Computational Boundary Sampling to Accelerate Optimization of Intensity-modulated Radiation Therapy.

Paras Babu Tiwari[1] ,Yao Xie[1] ,Yixin Chen[1] , and Joseph O. Deasy[2]
[1] Department of Computer Science and Engineering
Washington University in Saint Louis.
[2] Memorial Sloan-Kettering Cancer Center.
{pbtiwari,y.xie,ychen25g}@wustl.edu, DeasyJ@mskcc.org

March 27, 2013

Abstract

Intensity-modulated radiation therapy (IMRT) is a widely used cancer treatment method. The optimization of IMRT is computationally expensive because of a large number of variables and constraints. We propose a boundary sampling method to reduce the computational time and space required in optimization. In our approach, we identified and included in the optimization boundary voxels for each organ. In addition, we included approximately 20% of the inner voxels selected by grid sampling to better approximate the set of voxels. We used anonymized treatment plans for the head and neck (7), prostate (1), and lung (1) to compare the dose distributions of IMRT with sampling to that of IMRT without sampling. Although the dose distributions obtained with the sampling method were approximately 1.4% worse than those obtained in the method without sampling, the sampling method accelerated optimization of IMRT by 5 to 35 times, compared with no sampling. In computation, the sampling method used 1.5 to 1.8 times less memory than no sampling. These results suggest that the computational time and space used in the optimization of IMRT could be reduced by the sampling method with minimal effect on dose distributions to organs.

1 Introduction

First described in 1982 [6], intensity-modulated radiation therapy (IMRT) is now widely used in the United States and around the world for the clinical treatment of cancer. Since the application of computed tomographic (CT) imaging in treatment planning, IMRT has become a focus of research study.

IMRT is a technique that modulates the intensity of incoming beams of radiation that originate from different directions to achieve high-dose conformability on tumor targets while sparing critical organs that may be near the tumor (e.g., heart, rectum) from radiation. IMRT offers additional degrees of freedom, compared with conventional radiation beams of uniform intensity, which gives the therapy the potential to achieve high-quality treatment plans.

However, these added degrees of freedom also increase the complexity of IMRT, causing issues with computational memory and time. During inverse planning for IMRT, for instance, each of the radiation beams coming from different directions consists of hundreds of beamlets, each of which has an intensity that can be modulated by a multileaf collimator (MLC). The dose that each beamlet contributes to each voxel is stored in an influence matrix, in which each

row corresponds to the voxels of an organ and each column corresponds to the beamlets. The optimization process for IMRT requires referencing the influence matrix and multiplying with beamlet weights, which can exceed a computer's capacity for memory due to the size of the influence matrix [10]. Similarly, the large size of the influence matrix determines the large size of mathematical models for beam intensity optimization, which turns out to be time-consuming.

Although CT scan resolution can be high, the dose grid resolution can not be set too high because of the memory limits of a computer. Without sampling, an optimization cannot be run in native scan resolution. For this reason, CERR (a Computational Environment for Radiotherapy Research) [4] offers a way to decrease grid resolution. Other techniques have been developed to accelerate IMRT optimization. As mentioned above, the decreased speed in IMRT optimization is an effect of referencing a large influence matrix and multiplication with beamlet weights. Several groups have looked into reducing the matrix size. Thieke et al. developed an importance sampling technique that integrated the low doses that can occur in beamlets, which have rapid radial fall-offs, and that could potentially speed up IMRT optimization [14]. Considering the dose on adjacent or nearby voxels are related, some groups have utilized voxel clustering and voxel sampling. In 2005, Scherrer et al. introduced an adaptive voxel clustering method with a hierarchical initialization step [12]. This method reaches optimization quickly, but the initialization step is complex and not preferred for robust IMRT optimization. Voxel sampling has been used in beam angle optimization. Ferris et al. developed an adaptive algorithm to determine the sampling rate in different organs, while using three methods-grid sampling, uniform sampling, and rind sampling-to sample within each organ[5]. In IMRT optimization, Martin et al.[8] introduced a technique using sampled voxels to compute objective and gradient for a steepest descent algorithm , which can provide an optimum solution.

In this paper, we propose a voxel sampling method called boundary sampling. In this method, we include boundary voxels and a small percentage of the inner voxels of each organ during the optimization process. A voxel is considered a boundary voxel if there are no voxels surrounding it in at least one of the four directions (left, right, up, down). Otherwise, the voxel is considered an inner voxel. To consider the contributions of every beamlet in optimization, a sufficient number of voxels should be sampled to ensure that every beamlet passes through at least one voxel. We included all the boundary voxels of the organs, with the assumption that each beamlet must pass through the boundary of an organ. Therefore, boundary sampling ensures that the contribution of every beamlet is taken into consideration during optimization. In addition, this method includes a grid-based sampling technique to evenly select voxels.

To the best of our knowledge, boundary sampling is the first method of its kind to emphasize the boundary of organs in IMRT optimization. In this paper, we will also discuss a theory that supports the new sampling method.

The boundary sampling method is based on the prioritized prescription algorithm [2] [15]. The algorithm has four steps. The first step is to optimize the objective function on target structures, while sparing key sensitive organs from radiation. In the second step, the algorithm tries to find the best beamlet weights with the solution of the first step, while sparing all critical organs from radiation. The third step finds a solution around that of the second step to decrease dose to other normal organs. Finally, dose distributions for all organs are smoothed to avoid hot and cold spots.

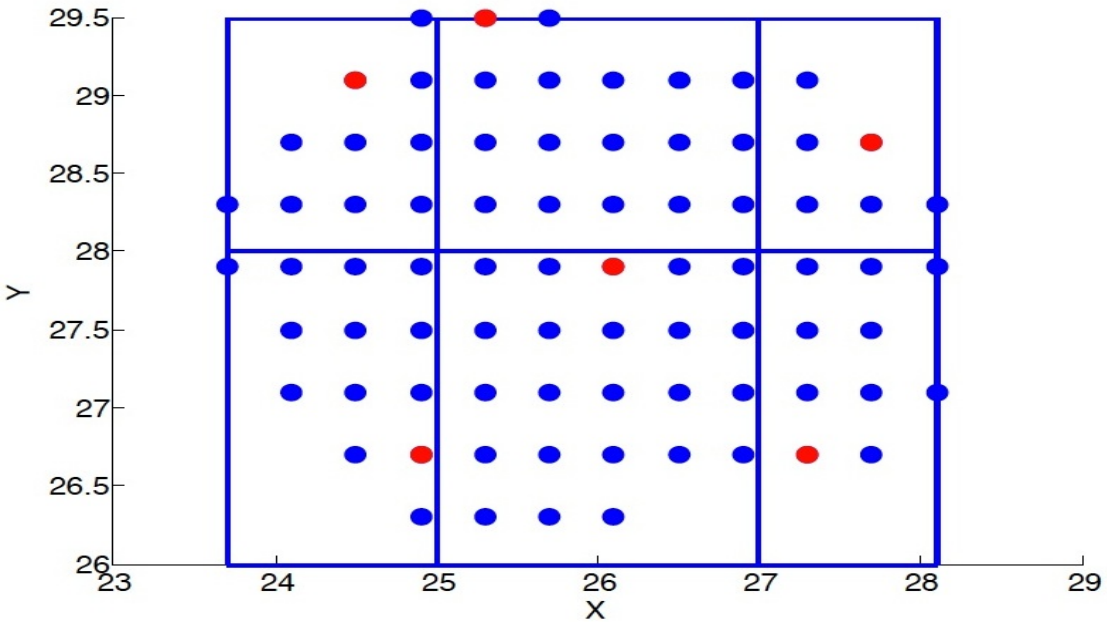
There are several reasons to base our method on the prioritized prescription algorithm. First, the IMRT optimization problem is hierarchical in nature. Objective functions on targets should have higher priority than those on sensitive organs, which should be more significant than those on other normal organs. Second, the prioritized prescription algorithm is better than a weighted-sum approach. The algorithm avoids the difficulties associated with selecting the best weight of the weighted-sum approach. Finally, the four steps of the prioritized prescription algorithm

help achieve high-dose distributions on target structures, since the last three steps find solutions around that of the first step. The weighted-sum approach has no such advantage.

We compared boundary sampling to no sampling using treatment plans for patients with head and neck, lung, and prostate cancer. The computational boundary sampling (CBS) method was 5 to 35 times faster than the no-sampling (NS) method. Similarly, CBS used around 1.5 to 1.8 times less memory that of the NS method. The quality of dose distributions with the CBS method was comparable with that of the NS method. We will also present a corresponding theory with the reasonable assumptions of the Lipschitz continuity of a beam profile and dose distributions of each beamlet.

2 Sampling Based Prioritized Optimization Algorithm

The prioritized optimization algorithm takes a substantial amount of time to solve the IMRT problem. In the sampling-based method, we consider only a subset of all voxels of structures, reducing the number of constraints and variables in IMRT optimization, which in turn significantly reduces the time needed to find an optimal solution. Figure 3.1 illustrates the voxels that are selected by the CBS method for a slice of a structure at a sampling rate of 10%. The voxels selected by the algorithm are shown in red, while those not selected by the algorithm are in blue. A grid that covers the slice was imagined, and some number of voxels were randomly selected from each grid cell.



Red: Selected Voxels.
Blue: Not Selected Voxels.

Figure 2.1: Illustration of Sampling method

2.1 Grid Sampling Algorithm

The grid sampling algorithm is used to select p% of voxels from each slice of a structure. A grid that covers the slice was imagined, and p% of voxels were randomly selected from each grid cell. The size of the grid was determined by the number of voxels in the structure. Each grid cell must have a number of voxels that equal at least p. The minimum number of grid cells is one and the maximum number of grid cells is MAXGRIDSIZE.We set the maximum grid size so that sampling is reasonably fast. In this study, we found that a grid with 100 cells sufficiently covers a slice of most organs. There were some organs, like skin, that have a large number of voxels and a grid of 100 cells might produce a coarse sampling. However, since these organs tend to be low priority, coarse sampling in such organs will not have much of an impact on their final dose distribution. Therefore, the number of grid cells for a structure is given by:

$$\text{Number of the grid cells} = \begin{cases} 100 & \text{if } \frac{|V_i|}{p} > 100 \\ 1 & \text{if } \frac{|V_i|}{p} \leq 1 \\ \frac{|V_i|}{p} & \text{Otherwise.} \end{cases} \tag{2.1}$$

The algorithm divides the total number of grid cells into a number of rows and columns based on the dimensions of the slice. The number of columns approximately equals the number of rows multiplied by the ratio of the x-dimension to the y-dimension. For example, if the x-dimension is twice more than the y-dimension then the number of columns is twice more than the number of rows. The interval x gives the increment in the x-axis and the interval y gives the increment in the y-axis. Starting from the bottom left corner of the grid, we drew p% voxels from a grid cell defined by [(min(x), min(y)),(min(x)+intervalx,min(y))]. The algorithm terminated once it reached the top rightmost grid cell.

To make things clear, let's take an example. Consider a slice has 800 number of voxels and we would like to draw 10% voxels.

Number of the grid cells = 800/10 = 80.

Suppose the slice's x-dimension is twice bigger than that of y-dimension. Therefore, row = floor(sqrt(80/2)) = 6, col = 80/6 = 13.

So the grid has 6 rows and 13 columns. Each grid cell has at least 10 voxels and we select one voxel (10% of 10 voxels) from each grid cell. Therefore total number of voxels selected equals 80, which is 10% of 800.

If the sampling rate is 20, there will be 40 grid cells, and each grid cell has at least 20 voxels. We select four voxels(20% of 20 voxels) from each grid cell. Therefore, the total number of voxels selected equals 160, which is 20% of 800.

The pseudocode for the grid sampling algorithm is given below. We processed the grid by row, selecting voxels from each cell in a row before moving to the next row. The innermost loop randomly selects p% of voxels from each grid cell.

Algorithm 1 Grid Sampling Algorithm

procedure GETPVOXELS($indicies$, x, y, z,p) ▷ Return sliceIndex
    n = Get the number of slices from the indicies variable;
    for $i \leftarrow 2, n-1$ **do**
        $|V_i|$ = Calculate the total number of voxel in i^{th} slice from the indicies variable;
        noGridCell = $\frac{|V_i|}{p}$; ▷ Get the number of cells in a grid;
        [row,col] = GetGridSize(x,y,noGridCell) ▷ Divide the slice into rows and cols such
                                                ▷ that noGridCell = row * col;
        gridElement = $\frac{|V_i|}{noGridCell}$;
        noFrmEachCell = gridElement * p%; ▷ Number of elements from each grid cell;
        [xdiff,ydiff] = GetXYRange(x,y);
        [intervalx,intervaly] = xdiff/col , ydiff/row;
        [xstart,ystart] = [min(x),min(y)];
        for $j \leftarrow 1, row$ **do**
            yend =ystart + intervaly;
            for $k \leftarrow 1, col$ **do**
                xend =xstart + intervalx;
                [xpop,ypop] = Get all voxel indicies in $[(xstart, ystart), (xend, yend)]$ gridcell;
                for $t \leftarrow 1, noFrmEachCell$ **do** ▷ Select p% voxels from each grid cell;
                    index = Get a random number in the range [1,size(xpop)];
                    [xindex,yindex] = Get indicies corresponding
                                        to [xpop(index),ypop(index)];
                    sliceIndex = sliceIndex ∪ [xindex,yindex,indicies(i)];
                end for
                xstart = xend;
            end for
            [xstart,ystart] = min(x),yend;
        end for
    end for
end procedure
procedure GETGRIDSIZE(x, y, $noGridCell$) **return** the number of rows and columns in the grid
    [xdiff,ydiff] = GetXYRange(x,y);
    times = xdiff/ydiff
    row = floor(sqrt(noGridCell/time))
    col = floor($noGridCell$/ row)
    return [row,col]
end procedure
procedure GETXYRANGE(x, y, $noGridCell$) **return** the size of x and y-axis
    return [max(x)-min(x), max(y)-min(y)]
end procedure

2.2 Optimization Formulation

Let T= $\{PTV_1, PTV_2...PTV_N\}$ denote the set of target structures, and let D_i^{pre} for $i \in T$ be the prescribed dose for the i^{th} target structure. Assume R^I denotes the set of critical organs for the first step, and D_i^{max} is the maximum dose of the i^{th} organ in R^I . Let the set of voxels in a structure be given by V_i, and $|V_i|$ denote the number of voxels in a structure. Assume $V' \subset V$ is a subset of voxels obtained by taking the union of the boundary voxels and the voxels collected by the p% sampling rate for a structure. Let A be an influence matrix, and w be the weight of the beamlets. Then the dose at the i^{th} voxel is given by:

$D_i = \sum_{j=1}^{n} A_{ij} * w_j$

2.3 Step I

The objective of the first step is to produce the prescribed dose distribution to target structures. This is achieved by minimizing the quadratic deviation of the actual dose from the prescribed dose. Generally, the physician does not want to underdose target structures because underdosing might lead to reoccurrence, but overdose does not. In this study, the minimum dose for targets must be greater than or equal to 95% of the prescribed dose. We achieved this goal by setting a constraint so that the difference between the actual dose and the prescribed dose would be less than 95% of the prescribed dose. This difference is denoted by variable t_i, and it is added in the objective function to maximize the minimum dose. The critical organs are also considered in this step,by setting a constraint so that no organ receives more than the maximum dose. . In our formulation, the prescribed dose and the maximum dose were D80 Gy and 1.15 * D80 Gy, respectively.

Let $F^I(w)$ denote the objective function for the Step I. The optimization would be:

Find w so as to

$$\text{minimize } F^I(w) = \sum_{i \in T} G_i(w) + t_i^2, \tag{1}$$

$$\text{where } G_i(w) = |V_i'|^{-1} \sum_{j \in V_i'} [D_j(w) - D_i^{pre}]^2 \qquad \forall i \in T. \tag{1a}$$

$$\text{subject to} \qquad D_i^{pre} - D_j(w) \leq t_i \qquad \forall i \in T, j \in v, \tag{1b}$$

$$0.05 D_i^{pre} \leq t_i \qquad \forall i \in T, \tag{1c}$$

$$D_j(w) \leq D_i^{max} \qquad \forall i \in R^i, j \in v_i, \tag{1d}$$

$$0 \leq w_k \qquad \forall k \in 1...N, \tag{1e}$$

$$\text{and } w_k \leq w^{max} \qquad \forall k \in 1..N \tag{1f}$$

2.4 Step II

In the second step, we minimize the mean tail dose of the hottest x%(MoHx) of the region of the critical organs without degrading much the dose computed for target structures in Step I. The formulation of MoHx for critical organs is based on the linear formulation of mean tail dose by Romejin et al[11]. We turned the results of Step I - the mean quadratic deviation from the prescribed dose and the minimum and maximum dose - into hard constraints. These hard constraints narrow the search space for finding an optimal solution. In fact, the solution for this step is identical to that of previous step.This means we could not make much progress in the second goal of reducing the dose to critical organs. Therefore, we introduced a slip factor 's'

that relaxes the constraint of the previous step and gives us more freedom to improve the second goal. In this study, we set the slip value to one.

Let w^I denote beamlets weight calculated in the Step I and R^{II} denote the set of critical organs for step II. Let D_i^{min} and D_i^{max}denote the minimum and maximum dose respectively for targets found in the Step I.

$$D_i^{min} = min\{D_j(w^I), j \in V_i\} \qquad \forall i \in T \tag{2.2}$$

$$\text{and } D_i^{max} = max\{D_j(w^I), j \in V_i\} \qquad \forall i \in T. \tag{2.3}$$

The minimization of the MoHx can be formulated as follows: Find w so as to

$$\text{minimize } F^{II}(w) = y_i^\alpha + \frac{1}{1-\alpha}\sum_{j=1}^{|V_i'|} p_{ji}^\alpha, \qquad \forall \alpha \in A_i, i \in R^{II}, j \in v_i \tag{2}$$

$$\text{subject to} \qquad D_j(w) - z_{ji} = 0 \qquad \forall i \in R^{II}, j \in V_i', \tag{2.a}$$

$$p_{ji}^\alpha \geq 0 \qquad \forall \alpha \in A_i, i \in R^{ii}, j \in v_i, \tag{2.b}$$

$$p_{ji}^\alpha - z_{ji} + y_i^\alpha \geq 0 \qquad \forall \alpha \in A_i, i \in R^{II}, j \in V_i', \tag{2.c}$$

$$D_j(w) \leq D_i^{max} \qquad \forall i \in R^I, j \in V_i', \tag{2.d}$$

$$G_i(w) \leq (1+s)G_i(w^I) \qquad \forall i \in T, \tag{2.e}$$

$$D_i^{min} \leq D_j(w) \leq D_i^{max} \qquad \forall i \in T, j \in V_i', \tag{2.f}$$

$$0 \leq w_k \qquad \forall k \in 1..N, \tag{2.g}$$

$$w_k \leq w^{max} \qquad \forall k \in 1..N \tag{2.h}$$

The α is the same as x in MoHx, and it indicates the hottest x% of voxels of structures in R^{II}. Equations (2.a) to (2.c) are taken from the linear formulation of the mean tail dose by Romejin et. al.

2.5 Step III

The third step deals with minimizing the mean dose to normal organs without degrading much the result found in the first and second steps. For this, we turned the objective of the first and second steps into constraints. We increased the slip factor quadratically,providing more freedom to improve the third goal.

Let R^{III} be the set of organs in Step III, w^{II} be the beamlets weight found in the Step II, and w be the beamlets weight that we want to compute for Step III. The mean dose for normal organs is defined as follow:

$\langle D(w)\rangle_i = |V_i|^{-1}\sum_{j\in V_i} D_j(w)$

Then, optimization would be : Find w so as to

$$\text{minimize } F^{III}(w) = \sum_{i \in R^{III}} \langle D(w) \rangle_i, \tag{3}$$

$$\text{subject to } D_j(w) \leq D_i^{max} \qquad \forall i \in R^I, j \in V_i', \tag{3.a}$$

$$G_i(w) \leq (1+s)^2 G_i(w^I) \qquad \forall i \in T, \tag{3.b}$$

$$D_i^{min} \leq D_j(w) \leq D_i^{max} \qquad \forall i \in T, j \in V_i', \tag{3.c}$$

$$y_i^\alpha + \frac{1}{1-\alpha} \sum_{j=1}^{|v_i|} p_{ji}^\alpha \leq M_{i\alpha}^{max}, \qquad \forall \alpha \in A_i, i \in R^{II}, j \in V_i', \tag{3.d}$$

$$D_j(w) - z_{ji} = 0, \qquad \forall i \in R^{II}, j \in V_i', \tag{3.e}$$

$$p_{ji}^\alpha \geq 0 \qquad \forall \alpha \in A_i, i \in R^{II}, j \in V_i', \tag{3.f}$$

$$p_{ji}^\alpha - z_{ji} + y_i^\alpha \geq 0 \qquad \forall \alpha \in A_i, i \in R^{II}, j \in V_i', \tag{3.g}$$

$$0 \leq w_k \qquad \forall k \in 1..N, \tag{3.h}$$

$$w_k \leq w^{max} \qquad \forall k \in 1..N \tag{3.i}$$

$$\tag{2.4}$$

The term $M_{i\alpha}^{max}$ in equation (3.d) is defined as $MOH\alpha$ for structure i , $\forall \alpha \in A_i, i \in R^{II}$.

2.6 Step IV

In the fourth and final step, the objective is to smooth the fluence map to avoid hot or cold spots without degrading the results of previous steps. This is accomplished by setting the results of previous steps as the constraints in this step and minimizing the sum of the square of the beam weights.

Let w^{III} be the beamlet weight found in the Step III of the optimization. The formula for this step is as follow:

Find w so as to

$$\text{minimize } F^{IV}(w) = \sum_{k=1}^{N} w_k^2 \tag{4}$$

$$\text{subject to } D_j(w) \leq D_i^{max} \qquad \forall i \in R^i, j \in V_i', \tag{4.a}$$

$$G_i(w) \leq (1+s)^3 G_i(w^I) \qquad \forall i \in T, \tag{4.b}$$

$$D_i^{min}(1-s_2) \leq D_j(w) \leq D_i^{max} \qquad \forall i \in T, j \in V_i', \tag{4.c}$$

$$y_i^\alpha + \frac{1}{1-\alpha} \sum_{j=1}^{|V_i'|} p_{ji}^\alpha \leq M_{i\alpha}^{max}, \qquad \forall \alpha \in A_i, i \in R^{II}, j \in V_i', \tag{4.d}$$

$$D_j(w) - z_{ji} = 0, \qquad \forall i \in R^{II}, j \in V_i', \tag{4.e}$$

$$p_{ji}^\alpha - z_{ji} + y_i^\alpha \geq 0 \qquad \forall \alpha \in A_i, i \in R^{II}, j \in V_i' \tag{4.f}$$

$$\langle D(w) \rangle_i \leq \langle D(w^{III} \rangle_i \qquad \forall i \in R^{III}, \tag{4.g}$$

$$0 \leq w_k \qquad \forall k \in \{1..N\}, \tag{4.h}$$

$$w_k \leq w^{max} \qquad \forall k \in \{1..N\} \tag{4.i}$$

$$\tag{2.5}$$

3 Theory

During the IMRT process, irradiation time is estimated by the number of monitor units (MU) that are needed. We call both of them IMRT complexity. There is a tradeoff between IMRT complexity and plan quality, such that one can reduce the complexity greatly while only slightly reducing the plan quality [3]. Also, reducing IMRT complexity can help to avoid the hot spot outside target tumors [1, 13].

Our theory is motivated by such research to consider another variable evaluating the number of MU needed and the smoothness of beam profile. The smoother the beam profile is, the lower the number of MU that are needed. It has been empirically shown that the complexity can be decreased to some extent. We believe it is reasonable to make an assumption that the beam profile that is needed clinically is somewhat smooth. Here we assume its smoothness is Lipschitz continuous. Theoretically, we consider each organ as a set of points and each voxel as a subset of it. Suppose there are N beam sources. Each beam consists of several beamlets in the same dirrection. For any voxel v in the organ, let di(v) be the dose received by point v from beam source i. Fix an organ below.

Assumption 1. *For each beam source i, its beam profile of the optimization solution is Lipschitz continuous with Lipschitz constant R_i. Let R be the maximum of these Lipschitz constants, then every beam profile is Lipschitz continuous with Lipschitz constant R.*

When a beamlet goes further in an organ, it is known that the dose will decrease rapidly and the dose distribution generated by the beamlet in the structure is infinitely smooth. We have our second assumption.

Assumption 2. *If every beamlet of beam source i has unit intensity, then they share the same dose distribution $e^{-a_i x}$ for a constant $a_i > 0$. Moreover, the dose distribution of each beamlet in beam source i is Lipschitz continuous with Lipschitz constant S_i. Let S be the maximum of these Lipschitz constants, then every beamlet is Lipschitz continuous with Lipschitz constant S.*

Remark. Here we assume the beamlets from one beam source 'hits' the organs after the same distance run.

A direct proposition from these two assumptions is:

Proposition 3. *For any beam source i, and any two points v and u in the same plane perpendicular to this beam direction, there exists $R_i > 0$, such that*

$$|d_i(v) - d_i(u)| \leq R_i|v - u|.$$

Proof. Let $g_i(v)$ and $g_i(u)$ be the intensity in the beam profile of the beamlet going through v and u respectively. Then $|g_i(v) - g_i(u)| \leq S_i|v - u|$ for some $S_i > 0$ by Assumption 1 (need ref). Also, $f_i(v) = g_i(v)e^{-a_i d}$ and $f_i(u) = g_i(u)e^{-a_i d}$, where d is the distance between beam source and the plane going through v, and u that is perpendicular to the beam direction. So $|f_i(v) - f_i(u)| \leq |g_i(v)e^{-a_i d} - g_i(u)e^{-a_i d}| = e^{-a_i d}|g_i(v) - g_i(u)| \leq e^{-a_i d}S_i$. Let $R_i = e^{-a_i d}S_i$, then we are done. □

For any beam source i and two points v and u in one organ, we can project the vector $\overrightarrow{vu}$ along the beamlet from beam source i going through v and the plane orthogonal to the beamlet. By Corollary 3, the difference of the dose received by v and u can be bounded.

Theorem 4. *If $d(v)$ and $d(u)$ are the dose received by points v and u from N equally spaced beam sources with beam weights of one IMRT optimization solution respectively, then $|d(v) - d(u)| \leq L|v - u|$ for a constant L, which is named the Sampling Lipschitz Constant (SLC).*

Proof. Let θ_i be the angle between the vector $\overrightarrow{vu}$ and the direction of beam source i. By Assumption 2 and Proposition 3,

$$|d_i(v) - d_i(u)| \leq S_i|v-u|cos\theta_i + R_i|v-u|sin\theta_i.$$

Summing up the doses from different beam sources,

$$\begin{aligned}
|d(v) - d(u)| =& |\sum_{i=1}^{N} d_i(v) - \sum_{i=1}^{N} d_i(u)| \\
\leq& \sum_{i=1}^{N} |d_i(v) - d_i(u)| \\
\leq& \sum_{i=1}^{N} (S_i|v-u|cos\theta_i + R_i|v-u|sin\theta_i) \\
\leq& \sum_{i=1}^{N} S|v-u|cos\theta_i + \sum_{i=1}^{N} R|v-u|sin\theta_i \\
=& |v-u|(S\sum_{i=1}^{N} cos\theta_i + R\sum_{i=1}^{N} sin\theta_i) \\
\leq& N(S+R)|v-u|,
\end{aligned}$$

where $S = max(S_1, S_2, \ldots, S_N)$, and $R = max(R_1, R_2, \ldots, R_N)$. Let $L = N(S+R)$. □

Suppose a target and several beams have SLC L. Then we have

$$|d(v) - d(u)| \leq L|v-u|.$$

If we sample $p\%$ of voxels in that organ, then the total volume of that target can be viewed as the sum of the volumes of each sampled voxel. In this way, we can estimate the distance between two neighboring voxels with the assumption that the voxels are evenly sampled. Specifically, the distance can be estimated by

$$2 * (\frac{256 * 256 * numSlice}{p * numVoxel})^{1/3}.$$

If $L * 2 * (\frac{256*256*numSlice}{p*numVoxel})^{1/3}$ is within 1% of $D95$ for the sampled voxels, then $D95$ for the sampled voxels is also within 1%of the $D95$ for all voxels.

4 Experimental Result

We ran the CBS and NS methods on nine different anonymized treatment plans: head and neck (7), prostate (1), and lung (1). The treatment plan has 256 × 256 voxels of size 0.2 × 0.2 cm, and a thickness of 0.5 cm in each slice. Figure 5.1 shows CERR's IMRTP configuration to create a treatment plan. We used seven equispaced beams that had 6MV of energy and default values for other parameters. We used the Mosek[9] solver to solve the optimization problem, as it can solve problems with a quadratic convex function and constraints. We ran the experiment in the engineering cloud cluster maintained by Washington University in Saint Louis, MO.

We divided the voxels into boundary voxels and inner voxels. Figures 5.2 shows the dose-volume histogram(DVH) of different treatment plans. Figures 5.4 to 5.7 show the dose distributions to different organs using a sampling rate of 10% to 100%. The D95 of target structures were around 3% worse than that of the NS method for the sampling rate of 10% . As seen in the DVH (Figures 5.2), the dose distribution to critical and normal organs obtained by the CBS method was always less than that obtained by the NS method. The difference in dose distribution decreased to zero as the sampling rate increased to 100%. We can see that the sampling rate of 20% produced a dose distribution in target structures that is comparable to that produced by the NS method, with a difference of less than 1.4%. Although the percentage difference in MoHx (x=5 for head and neck; x=50 for lung; and x=84 for prostate) to critical organs and the mean dose to normal organs was higher in Figures 5.6 and 5.7, the dose to these organs by CBS was less than that using the NS method. Figure 5.8 shows the running time of head and neck treatment plans. Compared with the NS method, a sampling rate of 20% resulted in the running time accelerated by approximately 5 to 35 times. Therefore, we estimated that a sampling rate of 20% produces a good dose distribution within a reasonable amount of time.

The dose distributions to D95 obtained with the CBS method (for sampling rate 20%) was worse than that obtained with the NS method. Figures 5.4 to 5.5 show the percentage by which D95 of the CBS method was worse than that of the NS method. The D95 to the highest-dose target structure and regional-field target structures obtained by the CBS method was less than 1%, and 1.4% worse than that obtained by the NS method,respectively. Figures 5.6 and 5.7 show the percentage by which MoHx to a critical organ and mean dose to a normal organ obtained from the CBS method (for sampling rate 20%) was different than that obtained from the NS method. The maximum improvement on MoHx was 25% for treatment plan HNTCP79, while the minimum improvement was 3% for the prostate treatment plan. Similarly,the mean dose to the normal organ was improved in the range of 2% to 35%.

We used CERR to subtract the dose distributions of the CBS (for sampling rate 20%) and NS methods. Figure 5.3 shows a screenshot taken from CERR's visualization of the subtracted dose. Although there was a small dose difference, CERR treated it as zero. Therefore, the dose distributions in the subtracted images are seen as zero.

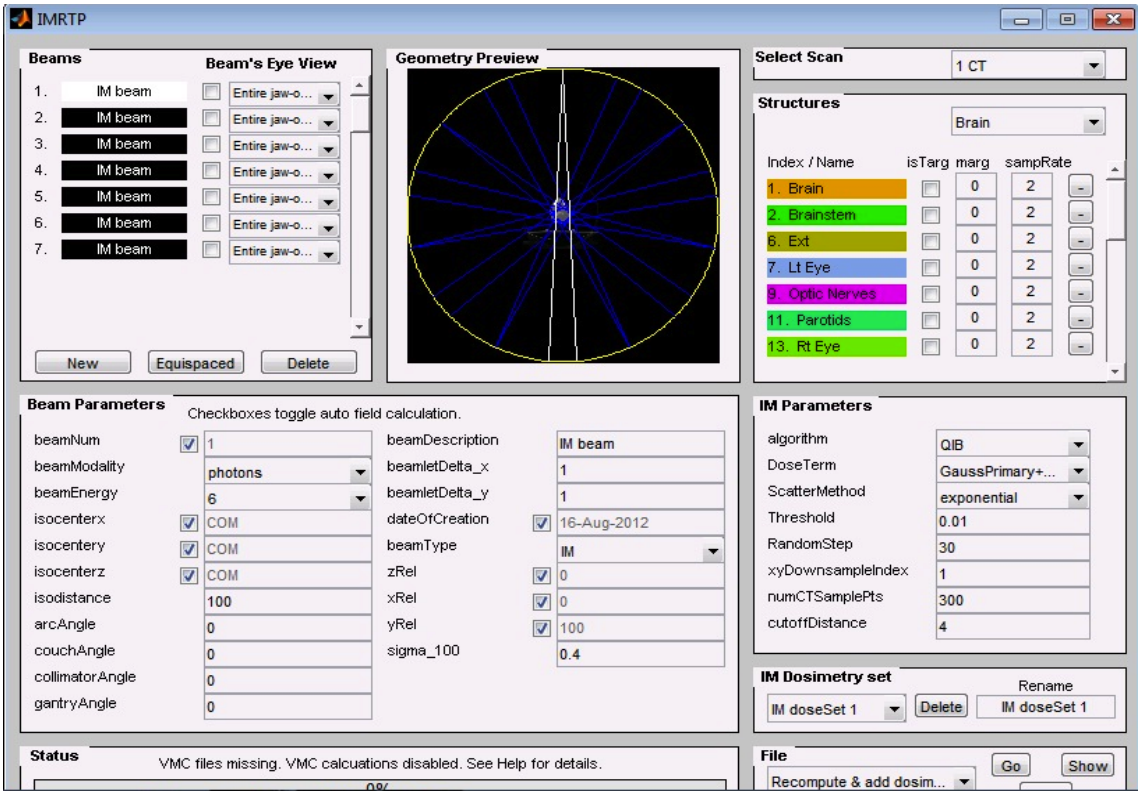
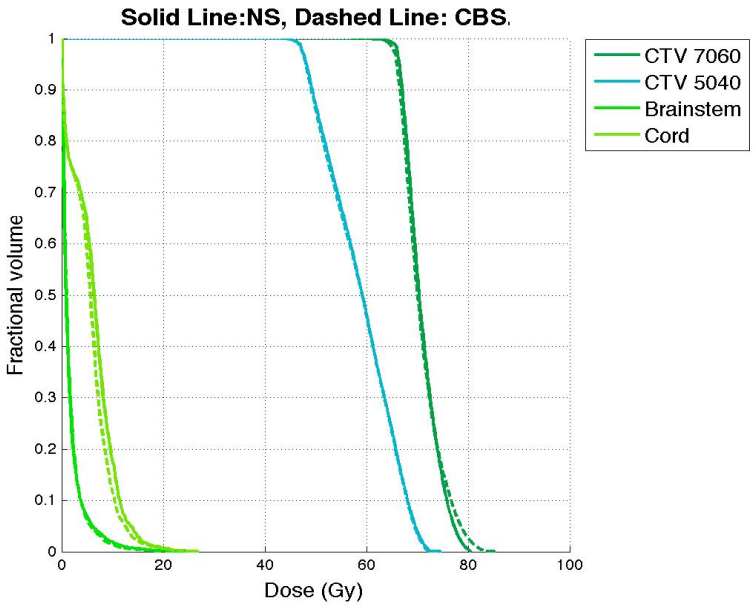
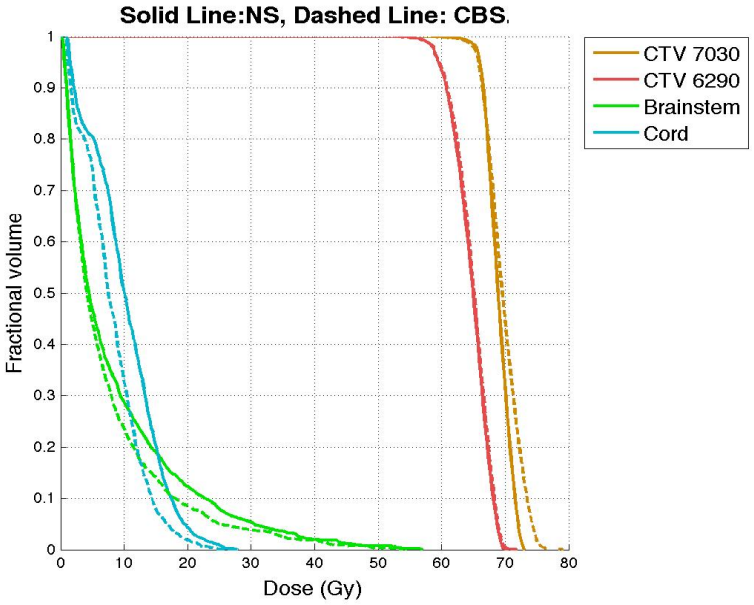


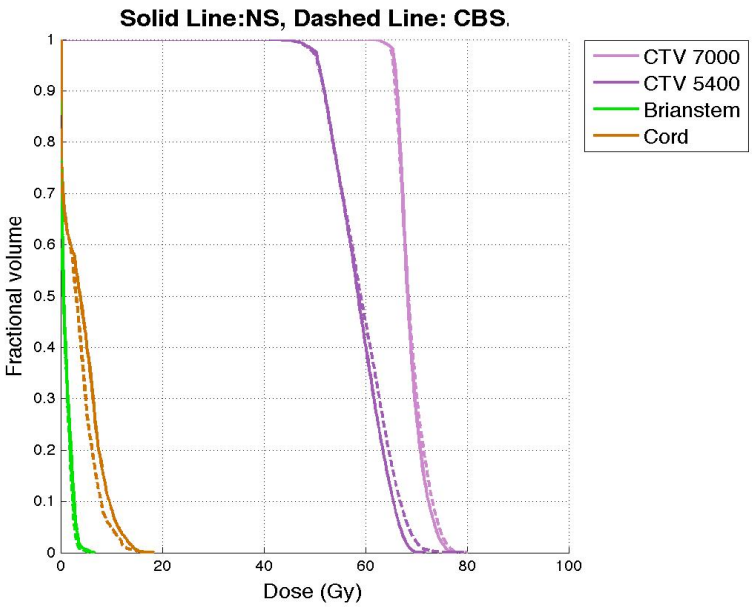
Figure 4.1: CERR configuration.



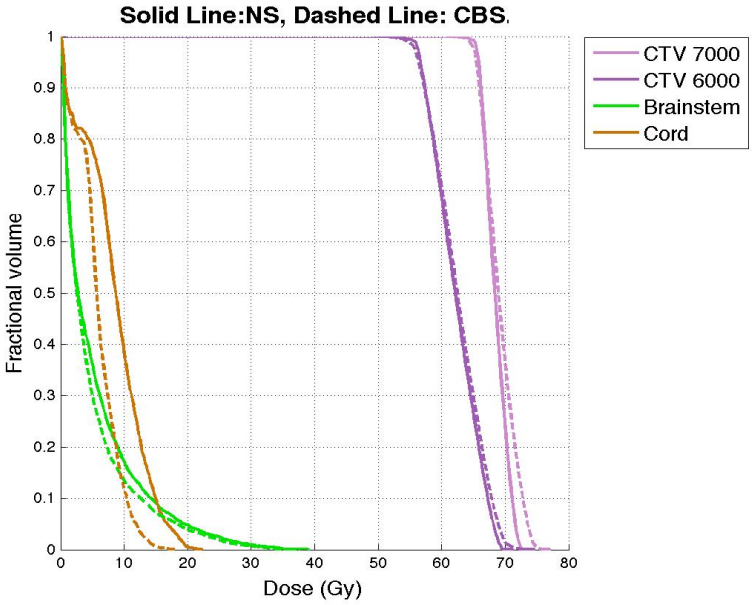
a) HNTCP5



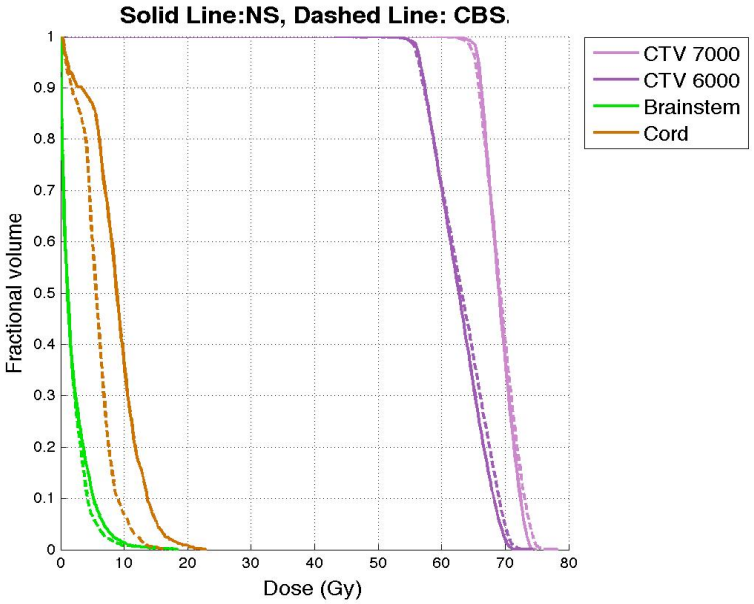
b) HNTCP25



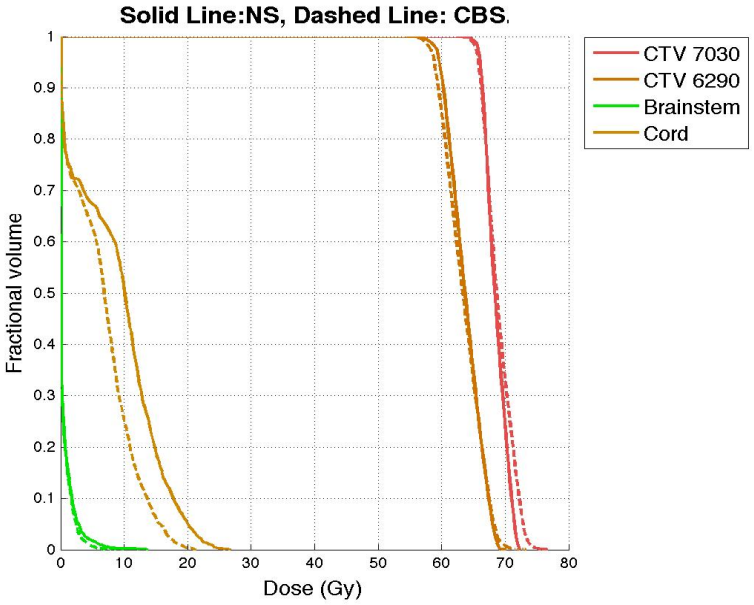
c) HNTCP104



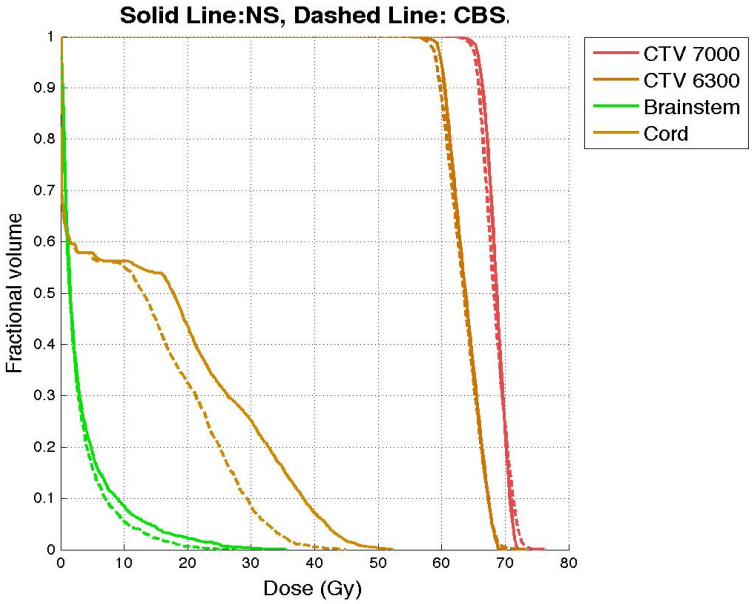
d) HNTCP108



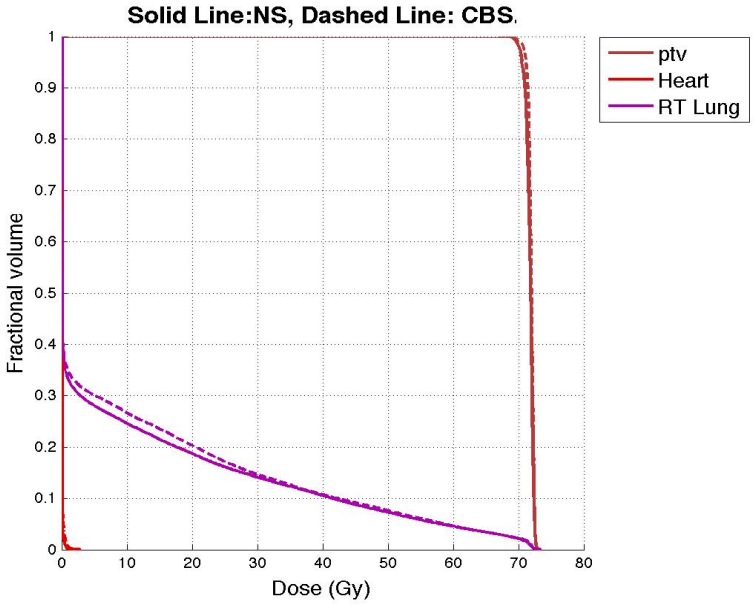
e) HNTCP113



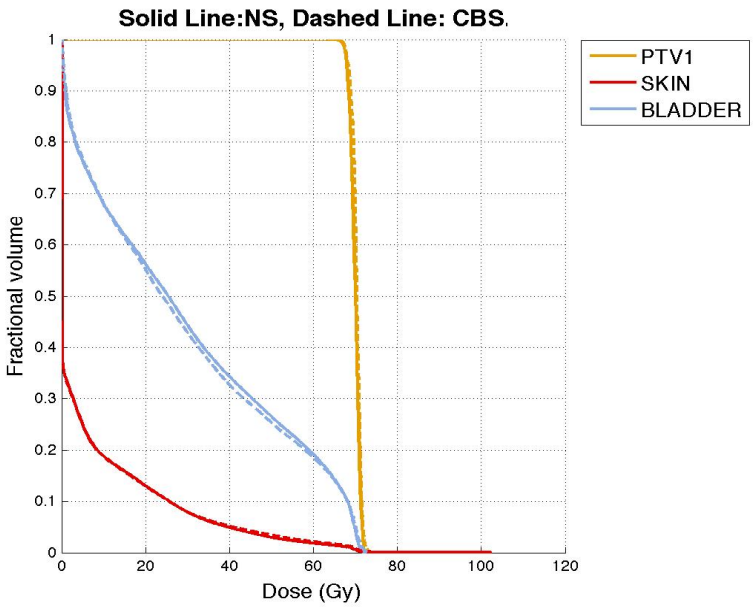
f) HNTCP30



g) HNTCP79



h) HNTCP1009



i) HNTCP3

Figure 4.2: The dose-volume histogram of different plans computed using the CBS and NS method.

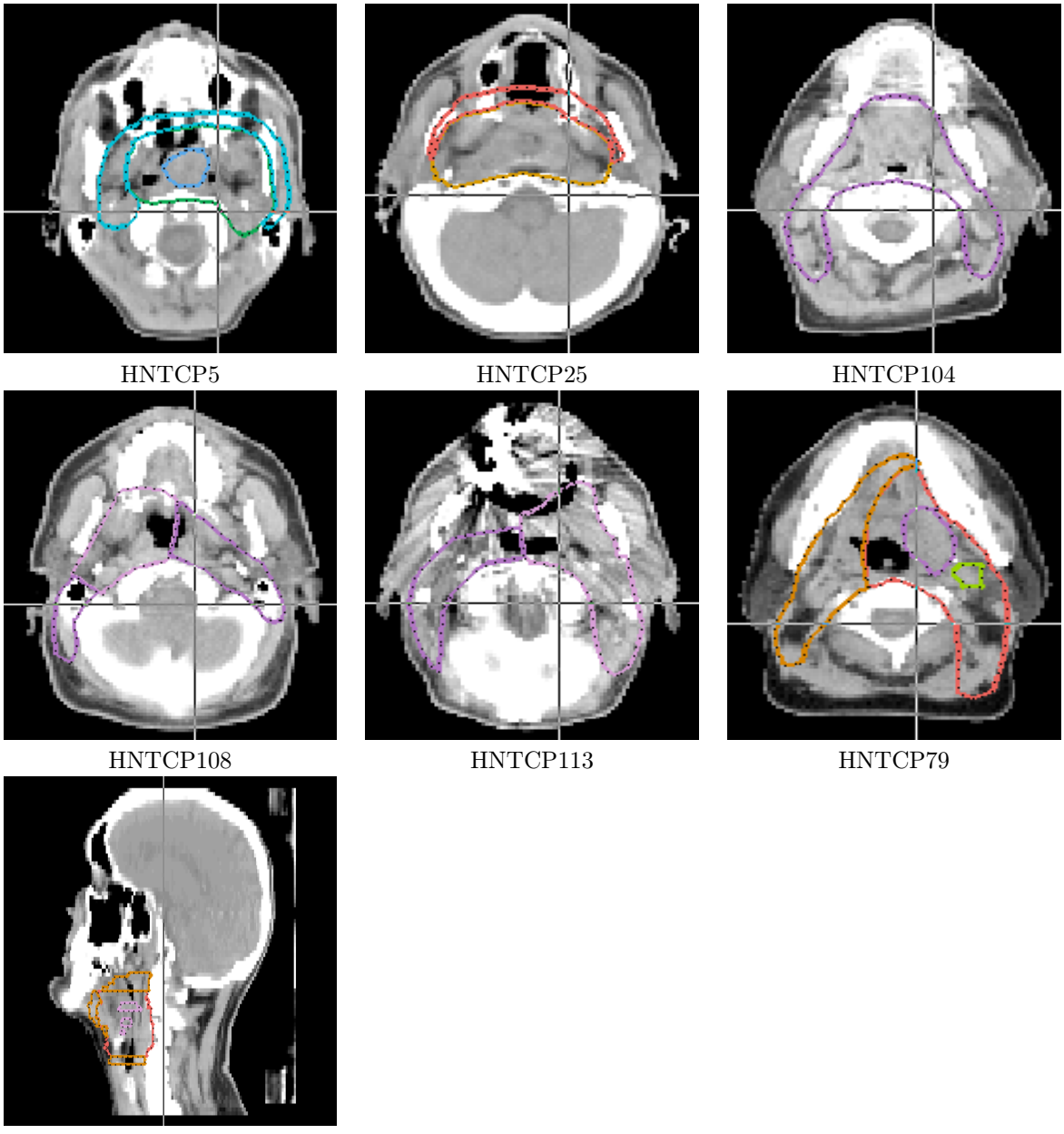


Figure 4.3: CERR screenshot of the dose obtained by subtracting the doses obtained from the CBS (sampling rate 20%) and NS methods.

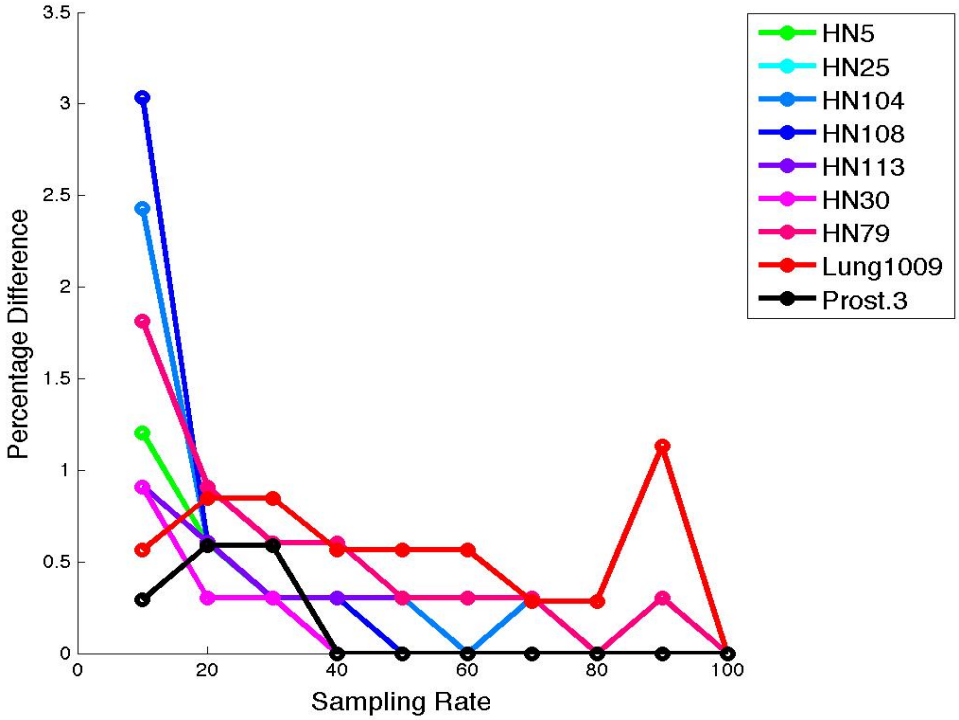


Figure 4.4: Percentage difference in D95 of the CBS method with different sampling rates, compared with the NS method for highest dose target structure.

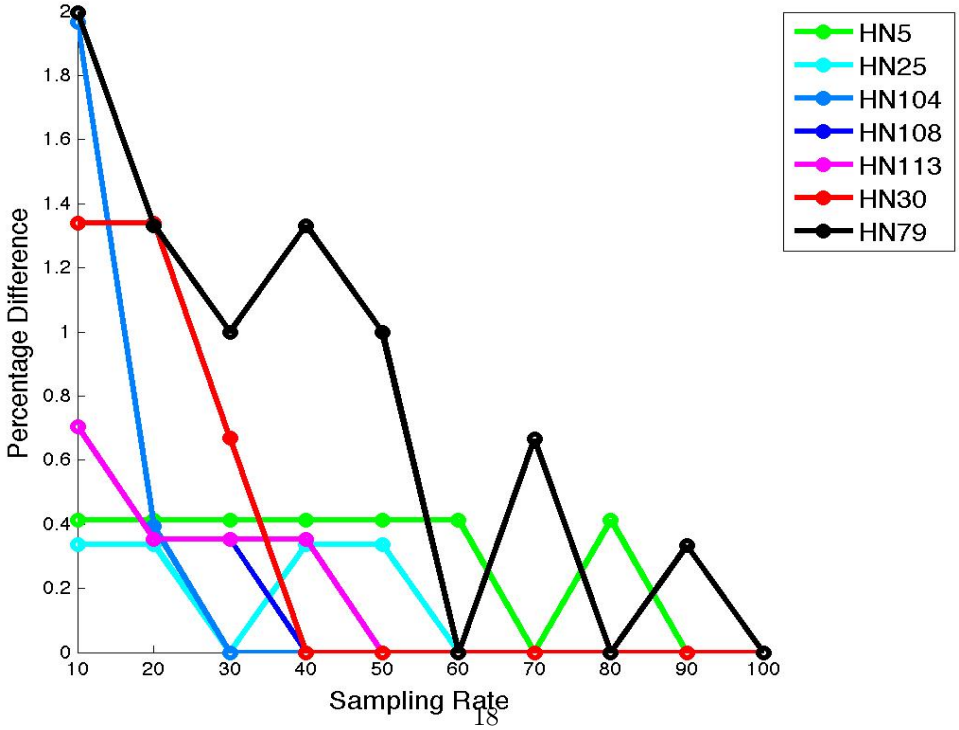


Figure 4.5: Percentage difference in D95 of the CBS method with different sampling rates, compared with the NS method for a regional-field target structure.

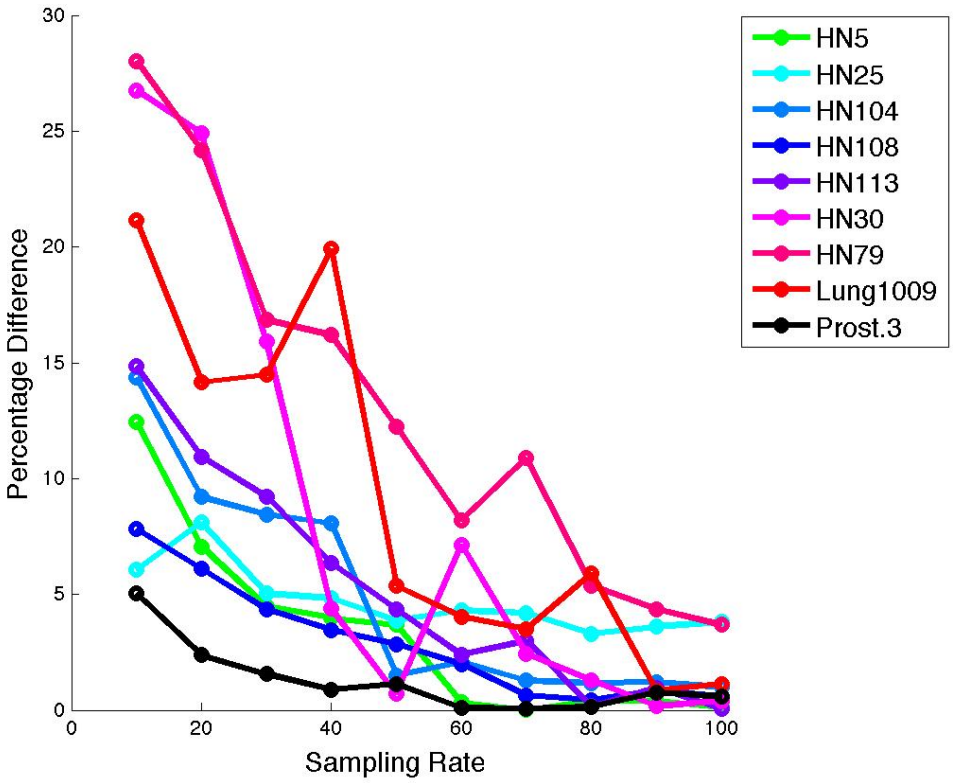


Figure 4.6: Percentage difference in Mohx to critical structures in the CBS method with different sampling rates, compared with the NS method. Critical structures are: for the head and neck case, brainstem, x=5; lung case, heart, x=50; and prostate case, rectum, x=84.

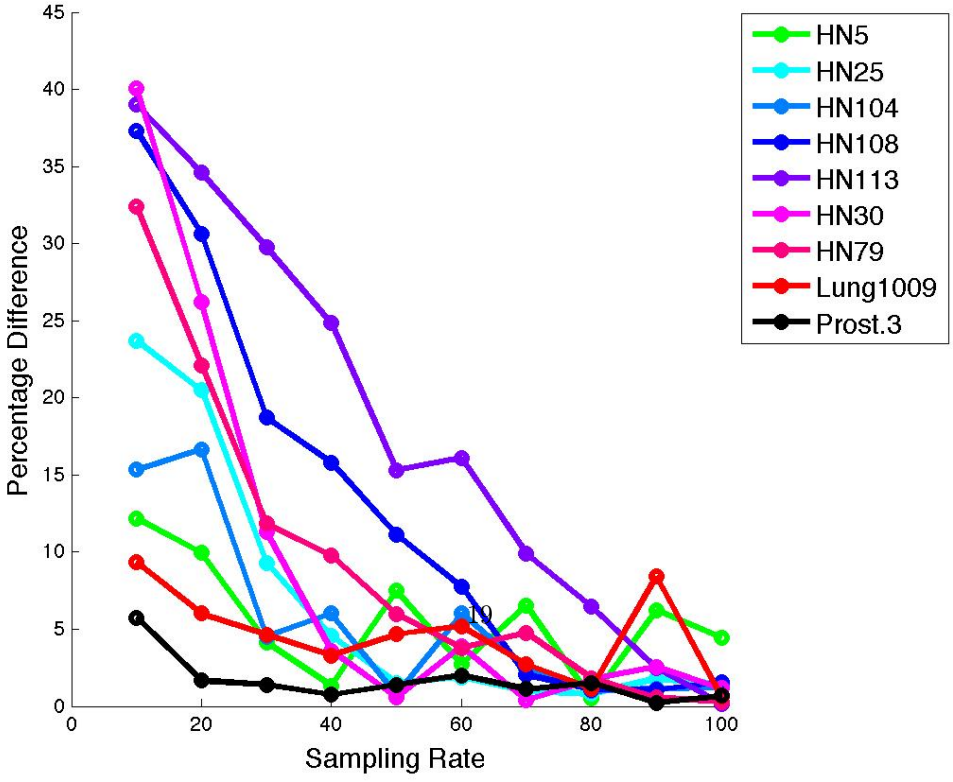


Figure 4.7: Percentage difference in mean dose to normal organs using the CBS method with different sampling rates, compared with the NS method. Normal organs are: for the head and neck case, spinal cord; lung case, right lung; prostate case, bladder.

4.1 Optimization Run Time and Resources used

Figures 5.8 and 5.9 show the time and memory used by the CBS and NS methods. As shown in Table 3, the CBS method was significantly faster than the NS method. Treatment plan HNTCP79 achieved the maximum improvement in running time, whereas the least improvement was seen in treatment plan HNTCP5. The CBS method was around 35 times faster than the NS method for treatment plan HNTCP79, and it was five times faster for treatment plan HNTCP5.

The memory used by the CBS method was less than that used by the NS method. As shown in the Figure 5.9, the NS method used around 1.5 to 1.8 times more memory than the CBS method.

Plan	CBS(Minutes)	NS(Minutes)	Speedup
5.0	76.2	443.1	5.8
25.0	33.0	698.5	21.2
104.0	30.7	232.9	7.6
108.0	37.2	374.5	10.1
113.0	158.5	2423.2	15.3
30.0	12.9	164.8	12.7
79.0	65.0	2296.1	35.3
1009.0	105.4	3498.5	33.2
3.0	14.6	234.2	16.0

Table 3: The acceleration of running time with the CBS method(for sampling rate 20%) compared with the NS method.

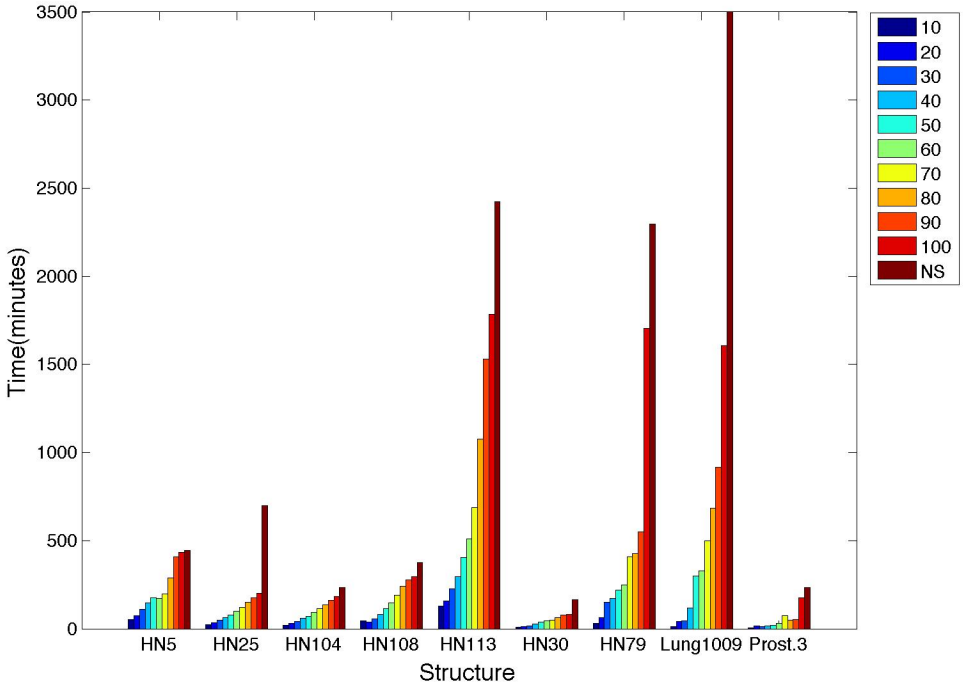


Figure 4.8: Running time with the CBS(for different sampling rates) and NS methods for head and neck treatment plans.

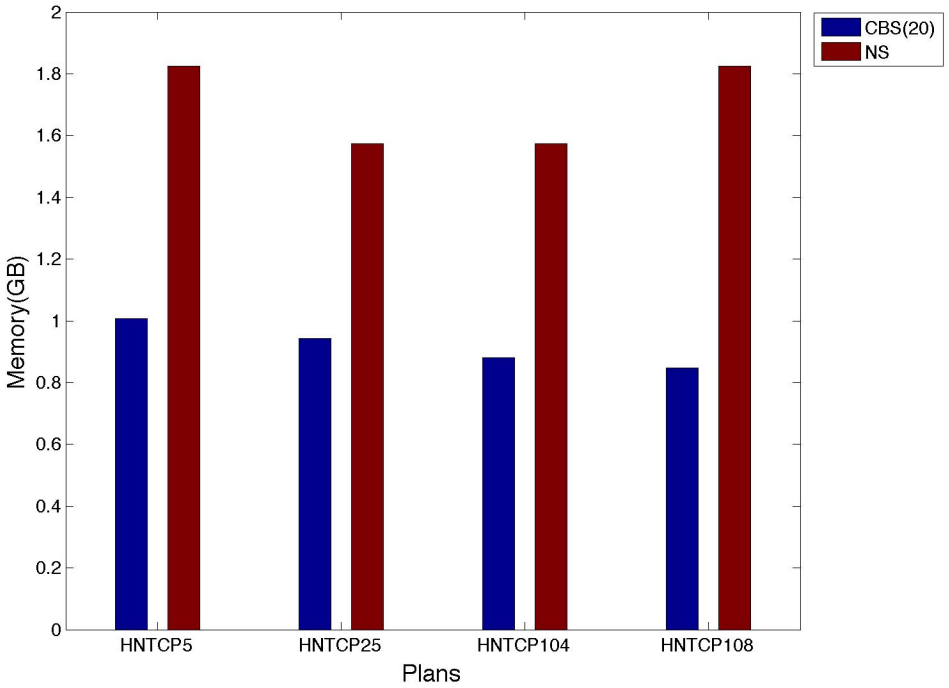


Figure 4.9: Memory used by the CBS and NS methods.

5 Discussion

This study shows that CBS could significantly reduce the time and memory used by IMRT optimization, without substantially degrading the dose distributions to organs. We tried to include voxels that play an important role in controlling dose distributions to organs, such as those on the organ boundary, through which beamlets must pass. In addition, we included approximately 20% of inner voxels sampled using grid sampling to get a more realistic representation of voxels.

The D95 of target structures obtained from the CBS method was slightly worse than that obtained from the NS method. This could be because CBS uses sampling to compute the dose distributions. Similarly, we found that the MoH of a critical organ and the mean dose to normal organs obtained from the CBS was lower than that obtained from the NS method. This could be attributed to the fact that if the dose distribution to target structures is degraded, the dose distribution to critical and other organs improves. Figures 5.4 to 5.7 show that the percentage difference of dose decreased sharply for lower sampling rates, and for most of the treatment plans, it decreased slowly as the sampling rate increased. This shows that grid sampling uniformly sampled the voxels from structures.

The CBS method has far fewer variables and constraints in IMRT optimization than the NS method. Therefore, the time and memory required by the CBS method was less than that required by the NS method. The lung treatment plans took more time than the head and neck and prostate treatment plans. The heart, rectum, and brainstem were used as critical organs in lung, prostate, and head and neck cases, respectively. The number of voxels in the heart was greater than the number of voxels in the brainstem and rectum. Typically, the number of voxels in the brainstem and rectum was less than 2000, whereas the number of voxels in the heart was approximately 20000. The formulation in Step II adds new artificial variables (p and z) and constraints equal in number to the voxels in a critical organ. So for lung case, we added approximately 40000 new variables and 20000 new constraints in Step II.Therefore, the lung treatment plan took more time than other plans. However, we significantly reduced the computational time using the CBS method.

There are different approaches proposed in the literature to reduce IMRT optimization time. Martin et. al.[8] proposed an approach of randomly selecting some voxels and using these voxels to calculate estimates of the objective function and gradient for a randomized steepest descent algorithm. This approach achieved acceleration at approximately an order of magnitude. Rocha et. al. [10] and Ferris et. al. [5] studied the impact of sampling in radiation therapy treatment design.Their study showed that the result of optimization using sampling could be as good as that of optimization without sampling, if we choose an adequate sampling rate. Lu et. al.[7] compared the performance of random and grid sampling to calculate dose distributions to organs. They found that the dose calculated using grid sampling was better than that calculated using random sampling. These results thus lend further credence to our earlier suggestion that sampling could reduce optimization time without deteriorating dose distributions to organs. Our sampling approach is different from other methods because we always include boundary voxels and approximately 20% of inner voxels sampled using grid sampling. Our approach leads to significant improvement in running time, compared to other sampling approaches, without substantially degrading the dose distribution to organs.

In our sampling approach, we fixed the sampling rate at 20% to select inner voxels (voxels excluding boundary voxels) because it did not severely impact dose quality. However, we believe that dose quality would be better if we use an adaptive sampling rate instead of the fixed one. We chose to include all boundary voxels in the optimization, since, as mentioned above, boundary voxels play an important role in controlling dose quality.

6 Related Work

Other groups have conducted research on voxel sampling and we compare our contribution to theirs in this section.

Ferris et al [5] used a voxel sampling approach for beam angle optimization. They used a simple adaptive algorithm for inter-organ sampling. For intra-organ sampling, they compared three different methods-grid sampling, uniform sampling, and rind sampling-and found only small differences among them. Rind sampling is similar to our boundary sampling, but the major difference is that we used sampling for inverse plan optimization rather than beam angle optimization. In addition, we showed both empirical and theoretical results of the CBS method.

Martin et al [8] presented an approach in which sampled voxels are used to estimate objective and gradient during a steepest descent algorithm, with the purpose of speeding up the process. Their approach saved 10 times the computational time saved by the steepest descent algorithm. However, since they estimate gradient by sampling, their approach can only work for separable objectives/constraints that can be expressed as a sum of certain terms for each voxel. In addition, their approach can only manage weighted-sum formulations,whereas the CBS method can work on any formulation, and a prioritized optimization is tested in this paper. Finally, their paper did not include theoretical bounds for their approach.

Other than accelerating optimization algorithms directly, a common way to save computational time is to reduce the size of the influence matrix. To this end, Thieke et al [14] used importance sampling to sample the voxels outside the beam penumbra.

There are also some approaches that attempt to speed up the IMRT process without sampling voxels. Scherrer et al [12] presented an adaptive clustering method to cluster voxels. A preprocessed hierarchy of aggregated dose is used as a basis for adapted approximations. The optimum solution can be obtained quickly. However, the initial hierarchy construction required by this method is complicated and is not suitable for robust optimization.

References

[1] M Alber and F Nüsslin. Intensity modulated photon beams subject to a minimal surface smoothing constraint. *Physics in Medicine and Biology*, 45(5):N49–N52, May 2000.

[2] V. H. Clark, Y. Chen, J. Wilkens, J. R. Alaly, K. Zakaryan, and J. O. Deasy. IMRT treatment planning for prostate cancer using prioritized prescription optimization and mean-tail-dose functions. *Linear algebra and its applications*, 428(5-6):1345–1364, March 2008. PMID: 18974791 PMCID: PMC2574493.

[3] David Craft, Philipp Süss, and Thomas Bortfeld. The tradeoff between treatment plan quality and required number of monitor units in intensity-modulated radiotherapy. *International journal of radiation oncology, biology, physics*, 67(5):1596–1605, April 2007. PMID: 17394954.

[4] Joseph O Deasy, Angel I Blanco, and Vanessa H Clark. CERR: a computational environment for radiotherapy research. *Medical physics*, 30(5):979–985, May 2003. PMID: 12773007.

[5] Michael C. Ferris, Rikhardur Einarsson, Ziping Jiang, and David Shepard. Sampling issues for optimization in radiotherapy. *Annals of Operations Research*, 148:95–115, October 2006.

[6] I. Lax and A. Brahme. *Rotation therapy using a novel high-gradient filter.* 1982.

[7] X.-Q. Lu and L. M. Chin. Sampling techniques for the evaluation of treatment plans. *Medical Physics*, 20(1):151–161, 1993.

[8] Benjamin C Martin, Thomas R Bortfeld, and David A Castanon. Accelerating IMRT optimization by voxel sampling. *Physics in Medicine and Biology*, 52:7211–7228, December 2007.

[9] Mosek. The MOSEK optimization toolbox for MATLAB.

[10] Humberto Rocha, Joana M. Dias, Brigida C. Ferreira, and Maria Carmo Lopes. Influence of sampling in radiation therapy treatment design. *Computational Science and Its Applications - ICCSA 2011*, 6784:215–230, 2011.

[11] H. Edwin Romeijn, Ravindra K. Ahuja, James F. Dempsey, and Arvind Kumar. A new linear programming approach to radiation therapy treatment planning problems. *Operations Research*, 54(2):201–216, March 2006.

[12] Alexander Scherrer, Karl-Heinz Küfer, Thomas Bortfeld, Michael Monz, and Fernando Alonso. IMRT planning on adaptive volume structures—a decisive reduction in computational complexity. *Physics in Medicine and Biology*, 50:2033–2053, May 2005.

[13] Xuepeng Sun and Ping Xia. A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning. *Medical physics*, 31(5):1158–1165, May 2004. PMID: 15191305.

[14] Christian Thieke, Simeon Nill, Uwe Oelfke, and Thomas Bortfeld. Acceleration of intensity-modulated radiotherapy dose calculation by importance sampling of the calculation matrices. *Medical Physics*, 29:676, 2002.

[15] Jan J Wilkens, James R Alaly, Konstantin Zakarian, Wade L Thorstad, and Joseph O Deasy. IMRT treatment planning based on prioritizing prescription goals. *Physics in medicine and biology*, 52(6):1675–1692, March 2007. PMID: 17327656.