Accounting for the tongue-and-groove effect using a robust direct aperture optimization approach

Ehsan Salaria
Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611-6595

Chunhua Men
Department of Radiation Oncology, University of California San Diego, La Jolla, California 92037-0843

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

(Received 2 August 2010; revised 4 January 2011; accepted for publication 6 January 2011; published 14 February 2011)

Purpose: Traditionally, the tongue-and-groove effect due to the multileaf collimator architecture in intensity-modulated radiation therapy (IMRT) has typically been deferred to the leaf sequencing stage. The authors propose a new direct aperture optimization method for IMRT treatment planning that explicitly incorporates dose calculation inaccuracies due to the tongue-and-groove effect into the treatment plan optimization stage.

Methods: The authors avoid having to accurately estimate the dosimetric effects of the tongue-and-groove architecture by using lower and upper bounds on the dose distribution delivered to the patient. They then develop a model that yields a treatment plan that is robust with respect to the corresponding dose calculation inaccuracies.

Results: Tests on a set of ten clinical head-and-neck cancer cases demonstrate the effectiveness of the new method in developing robust treatment plans with tight dose distributions in targets and critical structures. This is contrasted with the very loose bounds on the dose distribution that are obtained by solving a traditional treatment plan optimization model that ignores tongue-and-groove effects in the treatment planning stage.

Conclusions: A robust direct aperture optimization approach is proposed to account for the dosimetric inaccuracies caused by the tongue-and-groove effect. The experiments validate the ability of the proposed approach in designing robust treatment plans regardless of the exact consequences of the tongue-and-groove architecture. © 2011 American Association of Physicists in Medicine. [DOI: 10.1118/1.3547722]

Key words: IMRT treatment planning, aperture modulation, tongue-and-groove effect

I. INTRODUCTION

Intensity-modulated radiation therapy (IMRT) treatment planning is concerned with the design of a treatment plan for individual cancer patients. Such a treatment plan consists of a collection of apertures to be formed by a multileaf collimator (MLC) system along with associated intensities. Leaves in most commercial MLCs have a tongue-and-groove design that helps reduce interleaf leakage. However, the exposed leaf stepped sides (tongues) may undesirably block or scatter part of the radiation, resulting in underdosing of targets. It has been clinically indicated that the tongue-and-groove effect may cause underdosing as large as 10%–25% (see Refs. 1–8). On the other hand, accurately estimating the dosimetric effects of the tongue-and-groove architecture is a difficult task, especially with the commonly used beamlet-based dose models. The goal of this paper is to develop, implement, and test a robust method that takes dosimetric inaccuracies with respect to the MLC architecture into account explicitly. We will achieve this by employing upper and lower bounds on the dose distribution delivered to the patient and tailoring the optimization model to find a treatment plan that is clinically attractive with respect to these bounds, i.e., one that is of high-quality regardless of the exact effect of the MLC tongues. By testing our approach on ten clinical cases of head-and-neck cancer, we show that our approach is successful in the sense that tight dose distribution bounds can be achieved even under coarse and easily obtainable bounds on the dosimetric effects. In contrast, a treatment plan optimization approach that does not take the dosimetric inaccuracies into account yields dose distribution bounds that are loose, making it hard to accurately assess the treatment plan quality.

IMRT treatment plan optimization is traditionally performed in two sequential stages: (i) Fluence map optimization (FMO) and (ii) leaf sequencing (LS). In particular, each orientation of the accelerator head defines a rectangular beam, each of which is conceptually discretized into a set of beamlets. The FMO problem then involves determining the optimal intensities for all beamlets. Given the optimal fluence map, the LS problem decomposes the fluence map for each beam into a manageable set of deliverable apertures. Both the FMO and the LS problems have been extensively...
Studies in literature; for modeling and solution approaches to FMO, we refer to the review papers by Shepard et al. and Romeijn and Dempsey. More specifically, Lee et al. studied mixed integer programming approaches; Hamacher and Küfer proposed a multicriteria approach to the problem; and Romeijn et al. developed convex programming models. Formally, the LS problem is to determine a set of deliverable apertures for delivering a fluence map that is optimal with respect to, typically, beam-on time, number of apertures used, or total treatment time. If the objective is to minimize beam-on time and any row-convex aperture is deliverable, the LS problem is efficiently solvable (see, e.g., Refs. 17–20). In addition, Baatar et al., Boland et al., Kamath et al., Dai and Hu, and Siochi studied the problem under additional MLC hardware constraints. In contrast, as Baatar et al. showed, the problem of decomposing a fluence map into the minimum number of apertures is NP-hard. This has led to the development of a large number of heuristics for solving this problem, such as in Refs. 21, 22, 25, and 26. Additionally, Engel, Kalinowski, and Lim developed heuristics to minimize the number of apertures while constraining the total beam-on time to be minimal. Finally, Taşkın et al. proposed an integer programming approach to minimize the total treatment time. The tongue-and-groove effect has been widely addressed in literature in the LS stage. In particular, LS algorithms are used. To address this issue, an integrated approach to the problem; and Romeijn et al. developed convex programming models. Formally, the LS problem is to select a set of deliverable apertures that yield a high-quality treatment plan. Despite the tongue-and-groove architecture of the MLC, this problem is efficiently solvable (see, e.g., Refs. 17–20).

II. METHODS

II.A. A robust DAO model

We are assuming that a patient is irradiated from several predetermined beam directions. Each of these beams is discretized into a grid of beamlets, and the collection of all beamlets is denoted by $N$. Furthermore, $K$ represents the set of all deliverable apertures. Each aperture consists of a number of exposed beamlets, so we denote the set of beamlets in aperture $k \in K$ by $A_k \subseteq N$. In order to evaluate the dose distribution delivered to the patient, the patient geometry is discretized into a set $V$ of the so-called voxels. The dose delivered to voxel $j \in V$ by aperture $k \in K$ at unit intensity is the so-called aperture dose deposition coefficient denoted by $D_{kj}$. We associate a decision variable $z_j$ with each voxel $j \in V$ representing the dose received by voxel $j \in V$. Moreover, we let decision variable $y_k$ represent the intensity of aperture $k \in K$. Voxel doses can then be expressed as a linear function of the aperture intensities through the dose deposition coefficients as follows:

$$z_j = \sum_{k \in K} D_{kj} y_k, \quad j \in V.$$  

II.A.1. Bounds on the aperture dose deposition coefficients

Typically, the aperture dose deposition coefficients are expressed implicitly in terms of the beamlet dose deposition coefficients $D_{ij}$, the dose delivered to voxel $j \in V$ by beamlet $i \in N$ at unit intensity,

$$D_{kj} = \sum_{i \in A_k} D_{ij}, \quad k \in K, \quad j \in V. \quad (1)$$

This expression only provides an approximation to the aperture dose deposition coefficients (see, e.g., Refs. 41 and 43–45). In this paper, we will focus on the fact that Eq. (1) ignores the presence of the tongue-and-groove architecture of the leaves. In particular, beamlets on the boundary of aperture $k \in K$ that expose a tongue are partially blocked; Fig. 1 illustrates this issue. This will cause inaccuracies in evalu-
ing the dose distribution $z_j$, $j \in V$ when using the aperture dose deposition coefficients in Eq. (1). In fact, the right-hand side of Eq. (1) is an upper bound on $D_{kj}$ since it effectively assumes that no leaf tongue is present; in the remainder, we will denote this upper bound by

$$\bar{D}_{kj} = \sum_{i \in A_k} D_{ij}. \quad (1)$$

Thus, the realized dose distribution will be smaller than the one that is planned based on the approximation in Eq. (1), potentially causing underdosing of the target voxels.

To be able to measure the effect of the leaf tongue, let $D_{kj}^\delta$ denote the dose delivered to voxel $j \in V$ by beamlet $i \in N$ at unit intensity from which a strip of width $\delta$ is removed. (Clearly, $D_{kj}^\delta = D_{ij}$ and $D_{kj}^\delta < D_{kj}$ if $\delta > \delta'$.) We then define

$$\underline{D}_{kj}^\delta = \sum_{i \in A_k \delta_k} D_{ij} + \sum_{i \in \hat{A}_k \delta_k} D_{ij}, \quad j \in V, \quad k \in K,$n

where $\delta_k \subseteq A_k$ is the set of beamlets in aperture $k$ that expose a tongue. Now if the leaf tongue has width $\delta$, then

$$D_{kj} = \underline{D}_{kj}, \quad k \in K, \quad j \in V.$$

However, since the width of the leaf tongue is typically on the order of 0.5 mm (see, e.g., Ref. 46 or Ref. 47), it is difficult for commonly used approximate dose models to accurately account for its effect on the beamlet dose deposition coefficients. Our approach will therefore be to obtain a lower bound for these coefficients by overestimating the width of the tongue to allow for an accurate estimation of $D_{kj}^\delta$. Clearly, when $\delta$ is chosen to be larger than the width of the tongue, $\underline{D}_{kj}^\delta$ provides a lower bound on $D_{kj}$ since it effectively assumes that the leaf tongue is larger (and therefore blocks more dose) than the physical leaf tongue. For convenience, we will, in the remainder, denote the matrices of aperture dose deposition coefficients as well as their lower and upper bounds by $\bar{D}$, $\underline{D}$, and $\underline{D}$.

II.A.2. Treatment plan evaluation criteria

Now suppose a collection of treatment plan evaluation criteria, say $L$, has been identified to measure the treatment plan quality and are expressed as functions of the dose distribution: $G_{\ell}: \mathbb{R}^{|V|} \rightarrow \mathbb{R}$ for $\ell \in L$. Without loss of generality, we assume that smaller values are preferred to larger values. For convenience, we also assume that all criteria are convex. This is the case for many criteria proposed in literature, such as voxel-based penalties, equivalent uniform dose, tail means, or conditional value-at-risk, etc. In addition, it has been shown that in a multicriteria framework where the relative weights of the different criteria are varied, many others can equivalently be replaced by convex ones; e.g., tumor control probability (TCP) and normal tissue complication probability (NTCP) (see Ref. 48). Moreover, we assume that the set $L$ can be partitioned into two subsets $L = L \cup \bar{L}$; the criteria that deal with the effects of underdosing correspond to $\ell \in L$, while the criteria that deal with overdosing correspond to $\ell \in \bar{L}$. This is usually easy to do, especially if each of the criteria is a function of the dose distribution in a specific structure only. Finally, we make the mild assumption that each of the criteria is monotone in each voxel dose. In particular, if $z, z'$ are two dose distributions such that $z \preceq z'$, then $G_{\ell}(z) \leq G_{\ell}(z')$ for all $\ell \in L$ and $G_{\ell}(z) \geq G_{\ell}(z')$ for all $\ell \in \bar{L}$. In other words, uniformly increasing the dose distribution cannot deteriorate an underdosing criterion or improve an overdosing criterion.

II.A.3. Robust DAO

We next propose to use a robust DAO approach to treatment planning, i.e., an approach that identifies a treatment plan that is of high-quality for all values of the aperture dose deposition coefficients within the bounds derived in Sec. II A 1 (see, e.g., Refs. 49 and 50). This is an optimization problem formulated in terms of all deliverable apertures and their associated intensities,

$$\text{minimize } \quad \max_{D \in \underline{D} \leq D \leq \bar{D}} \sum_{\ell \in L} \gamma_\ell G_{\ell} \left( \sum_{k \in K} D_k y_k \right), \quad (R)$$

subject to

$$y_k \geq 0, \quad k \in K,$$

where the values $\gamma_\ell$ ($\ell \in L$) are the (non-negative) weights associated with the criteria and $D_k = (D_{kj}; j \in V)$ is the vector of aperture dose deposition coefficients for aperture $k \in K$. The objective function therefore measures the worst-case weighted sum of all criteria over all aperture dose deposition coefficients within the specified bounds. [In principle, upper bounds on any or all of the convex criteria functions $G_{\ell}$ ($\ell \in L$) could be accommodated as well; however, for ease of exposition, we will not explicitly incorporate this model extension.] The objective function of (R) has a mathematically inconvenient and cumbersome form. In the remainder of this section, we will therefore propose two reformulations of this model as tractable convex optimization problems. The first is a reformulation that uses the mathematical properties of the treatment plan evaluation criteria to derive a conservative bound on the objective function. The second one is equivalent to (R), but applies only if the treatment plan evaluation criteria are convex voxel-based penalty functions. Both reformulations rely on new decision variables $\bar{z}_{ij}$ and $\underline{z}_{ij}$ that represent a lower and an upper bound on the dose received by voxel $j \in V$, respectively.

Medical Physics, Vol. 38, No. 3, March 2011
sarily, the function the voxel-based penalty functions. Typically, but not neces-
subject to
at the upper bound of the delivered dose distribution. This evaluation criterion concerned with overdosing is evaluated
the delivered dose distribution, while any treatment plan evaluation criterion concerned with underdosing is evaluated at the lower bound of
conservatively: Any treatment plan evaluation criterion con-
let the objective function measure the treatment plan quality
we have implicitly incorporated the criterion weights
Fj
II.A.3.a. Reformulation 1. The first reformulation is based
on the following intuitive and appealing argument. First, re-
call that \( \bar{z} \) and \( \bar{z} \) are the lower and upper bounds on the delivered dose distribution, respectively. Then suppose we let the objective function measure the treatment plan quality conservatively: Any treatment plan evaluation criterion concerned with underdosing is evaluated at the lower bound of the delivered dose distribution, while any treatment plan evaluation criterion concerned with overdosing is evaluated at the upper bound of the delivered dose distribution. This then leads to the following DAO model:

\[
\begin{align*}
\text{minimize} & \quad \sum_{\ell \in L} y_{\ell}G_{\ell}(\bar{z}) + \sum_{\ell \in L} y_{\ell}G_{\ell}(\bar{z}), \\
\text{subject to} & \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad y_k \geq 0, \quad k \in K.
\end{align*}
\]

II.A.3.b. Reformulation 2. Now suppose that the treatment plan evaluation criteria are all voxel-based penalty functions. This means that the objective function of (R), for fixed \( D \), can be written as

\[
\sum_{j \in V} F_j \left( \sum_{k \in K} D_{kj}y_k \right),
\]

where the functions \( F_j : \mathbb{R} \rightarrow \mathbb{R} \) are convex \((j \in V)\). Note that we have implicitly incorporated the criterion weights \( y_{\ell} \) into the voxel-based penalty functions. Typically, but not necessarily, the function \( F_j \) will depend only on the structure containing the voxel \( j \in V \).

Now let us evaluate, for each voxel, the penalty function at both the upper and the lower bounds of the voxel dose and associate the maximum of these two penalties with this voxel. This yields the following DAO model:

\[
\begin{align*}
\text{minimize} & \quad \max\{F_j(\bar{z}_j), F_j(\bar{z}_j)\}, \\
\text{subject to} & \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad y_k \geq 0, \quad k \in K.
\end{align*}
\]

In other words, Eqs. (2) and (3) provide lower and upper bounds on the dose distribution.

II.A.3.a. Reformulation 1. The first reformulation is based on the following intuitive and appealing argument. First, re-
call that \( \bar{z} \) and \( \bar{z} \) are the lower and upper bounds on the delivered dose distribution, respectively. Then suppose we let the objective function measure the treatment plan quality conservatively: Any treatment plan evaluation criterion concerned with underdosing is evaluated at the lower bound of the delivered dose distribution, while any treatment plan evaluation criterion concerned with overdosing is evaluated at the upper bound of the delivered dose distribution. This then leads to the following DAO model:

\[
\begin{align*}
\text{minimize} & \quad \sum_{\ell \in L} y_{\ell}G_{\ell}(\bar{z}) + \sum_{\ell \in L} y_{\ell}G_{\ell}(\bar{z}), \\
\text{subject to} & \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad y_k \geq 0, \quad k \in K.
\end{align*}
\]

II.A.3.b. Reformulation 2. Now suppose that the treatment plan evaluation criteria are all voxel-based penalty functions. This means that the objective function of (R), for fixed \( D \), can be written as

\[
\sum_{j \in V} F_j \left( \sum_{k \in K} D_{kj}y_k \right),
\]

where the functions \( F_j : \mathbb{R} \rightarrow \mathbb{R} \) are convex \((j \in V)\). Note that we have implicitly incorporated the criterion weights \( y_{\ell} \) into the voxel-based penalty functions. Typically, but not necessarily, the function \( F_j \) will depend only on the structure containing the voxel \( j \in V \).

Now let us evaluate, for each voxel, the penalty function at both the upper and the lower bounds of the voxel dose and associate the maximum of these two penalties with this voxel. This yields the following DAO model:

\[
\begin{align*}
\text{minimize} & \quad \max\{F_j(\bar{z}_j), F_j(\bar{z}_j)\}, \\
\text{subject to} & \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad y_k \geq 0, \quad k \in K.
\end{align*}
\]

In Appendix A 2, we show that (P) is equivalent to (R) for voxel-based penalty objectives. Moreover, this optimization problem is of comparable size to \((P^c)\).

II.B. Column generation algorithm

The total number of potential apertures that needs to be included in (P) (i.e., the cardinality of the set \( K \)) is very large. For example, for an MLC in which all possible leaf settings are allowable, a beamlet grid of \( 20 \times 20 \) for each of five beams yields a total of about \( 2 \times 10^{16} \) deliverable apertures. Since this means that it is intractable to solve problems (P) and \((P^c)\) directly, we employ a column generation ap-

In Appendix A 1, we show that this reformulation is indeed, if anything, more conservative than (R), i.e., it is robust against larger deviations in the dose deposition coefficients. Moreover, this optimization problem is only marginally larger than a traditional DAO model that does not account for the tongue-and-groove effect.

II.A.3.b. Reformulation 2. Now suppose that the treatment plan evaluation criteria are all voxel-based penalty functions. This means that the objective function of (R), for fixed \( D \), can be written as

\[
\sum_{j \in V} F_j \left( \sum_{k \in K} D_{kj}y_k \right),
\]

where the functions \( F_j : \mathbb{R} \rightarrow \mathbb{R} \) are convex \((j \in V)\). Note that we have implicitly incorporated the criterion weights \( y_{\ell} \) into the voxel-based penalty functions. Typically, but not necessarily, the function \( F_j \) will depend only on the structure containing the voxel \( j \in V \).

Now let us evaluate, for each voxel, the penalty function at both the upper and the lower bounds of the voxel dose and associate the maximum of these two penalties with this voxel. This yields the following DAO model:

\[
\begin{align*}
\text{minimize} & \quad \max\{F_j(\bar{z}_j), F_j(\bar{z}_j)\}, \\
\text{subject to} & \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad \bar{z}_j + \sum_{k \in K} D_{kj}y_k, \quad j \in V, \\
& \quad y_k \geq 0, \quad k \in K.
\end{align*}
\]

III. RESULTS

III.A. Patient cases

We used a set of ten head-and-neck cases to study our model. For all cases, we designed plans using five equi-
paced \(^{60}\)Co beams around the patient. The nominal size of each beam is \(40 \times 40\) cm\(^2\). The beams are discretized into beamlets with size of \(1 \times 1\) cm\(^2\), yielding on the order of 1600 beamlets. However, we reduced the number of beam-
lets considered in the model by defining a mask for each beam eliminating any beamlets that only have a negligible contribution to target coverage. We generated data for a
voxel grid with size of $4 \times 4 \times 4 \text{ mm}^3$ for all targets and critical structures. These data were used for the evaluation of all treatment plans; however, in the optimization model, we used a coarser resolution of $8 \times 8 \times 8 \text{ mm}^3$ for unspecified tissue (full resolution was used for all targets and other critical structures). Table I shows the problem dimensions for the ten cases.

Each case contains two planning target volumes, PTV1 and PTV2, with prescription doses of 73.8 and 54 Gy, respectively. To determine the clinical quality of the treatment plans obtained by our method, we employ the following clinical dose-volume histogram (DVH) criteria (from the treatment planning protocol used in the Department of Radiation Oncology at the University of Florida):

- **PTV1**
  - At least 99% should receive at least 93% of the prescribed dose (0.93 × 73.8 = 68.6 Gy).
  - At least 95% should receive at least the prescribed dose (73.8 Gy).
  - No more than 10% should receive more than 110% of the prescribed dose (1.1 × 73.8 = 81.2 Gy).
  - No more than 1% should receive more than 120% of the prescribed dose (1.2 × 73.8 = 88.6 Gy).
- **PTV2**
  - At least 99% should receive at least 93% of the prescribed dose (0.93 × 54 = 50.2 Gy).
  - At least 95% should receive at least the prescribed dose (54 Gy).
- Salivary glands (left and right parotid glands—LPG/RPG, left and right submandibular glands—LSG/RSG)
  - No more than 50% of each gland should receive more than 30 Gy.
- **Other structures**
  - Spinal cord (SC) should receive no more than 45 Gy.
  - Brainstem (BS) should receive no more than 54 Gy.
  - Unspecified tissue (UT) should receive no more than 60 Gy.

DVH constraints on additional critical structures, such as, for example, optic nerves and chiasm, were always easily satisfied in our experiments, so we have omitted them from our results and analysis.

### III.B. Treatment plan evaluation criteria

For our proof of concept, we used treatment plan evaluation criteria that consist of one-sided quadratic voxel-based penalty functions. In particular, let $T$ denote the set of targets and $S$ the set of all structures (including targets). Moreover, let $V_s \subseteq V$ denote the set of voxels in the structure $s \in S$. (For convenience, we will, in the optimization model, assume that each voxel is assigned to a single structure only; the discussion below can, without any problems, be generalized to situations where this is not the case.) We then associate a function penalizing overdosing with the voxels in all structures and a function penalizing underdosing with the voxels in all targets. More specifically, we let (with a slight abuse of notation) $L = T \times \{-\}$ and $L = S \times \{+\}$ and define

$$G_{(s, -)}(z) = \sum_{j \in V_s} F_j(z_j), \quad s \in T,$$

$$G_{(s, +)}(z) = \sum_{j \in V_s} \bar{F}_j(z_j), \quad s \in S,$$

where

$$F_j(z_j) = \frac{1}{|V_s|} (\max(0, T_j - z_j))^2, \quad j \in V_s, \quad s \in T,$$

$$\bar{F}_j(z_j) = \frac{1}{|V_s|} (\max(0, z_j - \bar{T}_j))^2, \quad j \in V_s, \quad s \in S,$$

where Eq. (7) quadratically penalizes underdosing below the underdosing threshold $T_j$ in target $s \in T$, while Eq. (8) quadratically penalizes overdosing above the overdosing threshold $\bar{T}_j$ in the structure $s \in S$. In terms of the notation of Sec. II A 3 b, we have the following voxel-based penalty functions:

$$F_j(z_j) = \gamma_{(s, -)} F_j(z_j) + \gamma_{(s, +)} \bar{F}_j(z_j), \quad j \in V_s; \quad s \in T,$$

$$F_j(z_j) = \gamma_{(s, +)} \bar{F}_j(z_j), \quad j \in V_s; \quad s \in S \setminus T.$$

We can then write this using a smaller number of treatment plan evaluation criteria by defining

$$G_s(z_s) = \sum_{j \in V_s} \max\{F_j(z_j), \bar{F}_j(z_j)\}, \quad s \in S.$$

### III.C. Tongue-and-groove effect bounds

To account for the inaccuracies caused by the tongue-and-groove effect, we provide a lower and an upper bound on the realized dose distribution according to Eqs. (2) and (3). For that purpose, we calculated the dose deposition coefficients for the actual beamlets as well as for the reduced beamlets from which a strip of width $\delta$ was removed. In commercial:

<table>
<thead>
<tr>
<th>Case</th>
<th>No. of structures</th>
<th>No. of beamlets</th>
<th>Model No. of voxels</th>
<th>Full No. of voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>813</td>
<td>17 108</td>
<td>85 017</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>1282</td>
<td>23 998</td>
<td>104 298</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>1320</td>
<td>36 288</td>
<td>189 234</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>1471</td>
<td>38 609</td>
<td>195 113</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>935</td>
<td>15 916</td>
<td>86 255</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>692</td>
<td>13 783</td>
<td>58 636</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>1044</td>
<td>21 241</td>
<td>102 262</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>1005</td>
<td>18 609</td>
<td>84 369</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>822</td>
<td>14 520</td>
<td>71 873</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>1721</td>
<td>40 198</td>
<td>148 294</td>
</tr>
</tbody>
</table>
MANUALLY TUNED THE MODEL PARAMETERS AND USED THIS SET OF PARAMETERS TO SOLVE THE PROBLEM FOR ALL TEN PATIENT CASES. ALL EXPERIMENTS WERE PERFORMED IN MATLAB 2009B ON A 2.33 GHz INTEL CORE 2 DUO COMPUTER WITH 2 GB OF RAM USING WINDOWS OPERATING SYSTEM.

AS MENTIONED EARLIER IN THIS SECTION, WE USE OUR COLUMN GENERATION ALGORITHM TO SOLVE THE INSTANCES OF (P), AND AT EACH ITERATION OF THE ALGORITHM, WE SOLVE THE PRICING PROBLEM TO DETERMINE IF THERE EXISTS ANY PROMISING APERTURE THAT CAN IMPROVE THE TREATMENT PLAN. AS THE ALGORITHM PROGRESSES AND THE NUMBER OF APERTURES EXPLICITLY INCORPORATED IN THE MODEL INCREASES, THE MARGINAL BENEFIT OF ADDITIONAL APERTURES BECOMES CLINICALLY INSIGNIFICANT. THEREFORE, WE TERMINATE THE ALGORITHM BY MONITORING THE CLINICAL DVH CRITERIA DESCRIBED ABOVE. MORE SPECIFICALLY, WE STOP THE ALGORITHM IF, IN THE LAST FIVE ITERATIONS, THE RANGE OF OBSERVED DVH CRITERION VALUE SPANS LESS THAN $\Delta$, WHERE $\Delta=0.2\%$ FOR TARGETS AND $\Delta=2\%$ FOR CRITICAL STRUCTURES. ON THE AVERAGE, IT TAKES AROUND 3–5 MIN FOR THE TRADITIONAL DAO METHOD AND 7–9 MIN FOR THE ROBUST DAO METHOD PER CLINICAL CASE TO CONVERGE.

TABLE II. LOWER AND UPPER BOUNDS FOR TARGET DVH CRITERIA OBTAINED BY THE NEW AND TRADITIONAL MODELS FOR $\delta=0.5$ MM (IN % VOLUME).

<table>
<thead>
<tr>
<th>Case</th>
<th>New model</th>
<th>Traditional model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTV1</td>
<td>68.6 Gy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68.6 Gy</td>
</tr>
<tr>
<td>1</td>
<td>(100,100)</td>
<td>(95.98)</td>
</tr>
<tr>
<td>5</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
<tr>
<td>6</td>
<td>(100,100)</td>
<td>(96.99)</td>
</tr>
</tbody>
</table>

III.D. IMPLEMENTATION AND RESULTS

THE OPTIMIZATION PROBLEMS OF THE FORM (P) WERE SOLVED FROM OUR UFORT TREATMENT PLANNING SYSTEM USING OUR CUSTOM PRIMAL-DUAL INTERIOR POINT ALGORITHM (SEE REF. 51). WE MANUALLY TUNED THE MODEL PARAMETERS (I.E., THE UNDERDOSING AND OVERDOSING THRESHOLDS AS WELL AS THE WEIGHTS ASSOCIATED WITH THE EVALUATION CRITERIA) BASED ON THE TWO CASES. WE THEN USED THIS SET OF PARAMETERS TO SOLVE (DIFFERENT VARIANTS OF) THE PROBLEM FOR ALL TEN PATIENT CASES. ALL EXPERIMENTS WERE PERFORMED IN MATLAB 2009B ON A 2.33 GHz INTEL CORE 2 DUO COMPUTER WITH 2 GB OF RAM USING WINDOWS OPERATING SYSTEM.

AS MENTIONED EARLIER IN THIS SECTION, WE USE OUR COLUMN GENERATION ALGORITHM TO SOLVE THE INSTANCES OF (P), AND AT EACH ITERATION OF THE ALGORITHM, WE SOLVE THE PRICING PROBLEM TO DETERMINE IF THERE EXISTS ANY PROMISING APERTURE THAT CAN IMPROVE THE TREATMENT PLAN. AS THE ALGORITHM PROGRESSES AND THE NUMBER OF APERTURES EXPLICITLY INCORPORATED IN THE MODEL INCREASES, THE MARGINAL BENEFIT OF ADDITIONAL APERTURES BECOMES CLINICALLY INSIGNIFICANT. THEREFORE, WE TERMINATE THE ALGORITHM BY MONITORING THE CLINICAL DVH CRITERIA DESCRIBED ABOVE. MORE SPECIFICALLY, WE STOP THE ALGORITHM IF, IN THE LAST FIVE ITERATIONS, THE RANGE OF OBSERVED DVH CRITERION VALUE SPANS LESS THAN $\Delta$, WHERE $\Delta=0.2\%$ FOR TARGETS AND $\Delta=2\%$ FOR CRITICAL STRUCTURES. ON THE AVERAGE, IT TAKES AROUND 3–5 MIN FOR THE TRADITIONAL DAO METHOD AND 7–9 MIN FOR THE ROBUST DAO METHOD PER CLINICAL CASE TO CONVERGE.

TABLE II COMPARES THE DVH CRITERIA ASSOCIATED WITH TARGET COVERAGE OBTAINED BY THE ROBUST AND TRADITIONAL DAO METHODS USING $\delta=0.5$ FOR THREE CLINICAL CASES. FOR ALL CASES AND DVH CRITERIA, WE SHOW BOTH THE LOWER AND THE UPPER BOUNDS.

MLCs, THE TONGUE-AND-GROOVE OFFSET OF THE LEAF CAN BE AS SMALL AS 0.5 MM (SEE REF. 47). WE THEREFORE GENERATED VALUES OF $D_{ij}^\delta$ FOR $\delta \in \{0.5, 1, 3, 5\}$ (WHERE $\delta$ IS MEASURED IN MM) AND SOLVED THE DAO MODEL FOR EACH OF THESE VALUES. SINCE IT IS HARD TO ACCURATELY QUANTIFY THE EFFECT OF A VERY SMALL VALUE OF $\delta$, WE DO NOT ADVOCATE USING $\delta=0.5$ IN A CLINICAL SETTING. HOWEVER, WE USED THIS VALUE TO ASSESS THE IMPORTANCE OF EVEN SMALL DEVIATIONS FROM THE “IDEAL” (I.E., WITHOUT LEAF TONGUE) DOSE DEPOSITION COEFFICIENTS. NEXT, WE COMPARED THE QUALITY OF THE TREATMENT PLANS FOR THE COARSER BOUNDS OBTAINED USING THE LARGER VALUES $\delta \in \{1, 3, 5\}$ TO THE TREATMENT PLAN OBTAINED WITH THE TRADITIONAL MODEL ($\delta=0.5$).

FIG. 2. ISODOSE CURVES (DASHED LINES) FOR 73.8, 54, AND 30 GY ON A TYPICAL CT SLICE CORRESPONDING TO THE OPTIMAL TREATMENT PLANS OBTAINED FOR CASE 1 USING THE TRADITIONAL AND ROBUST MODELS WITH $\delta=0.5$. (a) AND (b) CORRESPOND TO THE UPPER AND LOWER BOUNDS OBTAINED USING THE TRADITIONAL MODEL, AND (c) CORRESPONDS TO THE LOWER BOUND OBTAINED USING THE ROBUST MODEL.

MEDICAL PHYSICS, VOL. 38, NO. 3, MARCH 2011
on the percent volume of a structure that receives at least the specified dose. However, due to the realistic size of the tongue width, the lower bounds are most representative of the realized delivered dose to the patient. The results therefore clearly show that treatment plans obtained by the traditional DAO model exhibit significant underdosing of PTV1 and also, to a somewhat lesser extent, of PTV2. In particular, the actual percent volume of PTV1 receiving at least the prescribed dose (73.8 Gy) in the traditional DAO model (that ignores the tongue-and-groove effect) varies from 79% to 92%, while the corresponding values in the robust DAO model are 95%–100%. This highlights the potential risk of underdosing the target when ignoring the tongue-and-groove effect during the treatment planning phase. In particular, Fig. 2 illustrates the isodose curves on a typical CT slice for clinical case 1. More specifically, isodose lines corresponding to the lower and upper bounds on the dose distribution obtained by the traditional and robust models are separately shown in the figure. For the traditional model, the lower bound on the dose distribution yields several cold spots in PTV1. However, the lower bound obtained by the robust model has significantly fewer cold spots and yields a better target coverage.

In the remainder of this section, we will focus on the robust DAO model with more conservative bounds on the dose distribution obtained with larger $\delta$ values. Tables III–VIII compare the results of our robust DAO method that accounts for the tongue-and-groove effect with results of the traditional DAO model on ten clinical cases for $\delta=1$, 3, and 5.

With respect to target coverage, the robust DAO model provides tight DVH bounds compared to the traditional DAO model. In contrast to the traditional DAO model, the robust DAO model is capable of maintaining acceptable DVH lower bounds on target coverage in the majority of the patient cases even for large values of $\delta$. It is important to note that the width of the MLC leaf tongues is (much) smaller than the values of $\delta$ that we used in our experiments. This means that the lower bounds calculated for both models can be expected to be quite loose, especially for $\delta=3$ or 5. This means that even in the few cases where the target coverage lower bounds fall slightly short of what is desired, the actual coverage should be clinically acceptable. This argument, of course, also applies to the traditional DAO model. However, for this model, the lower bounds are so far away from what is clinically acceptable that we cannot conclude that the

### Table III. Lower and upper bounds for critical-structure DVH criteria obtained by the new and traditional models for $\delta=5$ mm (in % volume).

<table>
<thead>
<tr>
<th>Case</th>
<th>LPG 60 Gy</th>
<th>RPG 60 Gy</th>
<th>LSG 60 Gy</th>
<th>RSG 60 Gy</th>
<th>BS 50 Gy</th>
<th>SC 50 Gy</th>
<th>UT 50 Gy</th>
<th>LPG 60 Gy</th>
<th>RPG 60 Gy</th>
<th>LSG 60 Gy</th>
<th>RSG 60 Gy</th>
<th>BS 50 Gy</th>
<th>SC 50 Gy</th>
<th>UT 50 Gy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(9,21)</td>
<td>(21,18)</td>
<td>(22,18)</td>
<td>(23,18)</td>
<td>(24,18)</td>
<td>(25,18)</td>
<td>(26,18)</td>
<td>(9,21)</td>
<td>(21,18)</td>
<td>(22,18)</td>
<td>(23,18)</td>
<td>(24,18)</td>
<td>(25,18)</td>
<td>(26,18)</td>
</tr>
<tr>
<td>2</td>
<td>(22,18)</td>
<td>(23,18)</td>
<td>(24,18)</td>
<td>(25,18)</td>
<td>(26,18)</td>
<td>(27,18)</td>
<td>(28,18)</td>
<td>(22,18)</td>
<td>(23,18)</td>
<td>(24,18)</td>
<td>(25,18)</td>
<td>(26,18)</td>
<td>(27,18)</td>
<td>(28,18)</td>
</tr>
<tr>
<td>3</td>
<td>(23,18)</td>
<td>(24,18)</td>
<td>(25,18)</td>
<td>(26,18)</td>
<td>(27,18)</td>
<td>(28,18)</td>
<td>(29,18)</td>
<td>(23,18)</td>
<td>(24,18)</td>
<td>(25,18)</td>
<td>(26,18)</td>
<td>(27,18)</td>
<td>(28,18)</td>
<td>(29,18)</td>
</tr>
<tr>
<td>10</td>
<td>(30,18)</td>
<td>(31,18)</td>
<td>(32,18)</td>
<td>(33,18)</td>
<td>(34,18)</td>
<td>(35,18)</td>
<td>(36,18)</td>
<td>(30,18)</td>
<td>(31,18)</td>
<td>(32,18)</td>
<td>(33,18)</td>
<td>(34,18)</td>
<td>(35,18)</td>
<td>(36,18)</td>
</tr>
</tbody>
</table>
treatment plans obtained without taking tongue-and-groove effects into account are clinically acceptable. In particular, employing the robust DAO model can significantly improve the uncertainties in PTV1 and PTV2 coverage due to the tongue-and-groove effect. For example, with the robust DAO model and \( \delta = 3 \) mm, the lower bound on PTV1 target coverage at the prescription dose of 73.8 Gy is at least 95% in eight out of ten cases, and 92% and 94% in the other two cases, respectively. For the traditional model, the corresponding lower bounds are below 10% in all cases.

With respect to structure sparing, for larger values of \( \delta \) the traditional DAO model provides smaller upper bounds on the corresponding DVH criteria than the robust DAO model. This can be explained by the fact that to ensure a desirable

<table>
<thead>
<tr>
<th>Case</th>
<th>New model</th>
<th>Traditional model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTV1</td>
<td>PTV2</td>
</tr>
<tr>
<td></td>
<td>68.6 Gy</td>
<td>73.8 Gy</td>
</tr>
<tr>
<td>1</td>
<td>(100,100)</td>
<td>(92,98)</td>
</tr>
<tr>
<td>2</td>
<td>(100,100)</td>
<td>(95,100)</td>
</tr>
<tr>
<td>3</td>
<td>(100,100)</td>
<td>(97,100)</td>
</tr>
<tr>
<td>4</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
<tr>
<td>5</td>
<td>(100,100)</td>
<td>(99,100)</td>
</tr>
<tr>
<td>6</td>
<td>(100,100)</td>
<td>(94,99)</td>
</tr>
<tr>
<td>7</td>
<td>(100,100)</td>
<td>(96,100)</td>
</tr>
<tr>
<td>8</td>
<td>(100,100)</td>
<td>(97,100)</td>
</tr>
<tr>
<td>9</td>
<td>(100,100)</td>
<td>(95,100)</td>
</tr>
<tr>
<td>10</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>New model</th>
<th>Traditional model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTV1</td>
<td>PTV2</td>
</tr>
<tr>
<td></td>
<td>68.6 Gy</td>
<td>73.8 Gy</td>
</tr>
<tr>
<td>1</td>
<td>(100,100)</td>
<td>(94,98)</td>
</tr>
<tr>
<td>2</td>
<td>(100,100)</td>
<td>(96,99)</td>
</tr>
<tr>
<td>3</td>
<td>(100,100)</td>
<td>(97,100)</td>
</tr>
<tr>
<td>4</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
<tr>
<td>5</td>
<td>(100,100)</td>
<td>(99,100)</td>
</tr>
<tr>
<td>6</td>
<td>(100,100)</td>
<td>(95,99)</td>
</tr>
<tr>
<td>7</td>
<td>(100,100)</td>
<td>(98,100)</td>
</tr>
<tr>
<td>8</td>
<td>(100,100)</td>
<td>(99,100)</td>
</tr>
<tr>
<td>9</td>
<td>(100,100)</td>
<td>(98,100)</td>
</tr>
<tr>
<td>10</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>New model</th>
<th>Traditional model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTV1</td>
<td>PTV2</td>
</tr>
<tr>
<td></td>
<td>68.6 Gy</td>
<td>73.8 Gy</td>
</tr>
<tr>
<td>1</td>
<td>(100,100)</td>
<td>(94,98)</td>
</tr>
<tr>
<td>2</td>
<td>(100,100)</td>
<td>(96,99)</td>
</tr>
<tr>
<td>3</td>
<td>(100,100)</td>
<td>(97,100)</td>
</tr>
<tr>
<td>4</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
<tr>
<td>5</td>
<td>(100,100)</td>
<td>(99,100)</td>
</tr>
<tr>
<td>6</td>
<td>(100,100)</td>
<td>(95,99)</td>
</tr>
<tr>
<td>7</td>
<td>(100,100)</td>
<td>(98,100)</td>
</tr>
<tr>
<td>8</td>
<td>(100,100)</td>
<td>(99,100)</td>
</tr>
<tr>
<td>9</td>
<td>(100,100)</td>
<td>(98,100)</td>
</tr>
<tr>
<td>10</td>
<td>(100,100)</td>
<td>(100,100)</td>
</tr>
</tbody>
</table>
target coverage for all aperture dose deposition coefficients in the range $D^\delta = D \subseteq \bar{D}$ (i.e., a desirable lower bound on target coverage), the robust DAO model is forced to select apertures that provide more dose to critical structures. However, please note that in the majority of the cases, the robust DAO model satisfies the DVH criterion for a given structure whenever the traditional DAO model satisfies this criterion. Not surprisingly, the difference between the bounds obtained by the robust and traditional models decreases with the value of $\delta$.

Figures 3 and 4 illustrate the DVHs of the optimal treatment plans obtained by the robust and traditional DAO models for case 5 with $\delta=3$ mm. For a given structure, the solid and dashed lines represent the upper and lower bounds on DVH values, respectively. The upper bounds correspond to an idealized dose distribution that ignores the presence of the leaf tongues, while the lower bounds indicate how much the actual delivered dose distribution may deviate from the optimized one.

Finally, we also evaluated the ability of the more conservative model (P3) in providing high-quality robust treatment plans that account for the tongue-and-groove effect. As noted before, when using treatment plan evaluation criteria that are voxel-based penalty functions (as described in Sec. III B), we do not need to use the more conservative model, but can derive an exact robust model. However, it is still interesting to compare the two approaches since the more conservative one should be used if other criteria are desired. These experiments showed that the differences between the conservative and the exact robust models in terms of DVH bounds are negligible. We illustrate the results of the two models for $\delta=3$ in Tables IX and X.

![Fig. 3. Lower (dashed) and upper (solid) bounds on the DVHs of the optimal treatment plan obtained by the robust model for case 5 using $\delta=3$ mm.](image-url)
IV. CONCLUSION

In this paper, we developed a robust direct aperture optimization method for IMRT treatment planning that explicitly accounts for the dosimetric inaccuracies caused by the tongue-and-groove architecture of the MLC leaves. Our approach does not rely on being able to accurately compute these dosimetric effects. Instead, we employ dose calcula-

![Dose Volume Histogram (Traditional Model)](image)

**Table IX.** Lower and upper bounds for target DVH criteria obtained by the conservative and exact robust models for $\delta=3$ mm (in % volume).

<table>
<thead>
<tr>
<th>Case</th>
<th>Conservative model</th>
<th>Robust model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(100,100)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(100,100)</td>
<td></td>
</tr>
</tbody>
</table>

**Table X.** Lower and upper bounds for critical-structure DVH criteria obtained by the conservative and exact robust models for $\delta=3$ mm (in % volume).

<table>
<thead>
<tr>
<th>Case</th>
<th>Conservative model</th>
<th>Robust model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(23,25)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(73,74)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(21,21)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(9,10)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(43,46)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(28,31)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(0,1)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(29,30)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(3,5)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(10,11)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Lower (dashed) and upper (solid) bounds on the DVHs of the optimal treatment plan obtained by the traditional model for case 5 using $\delta=3$ mm.
tions that overestimate the dimensions of the leaf tongues so that commonly used dose models can be applied. Our model then aims to find a robust treatment plan that can obtain treatment plans of high clinical quality regardless of the exact consequences of the tongue-and-groove architecture. Due to the computational difficulty of the resulting robust optimization model, we propose a reformulation that provides a conservative bound on the robust model as well as an exact and tractable reformulation for the case of voxel-based penalty criteria. We investigated the performance of the robust and traditional DAO models in the presence of the tongue-and-groove effect on ten clinical head-and-neck cancer cases. The experiments validate the ability of the proposed approach in designing robust treatment plans. Although the suggested robust optimization approach is presented to account for the tongue-and-groove effect, in principle, it could be applied to any source of dosimetric inaccuracies for which a lower and an upper bound on the beamlet dose deposition coefficients can be provided. Future research can extend the research in this paper by generalizing this approach to account for dosimetric inaccuracies caused by employing approximate dose calculation methods.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. DMI-0457394/CMMI-0852727.

APPENDIX A: REFORMULATIONS

1. Reformulation 1

It is easy to see that the objective function satisfies

\[
\max_{D \leq D \leq \bar{D}} \sum_{k \in L} \gamma_k G\left( \sum_{k \in K} D_{kj} y_k \right)
\]

\[
\leq \sum_{\ell \in L} \gamma_{\ell} \max_{D \leq D \leq \bar{D}} G\left( \sum_{k \in K} D_{kj} y_k \right). \tag{A1}
\]

Since for each \( \ell \in L \) the function \( G_\ell \) is nonincreasing, we have

\[
\max_{D \leq D \leq \bar{D}} G\left( \sum_{k \in K} D_{kj} y_k \right) = G\left( \sum_{k \in K} D_{\ell j} y_k \right),
\]

while for each \( \ell \in \bar{L} \) the function \( G_\ell \) is nondecreasing, we have

\[
\max_{D \leq D \leq \bar{D}} G\left( \sum_{k \in K} D_{kj} y_k \right) = G\left( \sum_{k \in K} \bar{D}_{kj} y_k \right).
\]

This implies that the objective function of (P) is an upper bound on that of (R).

2. Reformulation 2

In contrast to Eq. (A1), we can now exactly reformulate the objective function as follows:

\[
\max_{D \leq \bar{D}} \sum_{j \in V} F_j \left( \sum_{k \in K} D_{kj} y_k \right)
\]

\[
= \sum_{j \in V} \max_{D \leq \bar{D}} F_j \left( \sum_{k \in K} D_{kj} y_k \right),
\]

where \( D_{kj} = (D_{kj}; k \in K) \) is the vector of aperture dose deposition coefficients corresponding to voxel \( j \in V \) (and similarly for \( \bar{D}_{kj} \) and \( \bar{D}_{kj} \)). Then for each \( j \in V \), we have, due to the convexity of \( F_j \) and the non-negativity of \( y_k (k \in K) \),

\[
\max_{D \leq \bar{D}} F_j \left( \sum_{k \in K} D_{kj} y_k \right)
\]

\[
= \max \left\{ F_j \left( \sum_{k \in K} D_{kj} y_k \right), F_j \left( \sum_{k \in K} \bar{D}_{kj} y_k \right) \right\}.
\]

This implies that the objective function of (P) is equal to that of (R).

APPENDIX B: PRICING PROBLEM

1. Formulating the pricing problem

In this section, we formulate the pricing problem of the column generation algorithm described in Sec. II B and develop a solution approach to solve it. Let \( \pi_j \) and \( \bar{\pi}_j (j \in V) \) and \( \rho_k (k \in K) \) be the dual multipliers associated with constraints (4)–(6). In order to be able to accommodate both the two reformulations in Sec. II A 3, we will actually study a slightly more general model, which, with a slight abuse of notation, allows for the treatment plan evaluation criteria to be a function of both \( \bar{z} \) and \( \bar{z} \): \( G(\bar{z}, \bar{z}) \). Since the objective function is convex and the constraints are linear, the Karush-Kuhn-Tucker (KKT) conditions (see, e.g., Ref. 52) are necessary and sufficient conditions for optimality of (P). Assuming, for convenience, that the objective function is differentiable, they can be written as follows:

\[
\bar{z}_j = \sum_{k \in K} \bar{D}_{kj} y_k, \quad j \in V,
\]

\[
\bar{z}_j = \sum_{k \in K} \bar{D}_{kj} y_k, \quad j \in V,
\]

\[
\pi_j = \sum_{\ell \in L} \gamma_\ell \frac{dG(\bar{z}, \bar{z})}{d\bar{z}_j}, \quad j \in V,
\]

\[
\bar{\pi}_j = \sum_{\ell \in L} \gamma_\ell \frac{dG(\bar{z}, \bar{z})}{d\bar{z}_j}, \quad j \in V,
\]

\[
\rho_k = \sum_{j \in V} \bar{D}_{kj} \pi_j + \sum_{j \in V} \bar{D}_{kj} \bar{\pi}_j, \quad k \in K,
\]

\[
y_k \rho_k = 0, \quad k \in K,
\]

\[
y_k \rho_k \geq 0, \quad k \in K.
\]

(If the treatment plan evaluation criteria are convex but not everywhere differentiable, we can use the generalized KKT
conditions derived by Hiriart-Urruty. The analysis in the remainder of this section remains essentially unchanged. Any solution of the system above can be characterized by a vector of aperture intensities \( y \geq 0 \); this vector then determines \( z, \hat{z}, \pi, \hat{\pi}, \) and \( \rho \). Now let \( (\hat{y}; \hat{\pi}, \hat{\pi}, \hat{\rho}) \) be an optimal pair of primal and dual solutions to a restricted version of (P) in which only apertures in the set \( \hat{K} \subset K \) are considered. In other words, \( \hat{y}_k = 0 \) for \( k \in K \setminus \hat{K} \). This solution is optimal to (P) if and only if \( \hat{\rho}_k \geq 0 \) for all \( k \in K \) since in this case the set of KKT conditions for (P) will be fully satisfied. In order to check whether or not these constraints are satisfied, we formulate the so-called pricing problem as follows:

\[
\text{minimize}_{k \in K} \hat{\rho}_k = \sum_{j \in V} D_{ij}^\delta \hat{\pi}_j + \sum_{j \in V} \hat{D}_{kj} \hat{\pi}_j.
\]

Furthermore, we have

\[
\sum_{j \in V} D_{ij}^\delta \hat{\pi}_j + \sum_{j \in V} \hat{D}_{kj} \hat{\pi}_j = \sum_{j \in V} \left( \sum_{i \in A_k \setminus \partial_k} D_{ij} + \sum_{i \in \partial_k} D_{ij} \right) \hat{\pi}_j + \sum_{j \in V} \left( \sum_{i \in A_k \setminus \partial_k} D_{ij} \hat{\pi}_j + \sum_{j \in V} D_{ij} \hat{\pi}_j \right)
\]

We can interpret the objective function of the pricing problem as follows: as we increase the intensity of a beamlet \( i \in A_k \setminus \partial_k \), \( \Sigma_{j \in V} D_{ij} \hat{\pi}_j \) represents the per unit change to the objective function value due to underdosing effects, while \( \Sigma_{j \in V} D_{ij} \hat{\pi}_j \) represents the per unit change in the objective function value due to overdosing effects. Similarly, for a beamlet \( i \in A_k \setminus \partial_k \), \( \Sigma_{j \in V} D_{ij} \hat{\pi}_j \) and \( \Sigma_{j \in V} D_{ij} \hat{\pi}_j \) represent the analogous per unit changes to the objective function value, respectively.

2. Solving the pricing problem

In our pricing problem, the cost of aperture \( k \in K \), \( \hat{\rho}_k \), clearly depends on the beamlets that are exposed in the aperture (i.e., the set \( A_k \)). First, note that the pricing problem decomposes by beam direction; i.e., we can solve the pricing problem for each individual beam direction. Without incorporating the tongue-and-groove effect, Romeijn et al.\( ^{40} \) showed that the pricing problem decomposes by beamlet row, i.e., we can find the optimal solution to the pricing problem by independently finding an optimal pair of leaf settings for each beamlet row. However, when the tongue-and-groove effect is taken into account, the leaf settings in adjacent beamlet rows are important since these determine the set \( \partial_k \). We therefore formulate and solve the pricing problem for a given beam using a modification of the network flow model that was developed in Ref. 40 for solving the pricing problem under interdigitation constraints. In particular, consider a fixed beam with a beamlet grid of dimensions \( m \times n \) (i.e., \( m \) rows and \( n \) columns). We then create a network where each node corresponds to a potential leaf setting in a particular beamlet row; i.e., a typical node is characterized as \( (r, c_1, c_2) \), where \( r \) indicates the beamlet row and \( (c_1, c_2) \) represent the rightmost beamlet blocked by the left leaf and the leftmost beamlet blocked by the right leaf in row \( r \) (\( r = 1, \ldots, m \); \( c_1 = 0, \ldots, n \); and \( c_2 = 1, \ldots, n + 1 + c_1 < c_2 \)). (If \( c_1 = 0 \), the left leaf blocks no beamlets, and if \( c_2 = n + 1 \), the right leaf blocks no beamlets.) In addition, we define source node, say 0, and sink node, say \( (m + 1, 0, 1) \), representing the top (“beamlet row 0”) and bottom (soon “beamlet row \( m + 1 \)” of the aperture. We then define arcs from all nodes in beamlet row \( r \) to all nodes in a beamlet row \( r + 1 \) (\( r = 0, \ldots, m \)). Figure 5 illustrates the structure of this network for a small case of \( m = 2 \) rows and \( n = 2 \) columns.

Now note that there is a one-to-one correspondence between paths from the source node 0 to the sink node \( (m + 1, 0, 1) \) and deliverable apertures. Without loss of generality, assume that the tongue between rows \( r \) and \( r + 1 \) are part of the latter and hence partially block the former, i.e., row \( r \). Next, we assign a cost to each arc, which is defined as the cost of the exposed beamlets corresponding to the origin node of the arc. In particular, consider nodes \( (r, c_1, c_2) \) and \( (r + 1, c'_1, c'_2) \). Then \( c_1 + 1, \ldots, c_2 - 1 \) are the exposed beamlets at row \( r \) and \( \{c_1 + 1, \ldots, c'_1 \} \cup \{c'_2 - 2, \ldots, c_2 - 1\} \) are the ones among those that expose a tongue due to the leaf setting in row \( r + 1 \). [Note that if \( c'_1 \geq c' \), the former set is empty (i.e., the left leaf of row \( r + 1 \) does not expose a tongue), and similarly, if \( c_2 \leq c' \), the latter set is empty (i.e., the right leaf of row \( r + 1 \) does not expose a tongue) (see Fig. 6).] Representing the beamlet in row \( r \) and column \( c \) by \( (r, c) \), the cost associated with the arc from node \( (r, c_1, c_2) \) to node \( (r + 1, c'_1, c'_2) \) is equal to

![Fig. 5. The network model for the pricing problem under row-convexity constraints.](image)
The number of arcs in this network is

be solved in an amount of time that is proportional to the number of arcs in the network. Since it is easy to see that the total number of nodes in the network is \( O(mn^2) \), the problem can be solved in \( O(mn^2) \) time (see, e.g., Ref. 54).

Finally, we note that under interdigitation constraints (i.e., the left leaf of a row cannot overlap with the right leaf of an adjacent row) and connectedness constraints (i.e., the rows in which at least one beamlet is exposed are consecutive, relevant if the left and right leaves cannot entirely block a beamlet row and backup jaws are required), the pricing problem can be solved in a very similar way (and with the same running time) using slight modifications of the algorithms for these cases presented in Ref. 40.

\[ \sum_{c=1}^{c_1} \left( \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j + \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j \right) \]

(\( r = 1, \ldots, n \)). Finally, we assign a cost of zero to the arcs from the source node 0 to the nodes corresponding to beamlet row 1. The length of any path from the source node to the sink node is then equal to the objective function value in the pricing problem of the corresponding aperture. Therefore, the optimal solution to the pricing problem can be found by solving a shortest path problem in the network described above. Since the network is acyclic, this problem can be solved in an amount of time that is proportional to the number of arcs in the network. Since it is easy to see that the total number of nodes in this network is \( O(mn^2) \), the problem can be solved in \( O(mn^2) \) time (see, e.g., Ref. 54).

\[ \min_{c \leq c_0 + 1} \left( \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j + \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j \right) \]

\[ \sum_{c=1}^{c_2-1} \left( \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j + \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j \right) \]

\[ + \sum_{c=c_1+1}^{c_2-1} \left( \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j + \sum_{j \in V} D_{(r,c)}(j) \hat{c}_j \right) \]
Salari, Men, and Romeijn: Incorporating tongue-and-groove effect using robust direct aperture optimization


