9155/53/1/005.
[38] L. B. Marks et al., "Use of normal tissue complication probability models in the clinic," Int. J. Radiat. Oncol. Biol. Phys., vol. 76, no. 3, pp. S10–S19, 2010. doi: 10.1016/j.ijrobp.2009.07.1754.
[39] F. Carlsson, "Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy," Med. Phys., vol. 35, no. 9, pp. 3828–3838, Sep. 2008. doi: 10.1118/1.2964096.
[40] A. van't Riet, A. C. A. Mak, M. A. Moerland, L. H. Elders, and W. van der Zee, "A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: Application to the prostate," Int. J. Radiat. Oncol. Biol. Phys., vol. 37, no. 3, pp. 731–736, Feb. 1997. doi: 10.1016/S0360-3016(96)00601-3.
[41] N. Hodapp, "The ICRU Report 83: Prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT)," Strahlentherapie Onkologia, vol. 188, no. 1, pp. 97–99, Jan. 2012, doi: 10.1007/s00066-011-0015-x.
[42] Q. Wu and R. Mohan, "Algorithms and functionality of an intensity modulated radiotherapy optimization system," Med. Phys., vol. 27, no. 4, pp. 701–711, Apr. 2000. doi: 10.1118/1.598932.

**LIYUAN ZHANG** is currently pursuing the Ph.D. degree with the North University of China. Her research direction is the optimization of the plan of precision radiotherapy.

**ZHIGUO GUI** received the Ph.D. degree in signal and information processing from the North University of China, in 2004, where he is currently a Professor. His research interests include image processing and image reconstruction.

**JIE YANG** received the B.Sc. and M.Sc. degrees in computer science and technology, in 2004 and 2009, respectively, and the Ph.D. degree in information and communication engineering from the North University of China. Her research interests include programming, dose calculation, and planning optimization.

**PENGCHENG ZHANG** received the Ph.D. degree in computer science and technology from Southeast University, Nanjing, China, and the Ph.D. degree in traitement du signal et télécommunications from the Université de Rennes 1, Rennes, France, in 2014. He is currently involved in teaching and research at the North University of China. His research interests include medical image reconstruction, medical image analysis, dose calculation, and planning optimization.