

# Fast treatment plan modification with an over-relaxed Cimmino algorithm

Chuan Wu<sup>a)</sup>

*Department of Radiation Oncology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205*

Robert Jeraj

*Department of Medical Physics, University of Wisconsin, Madison, Wisconsin 53706, TomoTherapy Inc., 1240 Deming Road, Madison, Wisconsin 53717, and Jozef Stefan Institute, Ljubljana, Slovenia*

Weiguo Lu

*TomoTherapy Inc., 1240 Deming Road, Madison, Wisconsin 53717*

Thomas. R. Mackie

*Department of Medical Physics, University of Wisconsin, Madison, Wisconsin 53706 and TomoTherapy Inc., 1240 Deming Road, Madison, Wisconsin 53717*

(Received 10 February 2003; revised 17 September 2003; accepted for publication 17 October 2003; published 9 January 2004)

A method to quickly modify a treatment plan in adaptive radiotherapy was proposed and studied. The method is based on a Cimmino-type algorithm in linear programming. The fast convergence speed is achieved by over-relaxing the algorithm relaxation parameter from its sufficient convergence range of  $(0, 2)$  to  $(0, \infty)$ . The algorithm parameters are selected so that the over-relaxed Cimmino (ORC) algorithm can effectively approximate an unconstrained re-optimization process in adaptive radiotherapy. To demonstrate the effectiveness and flexibility of the proposed method in adaptive radiotherapy, two scenarios with different organ motion/deformation of one nasopharyngeal case were presented with comparisons made between this method and the re-optimization method. In both scenarios, the ORC algorithm modified treatment plans have dose distributions that are similar to those given by the re-optimized treatment plans. It takes us using the ORC algorithm to finish a treatment plan modification at least three times faster than the re-optimization procedure compared. © 2004 American Association of Physicists in Medicine.  
[DOI: 10.1118/1.1631913]

Key words: adaptive radiotherapy, Cimmino, treatment plan modification, optimization

## I. INTRODUCTION

In radiotherapy, the actual dose delivery suffers from various nonideal situations including setup errors and patient motion. As a result, the actual dose delivered in the patient might differ from the planned dose. Such dose errors can be remedied through adaptive procedures<sup>1</sup> by modifying the original treatment plan taking the previously detected dose error into account for subsequent dose delivery. For IMRT delivery, re-optimization is an intuitive and effective method to modify treatment plans in adaptive radiotherapy.<sup>2</sup> However, computationally, the re-optimization procedure may not always be considered favorable given the computing powers available at most clinical facilities. This re-optimization efficiency problem would be more pronounced for situations that require fast treatment plan modifications, such as intra-fraction patient motion correction, where ideally real time plan modification during dose delivery is desired. Therefore it is useful to study the possibility of implementing some fast treatment plan modification schemes in adaptive radiotherapy. On helical tomotherapy units, delivery modification based on the binary MLC leaf delivery sinograms<sup>3</sup> has been proposed. In delivery modification, the original optimized beam fluence in sinogram is quickly modified to account for simple patient offsets that can be defined by six patient offset parameters (three for patient translation and three for patient rotation). This technique lacks the ability to deal with general

patient motion involving organ/ROI deformation. Kapatoes *et al.*<sup>4</sup> studied a fast beam fluence adjustment strategy for treatment plan modification without re-optimization. In their work, the desirable dose change in certain voxels was mapped back to their contributing pencil beams geometrically through a conversion from dose to TERMA and then to incident pencil beam energy fluence. This method provides a very fast modification of the pencil beam energy fluence. However, the modified dose distributions are often not very favorable due to the approximate nature of the method.

In this work, a simultaneous projection feasibility algorithm—the Cimmino algorithm<sup>5</sup>—is used to increase plan modification speed by trading off the modification quality. The Cimmino algorithm is considered mainly because of its fast convergence speed, its implementation flexibility by assigning different weighting factors to different voxels, its successful application in solving various radiotherapy related problems,<sup>6–9</sup> and its potential for implementation on parallel computer architectures if desired.<sup>10,11</sup>

## II. MATERIALS AND METHODS

### A. An over-relaxed Cimmino algorithm

The Cimmino algorithm was introduced as an iterative linear algorithm to solve both linear equalities and inequalities.<sup>5</sup> To apply the Cimmino algorithm to radio-

therapy treatment plan modification problems, the following set of linear inequality problems is formatted:<sup>6</sup>

$$d_L \leq Dw \leq d_U$$

or

$$(d_L)_i \leq \sum_j D_{ij} w_j \leq (d_U)_i, \quad w \geq 0, \quad (1)$$

where  $(d_L)_i$  and  $(d_U)_i$  represent the lower and upper dose bounds assigned to voxel  $i$ . The dose given to voxel  $i$  is computed by matrix multiplication of the dose transfer matrix  $D$  and the beam weight vector  $w$ , whose components  $w_j$ 's are to be determined/modified. The Cimmino algorithm has the capability of searching for a feasible solution  $w$  of Eq. (1) if it exists, otherwise it converges to a least-square optimization solution related to the differences between the computed dose  $Dw$  and dose bounds  $d_L$  and  $d_U$ <sup>12,13</sup> [refer to Eq. (A2)]. Hoffner *et al.*<sup>14</sup> observed an interesting fact that by over-relaxing the relaxation parameter  $\lambda_k$  in the original Cimmino algorithm from its sufficient convergence interval  $(0, 2)$  to  $(0, \infty)$ , the convergence of the Cimmino algorithm can be greatly accelerated without significantly sacrificing the quality of the algorithm outputs.

In their over-relaxation scheme, the relaxation parameter  $\lambda'_k$  is increased from iteration  $k$  to  $k+1$  according to (the notation  $\lambda'_k$  is used in the over-relaxed Cimmino algorithm to differentiate it from its counter part  $\lambda_k$  in the original Cimmino algorithm):

$$\lambda'_{k+1} = \alpha_k \lambda'_k \quad (\alpha_k > 1) \quad (2)$$

as long as merit function  $G_k$  shows an improvement from the previous iteration. Their merit function  $G_k$  is given by<sup>14</sup>

$$G_k = \sum_i c_i \max(0, D^i w_k - (d_U)_i) + \sum_i c_i \max(0, (d_L)_i - D^i w_k), \quad (3)$$

where  $c_i$  is the voxel weighting factors assigned to voxel  $i$ . The  $G_k$  function in Eq. (3) measures the weighted linear accumulative deviation of the doses that exceed either lower or upper dose bounds. When the first  $G_{k+1} \geq G_k$  occurs at  $k=N$ , then the over-relaxation scheme is terminated and  $w_N$  is a good approximation of the final convergence result that would have been generated by using  $\lambda_k \in (0, 2)$  in the Cimmino algorithm. The over-relaxation scheme in Eq. (2) is proposed based on the assumption about the linearity of the change in beam weights  $w_k$ : multiple step updating of  $w_k$  in the original Cimmino algorithm with  $\lambda_k \in (0, 2)$  could be approximated by a single step updating of  $w_k$  with  $\lambda'_k > 2$ . To further explore this linearity assumption, the following continued over-relaxation procedure is proposed for iterations  $k > N$ , where  $N$  is the final iteration at which Eq. (2) terminates:

$$\lambda'_{k+1} = \beta_k \lambda'_k \quad (\beta_k < 1). \quad (4)$$

Or more explicitly, for iterations  $k > N$ , the subsequent iterations can be performed:

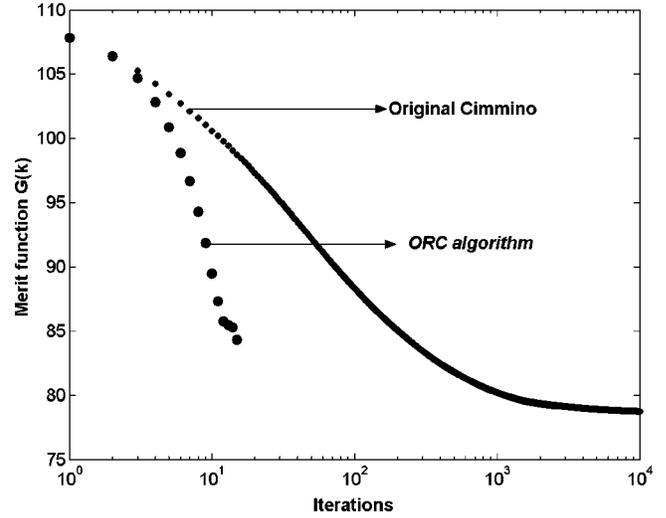


FIG. 1. The comparison between the original Cimmino algorithm ( $\lambda_k = 1.0$ ) and the ORC algorithm in a typical adaptive radiotherapy treatment plan modification case study (refer to Sec. III A for case details). The iteration indices ( $X$  axis) is plotted in log scale and the merit function  $G_k$  is plotted on  $Y$  axis.

*Step 1:* Update  $\lambda'_k$  according to Eq. (4); if  $\lambda'_{k+1} < 2$ , terminates, else continue to *step 2*.

*Step 2:* Updating  $w_k$  to  $w_{k+1}$  using Cimmino algorithm with  $\lambda'_{k+1}$ .

*Step 3:* If  $G_{k+1} < G_k$ , go to *step 1*; else, set  $w_{k+1} = w_k$ , go to *step 1*.

The introduction of the above-mentioned procedure for iterations  $k > N$  is based on the expectation that if a larger  $\lambda'_k$  fails to generate a better set of beam weights that corresponds to a better merit function score in the over-relaxation scheme, then a reduced  $\lambda'_k$  might be able to do so by giving a better linear approximation with smaller updating step length. A graphical representation of the convergence of the

TABLE I. Proposed parameter values for the ORC algorithm implementation in treatment plan modification.

ORC parameters	Parameter interpretations	Suggested values
$(d_L)_i$	Lower dose bound assigned to voxel $i$	Dose prescription in re-optimization if applicable
$(d_U)_i$	Upper dose bound assigned to voxel $i$	Dose prescription in re-optimization if applicable
$\lambda'_k$	Relaxation parameter	Varies according to Eq. (2) and Eq. (4); $\lambda'_1 = 1.0$
$\alpha_k$	ORC controlling parameter	$\alpha_k = 1.5$
$\beta_k$	Continued ORC controlling parameter	$\beta_k = 0.5$
$G_k$	Merit function	Quadratic dose objective function
$c_i$	Weighting factor assigned to voxel $i$	Derived equivalent weighting factors (Appendix)
Selected voxels	Voxels included in the ORC algorithm	All ROI voxels in plan
Selected beams	Beams $w_j$ 's updated by the ORC algorithm	Non-trivial beams in the original plan

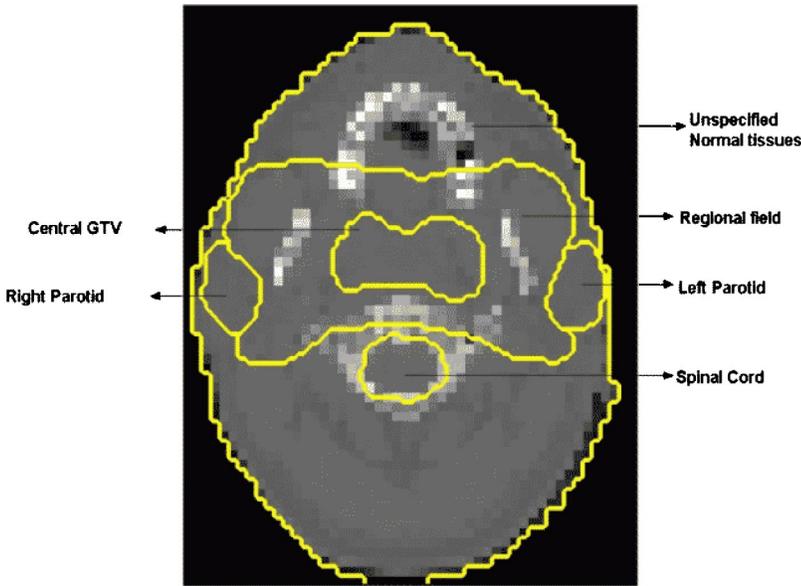


FIG. 2. One CT slice with contoured ROIs for the nasopharyngeal case setup. The ROIs include the central GTV, the surrounding regional field, the left and right parotids, the spinal cord and the rest is treated as unspecified normal tissues. The regional field is treated as part of the tumor. The parotids and spinal cord are treated as critical structures.

over-relaxed Cimmino (ORC) algorithm combining Eqs. (2) and (4) is shown in Fig. 1: the ORC algorithm rapidly lowers the merit function score  $G_k$  within 10–20 iterations to a value that is close to the final convergence result by the original Cimmino algorithm ( $\lambda_k=1.0$ ) after 10 000 iterations. In order to reach the same level of  $G_k$  value as the ORC algorithm, the original Cimmino algorithm would have to take approximately 10 times more iterations. The small distinctive tail after the tenth iteration on the ORC algorithm convergence curve corresponds to the further improvement made by the continued over-relaxation scheme with decreasing  $\lambda_k'$  as given by Eq. (4).

### B. Initial beam weights dependence

The main problem investigated in this work is how to quickly modify a treatment plan when needed. Because the ORC algorithm is used in an adaptive situation, it is natural to start the ORC algorithm from the originally planned beam weights  $w_0$ , instead of zero beam weights or a set of beam weights with equally weighted nonzero components. In this way, the ORC algorithm is expected to quickly generate an approximate correction for beam weights based on previously optimized  $w_0$ . The importance of a good starting point to the ORC algorithm was also briefly mentioned by Hoffner *et al.*<sup>14</sup> In their case studies, they stated that the first ten iterations were performed with  $\lambda_k < 2$  in order to get a “stable” state to start the subsequent ORC iterations. In the adaptive scenario of this work,  $w_0$  can always provide such a “stable” starting point for the ORC algorithm. Therefore, the ORC iterations can be started directly without first running a few regular Cimmino iterations with  $\lambda_k < 2$ . This could save significant computing time: in this work, it takes the ORC algorithm approximate 15 iterations to finish—compared to potential of 25 iterations (including 10 initial stabilization iterations), the number of Cimmino iterations is thus approximately reduced by 40%.

### C. Algorithm parameters and flexibility

To effectively use the ORC algorithm in adaptive radiotherapy, we propose selecting a set of algorithm parameters such that the ORC iterations can approximate an *unconstrained* quadratic re-optimization process (re-optimization with a quadratic dose objective function without any optimization constraints). Based on the experience of this work, by using the ORC algorithm with the set of algorithm parameters in Table I we are able to generate acceptable results.

In Table I, the lower/upper dose bounds assigned to each voxel  $d_L/d_U$  provide the driving force for the ORC algorithm. In order to approximate a quadratic dose objective function in re-optimization process, the prescribed dose values of corresponding re-optimization can be used as both the lower and upper dose bounds. For example, in a simplified two-fraction adaptive radiotherapy case where a dose error exists in the first fraction, the re-optimization dose prescription can be given as<sup>2</sup>

$$(d_L)_i = (d_U)_i = 2d_i^P - d_{1i}, \quad (5)$$

where  $d_i^P$  is the original prescribed dose per fraction to voxel  $i$  and  $d_{1i}$  is the dose actually delivered to it in the first fraction. The relaxation parameter  $\lambda_k'$  represents the weightings given to the updating term at iteration  $k$ . Its initial value is set to 1.0 and then  $\lambda_k'$  is varied systematically by Eqs. (2) and (4). The ORC controlling parameter  $\alpha_k$  in Eq. (2) is set to 1.5 for all iterations, based on the observed trade-off between the fast convergence speed (larger  $\alpha_k$ ) and possible better approximation of the original Cimmino algorithm (smaller  $\alpha_k$ ). In the continued ORC iterations with decreasing  $\lambda_k'$ , the value of the controlling parameter  $\beta_k$  in Eq. (4) is set to 0.5 for simplicity. The merit function  $G_k$  provides the criteria to measure the quality of the ORC iterations. To compare with the re-optimization process, the corresponding quadratic dose objective function in the re-optimization is used as  $G_k$ . When the equal lower/upper dose bound ( $d_L = d_U$ ) is used

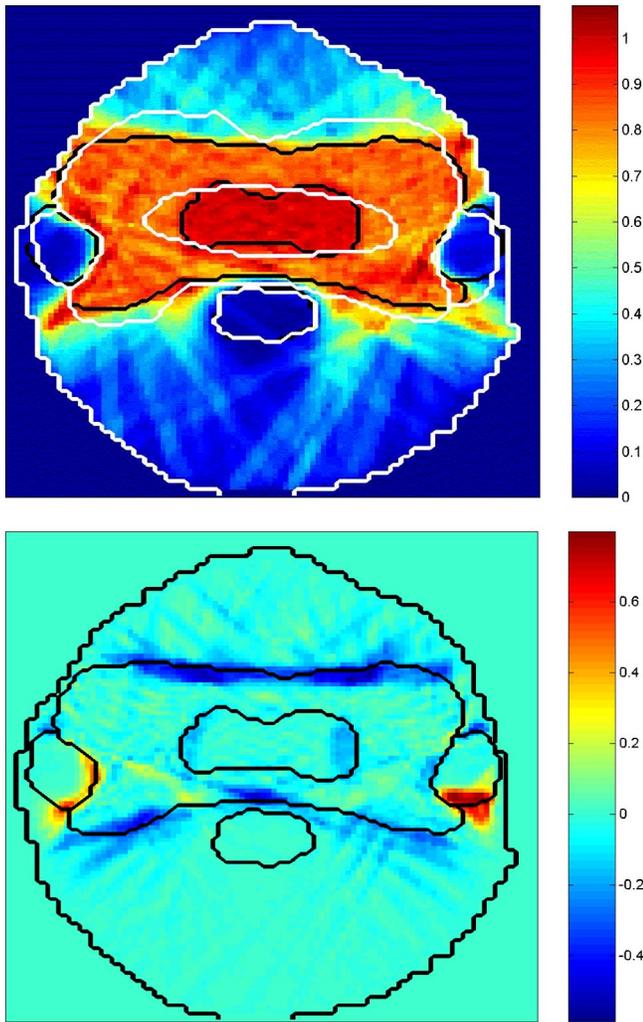


FIG. 3. The top panel shows the artificial ROI deformation introduced in the first fraction by manually recontouring the ROIs in white. The original contoured ROIs for the treatment planning is in black. The original optimized dose is also shown on the same image. The dose is normalized to the prescribed dose to the GTV per fraction. The bottom panel shows the registered dose error as a result of the top ROI deformation. The deformed ROI voxels given in the white contours are registered/mapped back to the original ROI voxels given in the original contours in black. And then the differential dose error between the delivered dose and the planned dose is computed. The distribution of cold/hot spots as well as their relative magnitude is shown on the bottom panel.

such as in Eq. (5), the problem given by Eq. (1) becomes infeasible under most situations. As mentioned before, the algorithm converges to the least-squares solution in the infeasible mode [refer to Eq. (A2)]. Based on this equivalency, a set of equivalent weighting factors  $c_i$  corresponding to the weighting factors  $I_i$  used in the quadratic dose objective function can be derived (refer to the Appendix) and is proposed for the ORC algorithm in this work. Also, the above suggested set of  $c_i$  values can provide a very good starting point for further tuning manipulations on  $c_i$  when desired. In addition to the above-noted explicit ORC algorithm parameters, the voxels that are included in a problem can significantly affect the convergence speed and the quality of the outcome.<sup>15</sup> In this work, all the ROI voxels are considered in

the ORC implementation in order to obtain good quality while sacrificing some computing time. If the convergence speed is the major concern, significantly faster iterations can be obtained by selecting only those voxels which have relatively large dose errors.

#### D. Case setup

A nasopharyngeal treatment (Fig. 2) using helical tomotherapy<sup>16</sup> was simulated with 51 equal-angularly spaced beam projections. Each of the projection angles has 64 binary multileaf collimator beam leaves, whose opening status and durations are optimized/modified as the treatment beam rotates helically around the patient. Two CT slices, which are treated during one full cycle of treatment beam rotation, are contoured and planned. The dose is computed by a convolution/superposition algorithm using a pixel size of  $0.18 \times 0.18 \text{ cm}^2$  in the CT plane and  $0.30 \text{ cm}$  along the CT slice direction. In the original treatment plan, the relative prescribed dose per fraction is such that the central GTV is 1.00 (corresponds to 70 Gy), the regional field is 0.86 (corresponds to 60 Gy) and elsewhere is zero. A typical weighted quadratic dose objective function is used both in the plan optimization and the subsequent re-optimization procedure. The ROI based weighting factors  $I_i$  is such that GTV: Regional field : Parotids : Spinal cord : Unspecified normal tissues = 8:5:1:1:0.2 [refer to Eq. (A1)]. Furthermore, in the constrained scenario (Sec. III C), a minimum dose of 50 Gy to the regional field and a maximum dose of 75 Gy to all ROIs is used. The optimization is preformed on a commercial optimization platform GAMS (General Algebraic Modeling System)<sup>17</sup> with its general gradient based CONOPT2 solver.<sup>18</sup> The ORC algorithm is implemented in MATLAB.<sup>19</sup>

### III. RESULTS

#### A. Treatment plan modification due to organ deformation

In this simplified two-fraction adaptive radiotherapy scenario, it is assumed that there exists an organ/ROI deformation in the first fraction and the patient geometry stays the same as the planning geometry in the second fraction. The original plan is to be modified for the second fraction in order to remedy this dose error. The artificial organ/ROI deformation was made by manually re-contouring the ROIs in the nasopharyngeal case as shown on the top panel of Fig. 3. The dose delivered to the deformed ROIs is mapped back to the original ROIs by a deformable registration method implemented by Lu.<sup>20</sup> The resultant registered dose error image is shown on the bottom panel of Fig. 3: major cold spots appear on both lateral edges of the central GTV, the anterior and posterior edges of the regional field, and major hot spots appear at the posterior part of the left parotid as well as along the edge of the right parotid.

To compare the ORC algorithm with the unconstrained re-optimization, both methods were applied to the above-mentioned ROI deformation scenario. The parameters in Table I were used for the ORC algorithm. The same set of

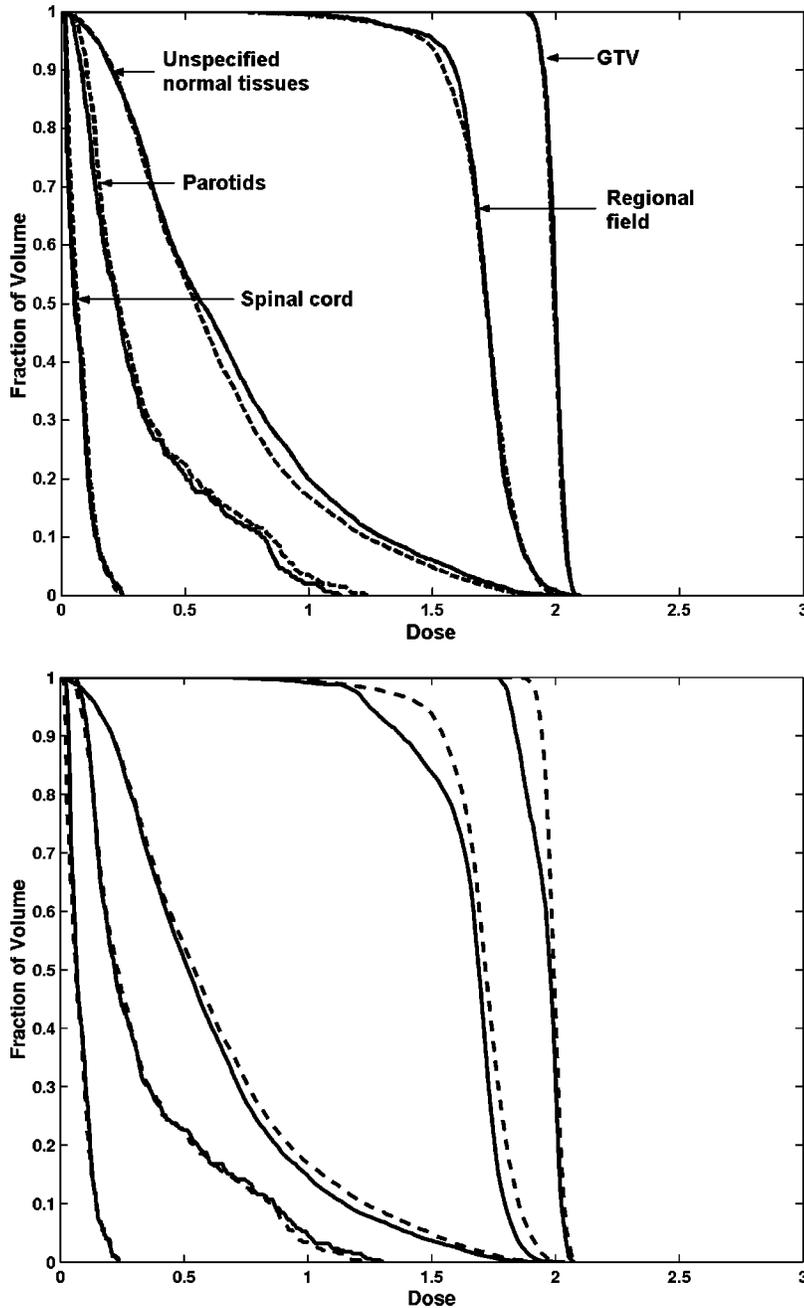


FIG. 4. The top panel compares the DVH delivered by the unconstrained re-optimized plan (solid line) and that by the ORC modified plan (dashed line) at the end of the treatment. The bottom panel compares the DVH delivered by the original plan without any modification (solid line) and that by the ORC modified plan (dashed line) at the end of the treatment. The dose is normalized to the prescribed dose to the GTV per fraction.

selected voxels and pencil beams were used in both methods. The DVH comparison between the dose delivered by the re-optimized plan and the ORC modified plan at the end of the treatment (two fractions) is shown on the top panel of Fig. 4: the two methods generate very similar DVH curves at the end of the treatment: the re-optimized DVH is slightly more uniform in the regional field and slightly lower in the parotids. This good agreement of the DVH curves indicates that when the suggested algorithm parameters in Table I are used, the ORC algorithm approximates the unconstrained re-optimization very well. The bottom panel of Fig. 4 compares the DVH delivered by the original plan without any modification and that by the ORC modified plan at the end of the treatment. The ORC modified plan delivers a significantly more uniform dose to the tumor while keeping a low dose to

sensitive structures and a slightly higher dose to the unspecified normal tissues. This demonstrates that the ORC algorithm is able to effectively remedy the cold spots in tumors without severely sacrificing the normal structures. The major goal of the adaptive radiotherapy treatment plan modification is therefore fulfilled.

Further comparison between the ORC modified plan and the re-optimized plan can be made based on the modified dose distributions for the second fraction as shown in Fig. 5. Compared to the dose error image in Fig. 3, it can be seen that the re-optimized plan delivers more spatially well-defined high dose patches to the right posterior edge of the regional field and to the severe cold spots located in the regional field posterior to the central GTV, while the ORC modified plan fails to correct them effectively. This occurs

mainly because the re-optimization process tends to generate accurate dose patterns by introducing beam weights of high frequencies (fine structures in optimized beam weights), which are found to be difficult to obtain from the ORC algorithm due to its approximation nature.

### B. Initial beam weights dependence

In treatment plan modification, it is of significant advantage to use the previous planned beam weights  $w_0$  (or previously modified beam weights) as the starting point for the ORC algorithm than to use uniform beam weights. To demonstrate this, the ORC was started from two different starting points (with the same set of algorithm parameters from Table I): the original optimized beam weights  $w_0$  and zero beam weights on the same organ deformation scenario as in Sec. III A.

The two DVH plots in Fig. 6 compare the ORC algorithm modified dose delivery started from two different initial beam weights with the same unconstrained re-optimized treatment plan. The ORC modified plan started from  $w_0$  delivers an overall better DVH than the plan started from zero beam weights: it delivers a significantly more uniform dose in the regional field and a lower dose to both the parotids and spinal cord, and a slightly higher dose to the unspecified normal tissues compared to the ORC modified plan started from zero beam weights.

### C. Treatment plan modification due to setup errors

In this similar adaptive radiotherapy scenario, it is assumed that there exists a setup error in the first fraction (Fig. 7, top) and the patient geometry stays the same as the planning geometry in the second fraction. To remedy the dose errors introduced in the first fraction (Fig. 7, bottom), the original plan is modified for the second fraction by either re-optimization or the ORC algorithm.

In this scenario, the flexibility of the ORC algorithm is demonstrated through varying the modified plan by tuning its weighting factors  $c_i$  (all the other ORC parameters are kept fixed for simplicity, although they may affect the ORC output as well). The same set of the selected voxels and pencil beams were used for both the ORC algorithm and the re-optimization. Tuning the weighting factors  $c_i$  of the ORC algorithm to reshape its modified DVH curves is a trial and error procedure, where the suggested equivalent  $c_i$  in Table I can be used as a good starting point.

The left panel in Fig. 8 compares the DVH given by the ORC modified plan with the suggested  $c_i$  values in Table I (dashed line) and that of the re-optimization with upper and lower dose constraints<sup>2</sup> (solid line): while the ORC modified plan delivers a lower dose to the parotids, it delivers a significantly less uniform dose to the regional field, where some severe cold spots need to be compensated. By increasing the weighting  $c_i$  assigned to the regional field voxels, more uniform DVH curve in it would be achieved. The right panel in Fig. 8 shows the comparison between the ORC modified plan with a set of tuned weighting factors  $c_i$  (dashed line) and that by the constrained re-optimization (solid line): the

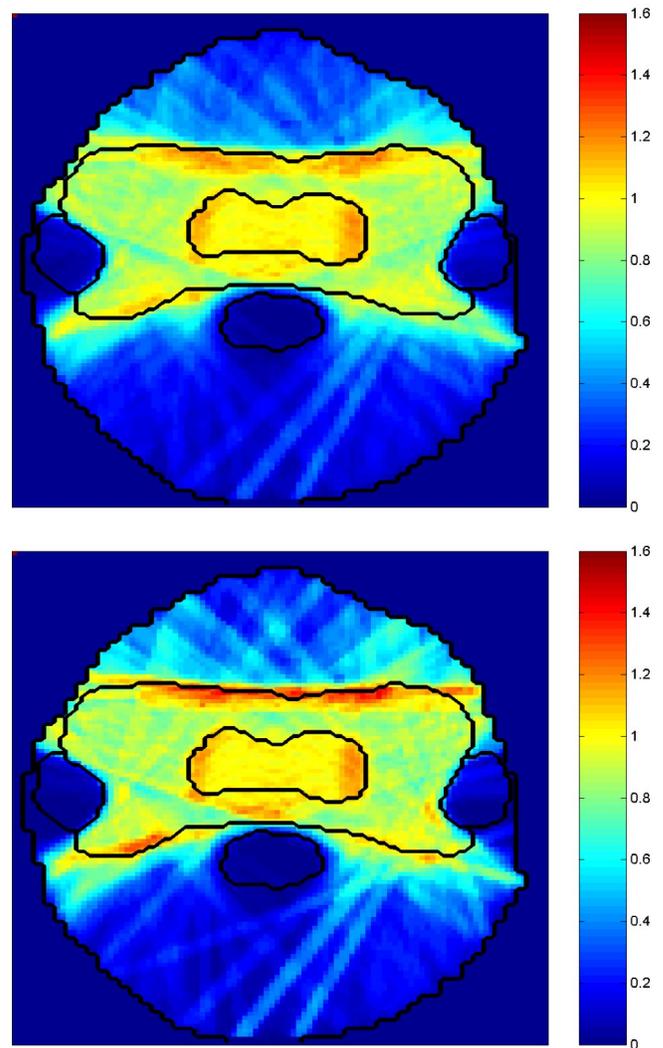


FIG. 5. The top panel shows the ORC modified dose distribution for the second fraction. The bottom panel shows the unconstrained re-optimized dose distribution for the second fraction. The dose is normalized to the prescribed dose to the GTV per fraction.

ORC algorithm with the tuned weighting factors generates a much more uniform dose coverage to the regional field at the expense of delivering slightly more dose to normal structures.

### D. Computing efficiency of the ORC algorithm

The issue of the computing efficiency of the ORC algorithm is investigated by comparing its computing time with that of the corresponding re-optimization processes. All compared procedures start from the original optimized beam weights  $w_0$ . The computer used in this work is a Pentium 4 2.4 GHz PC with 1.0 Gbytes of memory running Windows 2000. The computing time is tabulated in Table II, where the unconstrained case refers to the nasopharyngeal deformation scenario studied in Sec. III A and the constrained case refers to the setup error scenario discussed in Sec. III C. For a further comparison, the unconstrained re-optimization process was interrupted when its objective value approaches the

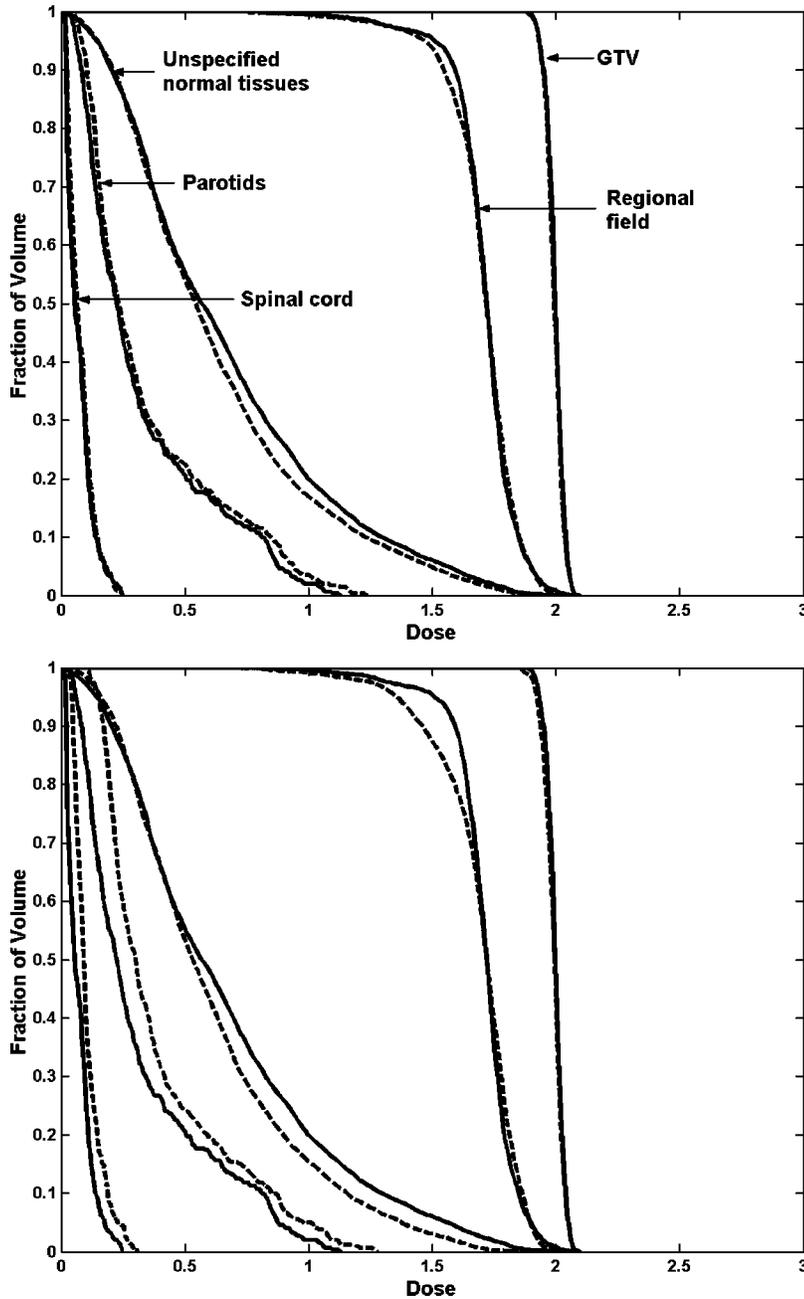


FIG. 6. The top panel compares the DVH by the ORC modified plan started from the original optimized beam weights  $w_0$  (dashed line) and by the unconstrained re-optimized plan (solid line) at the end of the treatment. The bottom panel compares the DVH by the ORC modified plan started from zero beam weights (dashed line) and by the same unconstrained re-optimized plan (solid line) at the end of the treatment. The dose is normalized to the prescribed dose to the GTV per fraction.

final  $G_k$  value of the ORC algorithm. By doing this, the approximate time needed for the re-optimization to generate a plan whose quality is close to the ORC modified plan can be recorded. This time is referred to as truncated unconstrained time in Table II. The ORC algorithm is at least 3 times faster than the unconstrained re-optimization, and it is approximately 30 times faster than the constrained re-optimization. Because the ORC algorithm converges so much faster than the constrained re-optimization, it allows us to intuitively adjust its weighting factors  $c_i$  (or other ORC algorithm parameters) by trial and error to generate acceptable modified plans, while still keeping the overall modification time short. For example, in our experience, it took only a few trials on different weighting factors  $c_i$  before a satisfactory modified plan is obtained. Furthermore, the consecu-

tive  $c_i$  factors adjustment can be completed very quickly due to the fact of that the pre-computed term  $\|D^i\|^2$  need not to be re-computed when  $c_i$  is varied.

#### IV. DISCUSSION

In radiotherapy inverse treatment planning, sometimes it is desirable to penalize the underdose and overdose in tumor differently, because underdose is considered a more serious problem than overdose from a radiobiological point of view. Different weightings  $c_i$  can be assigned to voxels depending on whether they are underdosed or overdosed in Cimmino algorithm.<sup>21</sup> In this work, almost all the voxels might be either underdosed or overdosed due to the same dose assigned to  $(d_U)$  and  $(d_L)$  as previously discussed. Assigning

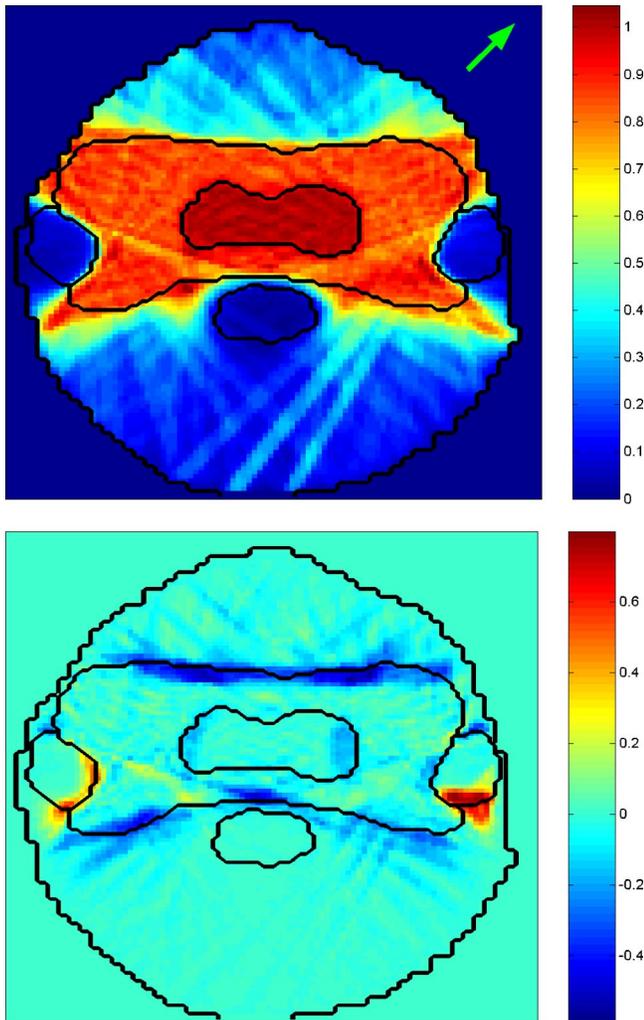


FIG. 7. The top panel shows the original plan as well as the introduction of setup errors in the first fraction. The arrow shows the direction where the patient is shifted in the CT plane due to the setup errors. The shift vector is 3 voxels (0.54 cm) in both lateral and AP/PA directions. The resultant dose error image is shown on the bottom panel. The distribution of cold spots and hot spots and their relative magnitude can be seen. The dose is normalized to the prescribed dose to the GTV per fraction.

different weighting factors based on whether voxels are underdosed or overdosed is equivalent to changing the weighting factors  $c_i$  of some of the voxels between iterations (such change happens whenever these voxels are updated from being underdosed to overdosed or vice versa during the ORC iterations). It was observed by us that the minimum dose delivered to the tumor can be increased, by assigning higher weighting factors  $c_i$  to the underdosed tumor voxels than the overdosed voxels. The convergence properties of varying  $c_i$  from iteration to iteration in the Cimmino algorithm need to be further studied.

A desirable property of the ORC algorithm in its future implementation is that the ORC algorithm has the potential advantage of parallel implementation.<sup>12,13</sup> This is expected to be helpful for realistic clinical cases when a very large number of voxels and pencil beams are involved.

The users of the ORC algorithm need to consider some

possible limitations of the algorithm in its clinical implementation. First, the suggested set of ORC algorithm parameters in Table I is meant to use the ORC algorithm to approximate an unconstrained re-optimization process. Although it has been successfully demonstrated the effectiveness of the ORC algorithm with this set of parameters, it does not guarantee that these are the “best” ORC algorithm parameters for any adaptive radiotherapy cases. In fact, many other good choices of the ORC algorithm parameters are possible. For example, in some of the cases we studied, good plan modification results were obtained by using different values for  $(d_L)$  and  $(d_U)$  to make Eq. (1) feasible. Second, the introduction of the ORC algorithm is based on a linear approximation of the original Cimmino iterations [Eqs. (2) and (4)]. However, there is no mathematical proof to guarantee that such over-relaxation technique always works, although we have never experienced a single exception in this study. In the case when the ORC algorithm fails, either the original Cimmino algorithm or the re-optimization approach can be invoked. Third, how close the ORC generated plan is to the unconstrained re-optimized plan is difficult to answer, because it is found to be case dependent. Furthermore, an additional convergence error (difference between the ORC modified plan and the re-optimization plan) would be introduced when using the ORC algorithm to approximate a constrained re-optimization by adjusting its weighting factors  $c_i$ . Due to the existence of such errors, an ORC modified plan might not be acceptable initially: for example, it is our experience that in some cases, when dose error to be corrected is relatively small (compared to the possible convergence error of the ORC algorithm), the ORC algorithm modified plans might not be closely comparable to the re-optimized plans. Under such situations, either different ORC algorithm parameters can be tried or re-optimization can be performed.

## ACKNOWLEDGMENTS

The authors thank Dr. Robert R. Meyer and Dr. Steven Wright at Department of Computer Sciences in University of Wisconsin–Madison and Dr. Censor Yair at Department of Mathematics in University of Haifa, Israel for their valuable communications, and Dr. Gustavo H. Olivera at TomoTherapy Inc. for computing the dose transfer matrix for this work. C. W. sincerely thanks Karen Berquist at the Department of Physics, College of William & Mary for her very careful reading and correction of the manuscript, which makes the writing of this work much clearer and better presented scientifically.

## APPENDIX

A generalized quadratic dose objective function in radiotherapy treatment plan optimization can be written as

$$f(w) = \sum_i \frac{I_i}{N_i} (d_i(w) - d_i^p)^2, \quad (\text{A1})$$

where  $d_i^p$  is the prescribed dose in voxel  $i$ ,  $I_i$  is the weighting

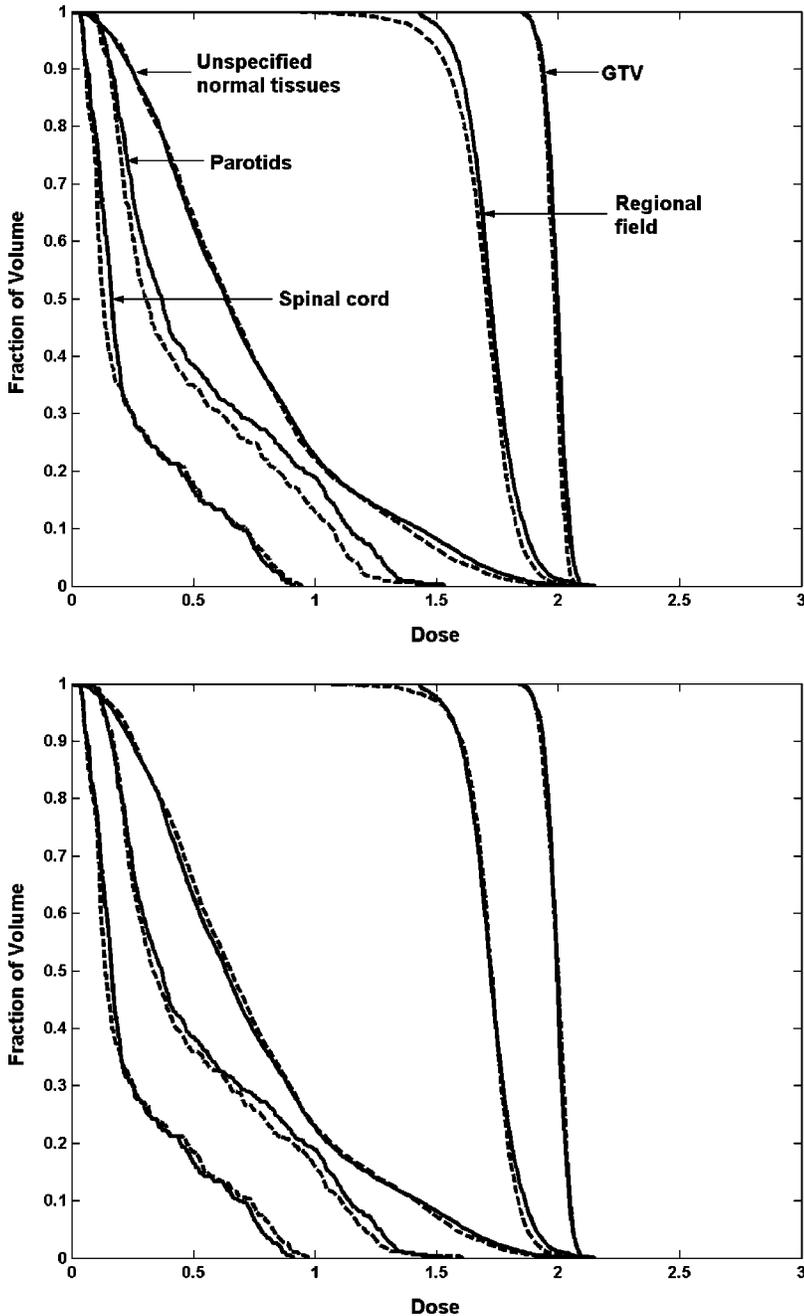


FIG. 8. The top panel compares DVH delivered by the ORC modified plan with the suggested set of weighting factors  $c_i$  from Table I (dashed line) and the hard dose constrained re-optimized plan (solid line) at the end of the treatment. The bottom panel compares the DVH delivered by the ORC modified plan with the tuned weighting factors (dashed line) and the same constrained re-optimized plan (solid line) at the end of the treatment. The suggested  $c_i$  values in Table I are calculated based on equivalent re-optimization weighting factors of  $I_i=8:5:1:1:0.2$  for GTV, regional field, parotids, spinal cord and normal tissues respectively (refer to Appendix B). The tuned  $c_i$  values correspond to  $I_i=8:14:1:1:0.2$  respectively. The dose is normalized to the prescribed dose to the GTV per fraction.

factor assigned to voxel  $i$ ,  $N_i$  denotes the number of ROI voxels to which the voxel  $i$  belongs, and  $w$  is the beam weights to be optimized. The Cimmino algorithm acts like a weighted least-squares optimization problem when Eq. (1) is

infeasible.<sup>12,13</sup> More explicitly, the Cimmino algorithm would converge to the minimization of the following function in the infeasible case:

$$g(w) = \sum_i c_i \|s_i(w)D^i\|^2, \tag{A2}$$

TABLE II. Computing time comparisons between the ORC algorithm and the direct re-optimizations. The time listed is in the units of seconds.

	Re-optimization (sec)	ORC (sec)
Unconstrained	~120	~30
Truncated unconstrained	~90	~30
Constrained	~840	~30

where  $s_i(w)$  is the projection of the beam weights vector  $w$  onto the constraints and  $\|a\|^2$  denotes the  $l_2$  norm of vector  $a$ . If given the equal lower and upper dose bound as the re-optimization dose prescription:

$$(d_L)_i = (d_U)_i = d_i^p. \tag{A3}$$

Substitute the expression for  $s_i(w)$  from the original Cimmino algorithm,<sup>6</sup> and note Eq. (A3):

$$g(w) = \sum_i c_i \left\| \frac{\langle D^i, w \rangle - d_i^p}{\|D^i\|^2} \right\|^2 \|D^i\|^2 = \sum_i c_i \frac{(d_i - d_i^p)^2}{\|D^i\|^2}. \quad (\text{A4})$$

Comparing  $g(w)$  in Eq. (A4) and  $f(w)$  in Eq. (A1), it can be concluded that if the two minimization problems given by Eqs. (A4) and (A1) were equivalent, and if there is a voxel to voxel correspondence between them, then at most there is a scaling factor  $\kappa$  between the weighting factors  $\{c_i\}$  in Eq. (B4) and  $\{I_i\}$  in Eq. (A1), i.e.,

$$\frac{I_i}{N_i} = \kappa \frac{c_i}{\|D^i\|^2}. \quad (\text{A5})$$

To determine the scaling factor  $\kappa$ , apply the normalization requirement of the  $c_i$  required by the Cimmino algorithm to Eq. (A5):

$$\kappa = \sum_i \frac{I_i}{N_i} \|D^i\|^2. \quad (\text{A6})$$

Combining Eqs. (A6) and (A5), an equivalent relation between  $\{I_i\}$  and  $\{c_i\}$  can be established as

$$c_i = \frac{I_i \|D^i\|^2}{N_i} \cdot \frac{1}{\sum_i \frac{I_i}{N_i} \|D^i\|^2}. \quad (\text{A7})$$

The set of  $c_i$  given by Eq. (A7) is used as the suggested values in Table I.

<sup>0</sup>Electronic mail: wuchuan@uams.edu

<sup>1</sup>D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Phys. Med. Biol.* **42**, 123–132 (1997).

<sup>2</sup>C. Wu, R. Jeraj, G. H. Olivera, and T. R. Mackie, "Re-optimization in adaptive radiotherapy," *Phys. Med. Biol.* **47**, 3181–3195 (2002).

<sup>3</sup>G. H. Olivera, E. E. Fitchard, P. J. Reckwerdt, K. Ruchula, and T. R. Mackie, "Delivery modification as an alternative to patient repositioning in tomotherapy," *Proceedings of ICCR 2000*, pp. 297–299.

<sup>4</sup>J. M. Kapatoes, G. H. Olivera, P. J. Reckwerdt, K. Ruchula, R. Jeraj, and

T. R. Mackie, "A fluence adjustment strategy for adaptation of dose distribution," *Med. Phys.* **28**, 1259 (2001); AAPM Annual Meeting, Salt Lake City, UT.

<sup>5</sup>G. Cimmino, "Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari," *La Ric. Sci. Roma XVI, Ser. II, Anno IX* **1**, 326–333 (1938).

<sup>6</sup>Y. Censor, M. D. Altschuler, and W. D. Powlis, "On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning," *Inverse Probl.* **4**, 607–623 (1988).

<sup>7</sup>P. Bloch, M. D. Altschuler, B. E. Bjarngard, and J. McDounough, "Determining clinical photon beam spectra from measured depth dose with the Cimmino algorithm," *Phys. Med. Biol.* **45**, 171–183 (2000).

<sup>8</sup>P. Kolmonen, J. Tervo, and T. Lahtinen, "Use of the Cimmino algorithm and continuous approximation for the dose deposition kernel in the inverse problem of radiation treatment planning," *Phys. Med. Biol.* **43**, 2539–2554 (1998).

<sup>9</sup>Y. Xiao, J. Galvin, M. Hossain, and R. Valicenti, "An optimized forward-planning technique for intensity modulated radiation therapy," *Med. Phys.* **27**, 2093–2099 (2000).

<sup>10</sup>T. Tommasini, "Cimmino type methods: Convergence parameters and parallel implementation," *Rivista di matematica pura ed applicata n.* **8**, 95–106 (1991).

<sup>11</sup>Y. Censor and S. A. Zenios, *Parallel Optimization—Theory, Algorithms, and Applications* (Oxford University Press, New York, 1997).

<sup>12</sup>A. N. Iusem and A. R. De Piero, "Convergence results for an accelerated nonlinear Cimmino algorithm," *Numer. Math.* **49**, 367–368 (1986).

<sup>13</sup>A. R. De Piero and A. N. Iusem, "A simultaneous projection method for linear inequalities," *Linear Algebr. Appl.* **64**, 243–253 (1985).

<sup>14</sup>J. Hoffner, P. Decker, E. L. Schmidt, W. Herbig, J. Rittler, and P. Weib, "Development of a fast optimization preview in radiation treatment planning," *Strahlenther. Onkol.* **172**, 384–394 (1996).

<sup>15</sup>S. M. Morrill, I. I. Rosen, R. G. Lane, and J. A. Belli, "The influence of dose constraint point placement on optimized radiation therapy treatment planning," *Int. J. Radiat. Oncol., Biol., Phys.* **19**, 129–141 (1990).

<sup>16</sup>T. R. Mackie, T. Holmes, S. Swerdloff, P. Reckwerdt, J. O. Deasy, J. Yang, B. Paliwal, and T. Kinsella, "Tomotherapy: A new concept for the delivery of conformal radiotherapy," *Med. Phys.* **20**, 1709–1719 (1993).

<sup>17</sup>A. Brooke, D. Kendrick, A. Meeraus, and R. Raman, "GAMS: A user's guide," 1998, website at <http://www.gams.com>.

<sup>18</sup>A. S. Drud, "CONOPT—A large scale GRG code," *ORSA J. Comput.* **6**, 207–216 (1994).

<sup>19</sup>MATLAB website at <http://www.mathworks.com>

<sup>20</sup>W. Lu, "Motion detection and correction for imaging guided radiation therapy," Ph.D. thesis, Department of Medical Physics, University of Wisconsin—Madison, 2001.

<sup>21</sup>G. Starksall and P. J. Eifel, "An interactive beam-weight optimization tool for three-dimensional radiotherapy treatment planning," *Med. Phys.* **19**, 155–163 (1992).