Optimization of intensity modulated beams with volume constraints using two methods: Cost function minimization and projections onto convex sets

Paul S. Cho[a)]
Department of Radiation Oncology, University of Washington, Seattle, Washington 98195-6043

Shinhak Lee
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

Robert J. Marks II
Department of Electrical Engineering, University of Washington, Seattle, Washington 98195-2500

Seho Oh
NeoPath Inc., Redmond, Washington 98052

Steve G. Sutlief and Mark H. Phillips
Department of Radiation Oncology, University of Washington, Seattle, Washington 98195-6043

(Received 6 February 1997; accepted for publication 27 January 1998)

For accurate prediction of normal tissue tolerance, it is important that the volumetric information of dose distribution be considered. However, in dosimetric optimization of intensity modulated beams, the dose–volume factor is usually neglected. In this paper we describe two methods of volume-dependent optimization for intensity modulated beams such as those generated by computer-controlled multileaf collimators. The first method uses a volume sensitive penalty function in which fast simulated annealing is used for cost function minimization (CFM). The second technique is based on the theory of projections onto convex sets (POCS) in which the dose-volume constraint is replaced by a limit on integral dose. The ability of the methods to respect the dose–volume relationship was demonstrated by using a prostate example involving partial volume constraints to the bladder and the rectum. The volume sensitive penalty function used in the CFM method can be easily adopted by existing optimization programs. The convex projection method can find solutions in much shorter time with minimal user interaction. © *1998 American Association of Physicists in Medicine.* [S0094-2405(98)01004-9]

Key words: beam intensity modulation, optimization, dose–volume constraint, convex set, cost function

I. INTRODUCTION

Optimization of intensity modulated beams can be designed to satisfy either dosimetric prescriptions or biological response specifications. While biological models can more directly measure clinical outcome, they still remain in the formative stages.[1] On the other hand, prescription in dosimetric unit has been well established as the clinical norm. Commonly, dosimetric strategies seek to deliver a high, uniform dose to the target while maintaining the organ dose to below tolerance. However, organ tolerance is more accurately assessed by the volumetric information of dose distributions.[2,3] Importance of the volume effect has been recognized and quantified as early as 1948 by Paterson, who measured the effect of volume on normal skin response to radiation.[4] In 1991, volume-dependent normal tissue tolerance data for various organs were tabulated and published by a task force under NCI contract.[5] The results of investigation for nearly 30 different critical tissues were categorized into three volume classes: 1/3, 2/3, and whole. For each partial volume, two dose levels, TD55/5 (the probability of 5% complication within 5 years) TD50/5 (the probability of 50% complication within 5 years), were reported. Reliability of partial volume data is expected to improve as more clinical results become available.

While much work has been done in the area of optimization of intensity modulation,[6] physical dosimetric optimization subject to volume restrictions has received little attention in the past. Langer *et al.* proposed a method using a mixed integer linear programming which was applied to compute beam weights of wedged and open fields.[7,8] However, the size of beam optimization problems which may be handled by this method and the applicability to general intensity modulation problem are unknown. More recently, Bortfeld *et al.* suggested a method of dose–volume constraint using a penalty function which is active only when the organ dose is within a certain interval.[9] Carol *et al.* reported the concept of an area cost function in which the cumulative dose volume histogram is subdivided into regions with different weights. The total cost is a weighted sum of the costs incurred in each region.[10]

In this paper we describe two methods of dosimetric optimization of intensity modulation with volume restrictions. The first method focuses on a formulation of a volume-

sensitive penalty function in which fast simulated annealing is used for cost function minimization (CFM). The second, the method of projections onto convex sets (POCS), does not use an objective function. Instead, the constraints are expressed in convex sets. Convergence to a solution is accomplished by alternately projecting between these convex sets. We apply these algorithms to the problem of intensity modulated beam optimization under dose–volume constraints.

II. METHODS

A simplistic strategy for imposing the dose–volume restrictions would be to partition the organ according to the dose–volume requirements and assign appropriate upper bounds to each segment. For instance, if the volume distribution requires that no part of an organ receive a dose higher than 50 Gy and that no more than 30% receive a dose greater than 40 Gy, then the organ can be divided into 30/70% segments with upper bounds set to 40 and 50 Gy, respectively. In this way, the problem is reduced to the case of multiple organs with different upper dose limits. After prepartitioning of the organ, any of the existing inverse methods can be applied to achieve conformal plan optimization with dose–volume constraints. The major difficulty with this approach is that arbitrary partitioning may not guarantee the best results. For instance, the region closer to the target volume will most likely receive a higher dose and, therefore, partitioning should be carried out to assign the higher of the two upper bounds to these areas. However, it may be difficult to approximate the optimal dose distribution within the organ *a priori*, especially for complex beam-anatomy configurations. Therefore, for this method to be effective, it would be necessary to perform optimization over all feasible organ volume partitions. This would be a cumbersome task. More practical methods are described in the following.

A. Volume-sensitive cost function minimization

Consider an organ at risk which requires that no part receive a dose greater than E_1 and that no more than $\gamma \times 100\%$ of volume receive a dose exceeding E_2. N represents the total number of uniformly sampled points over the volume. During the iterative optimization process, a dose D_k is calculated for each sample point. From these doses the fraction of the organ receiving a dose greater than E_2, denoted by γ', is determined. Then a volume-sensitive constraint in discrete form may be written as

$$P = \lambda V \sum_{k=1}^{N} \Psi_k, \tag{1}$$

where

$$\Psi_k = \begin{cases} (D_k - E_1)^2 + E_1, & D_k \geqslant E_1, \\ D_k, & D_k < E_1, \end{cases} \tag{2}$$

and

$$V = \begin{cases} \left(\dfrac{\gamma'}{\gamma}\right)^2, & \gamma' > \gamma \\ 1, & \gamma' \leqslant \gamma \end{cases}. \tag{3}$$

The volume-dependent term V has the value of unity when $\gamma' \leqslant \gamma$. Thus there is no additional penalty on organ irradiation when the volume constraint is met. However, the base penalty given by Eq. (2) is amplified as the fraction of volume elevated to dose above the limit E_2 begins to exceed the specified volume γ. The penalty function P is further weighted by the multiplier λ, which gives the ability to fine tune the constraint. The total cost function is formulated in the usual way[11] by summing the target objective and the penalty functions:

$$F = \frac{1}{N_T} \sum_{i=1}^{N_T} (D_i - TD)^2 + \sum_{j=1}^{M} \frac{P_j}{N_j}, \tag{4}$$

where TD is the prescribed dose to the target and D_i is the dose at a target sample point i. P_1 through P_M are the dose–volume constraints for M normal tissue regions given by Eq. (1). Each term is normalized to its respective volume which is proportional to the number of uniformly distributed sample points, $N_T, N_1, \ldots N_M$.

Minimization of the cost function was performed by the simulated annealing method.[12,13] The stochastic technique has the ability to escape the local minima solutions and therefore is suitable when the exact behavior of the cost function is unknown. Vectors representing changes in beam strength were generated with Cauchy probability density.[14] The n dimensional Cauchy distribution was sampled in spherical coordinates according to the scheme suggested by Mageras and Mohan.[15] The width parameter of distribution $W(l)$ at lth iteration was reduced according to $W(l) = W_0/(1 + l/R_w)$, where R_w is a parameter controlling the rate of collapse. Probability of accepting an uphill change in cost, $\exp(-\Delta F/T)$, was reduced by changing T after every nth iteration according to $T(l) = T_0/(1 + l/R_t)$ where R_t is a control parameter.

For the purpose of systematic tuning of the annealing parameters, it is useful to first determine the feasible range of the W_0. A reasonable initial value was obtained by choosing W_0 so that the peak of the joint Cauchy distribution corresponded to a beam strength step size in which the dose at the isocenter changed by a fraction of the maximum dose. We use the formula,

$$W_0 = \frac{0.5}{TPD} \sqrt{\frac{2n}{n-1}}, \tag{5}$$

where TPD is the total dose contributed from all the pencil beams at the isocenter and 0.5 means that the desired initial step size is 50% of the target prescription. With the W_0 selected, the initial temperature was determined after several trial runs while keeping the control parameters R_w and R_t fixed. Initially, R_w and R_t were assigned relatively large values and large iteration numbers were used to ensure convergence. Afterward, the parameters were adjusted to speed up the annealing process. The actual annealing parameters used in this study are tabulated in Table I.

TABLE I. Annealing parameters and values.

	Trial 1	Trial 2	Trial 3
W_0	0.02	0.02	0.02
T_0	0.2	0.2	0.2
R_w	1000	1000	1000
R_t	2000	2000	2000
λ_{bladder}	3.5	6.0	20.0
λ_{rectum}	3.0	9.0	25.0

B. Projections onto convex sets

The theory of convex projections developed by Bregman[16] and Gubin *et al.*[17] has been applied to such topics as sampling theory,[18] signal recovery,[19] artificial neural networks,[20] image processing,[21] and image restoration.[22] For biomedical applications, POCS has been applied to the problem of incomplete projection data in computerized tomography.[23] Recently, the mathematical foundation of POCS in radiotherapy applications has been presented.[24] For the specific case of intersecting linear varieties (hyperplanes), POCS is equivalent to Kaczmarz's alternating projection theorem.[25]

A set of vectors is said to be convex if a linear combination of any two vectors also lies within the set. More precisely, the definition states that a set C is convex if for every $\vec{x}_1 \in C$ and every $\vec{x}_2 \in C$, $\alpha\vec{x}_1 + (1-\alpha)\vec{x}_2 \in C$ for all $0 \leqslant \alpha \leqslant 1$. In other words, the line segment connecting $\vec{x}_1$ and $\vec{x}_2$ is totally enclosed in C. A useful property of POCS is that successive projections between two or more convex sets with nonempty intersection will yield a solution that satisfies all the sets involved. The convex projection method differs from other iterative optimization methods such as steepest descent,[26] quasi-Newton,[27] iterative filtered backprojection,[28] and conjugate gradient[29] in that it does not form a total cost function to be minimized. Instead, convergence to a solution is accomplished through the process of orthogonal projections whose directions are determined by minimizing the Euclidean distance between the convex constraint sets. The projection operation is analogous to taking the gradient in the cost function based iterative methods.

The framework of POCS theory applied in radiation therapy planning has been described previously,[24] which we summarize here. We begin by defining the dose vector domain in which the convex formulation is made. Assume that there are Q beams, each beam discretized to N beam elements. Let the weights of the beam elements from the kth beam be written as a column vector, $\vec{b}_k = [b_{k1}\ b_{k2} \cdots b_{kN}]^T$. For M sample points, the dose delivered to each point by the kth beam array is given by

$$\begin{bmatrix} d_k(1) \\ \vdots \\ d_k(M) \end{bmatrix} = \begin{bmatrix} a_{k1}^1 & \cdots & a_{kN}^1 \\ \vdots & \ddots & \vdots \\ a_{k1}^M & \cdots & a_{kN}^M \end{bmatrix} \begin{bmatrix} b_{k1} \\ \vdots \\ b_{kN} \end{bmatrix}, \tag{6}$$

where a_{kn}^m is the dose contribution per unit weight from the nth beam element in the kth beam array b_{kn} to the point m. Equation (6) is simplified by defining the dose computation matrix $\mathbf{A}_k$ and dose vector $\vec{d}_k$ as

$$\vec{d}_k = \mathbf{A}_k \vec{b}_k . \tag{7}$$

Then the total dose vector from Q beams is given by

$$\vec{t} = \begin{bmatrix} t(1) \\ \vdots \\ t(M) \end{bmatrix} = \sum_{k=1}^{Q} \vec{d}_k = \sum_{k=1}^{Q} \mathbf{A}_k \vec{b}_k . \tag{8}$$

Now, the dose vector domain in which the constraints are formulated is formed by concatenating the individual dose vectors $\vec{d}_k$ for Q beams,

$$\vec{d} = \begin{bmatrix} d_1(1) & \cdots & d_Q(1) \\ \vdots & \ddots & \vdots \\ d_1(M) & \cdots & d_Q(M) \end{bmatrix} = [\vec{d}_1^T \vec{d}_2^T \cdots \vec{d}_Q^T]^T . \tag{9}$$

A number of useful constraints can be constructed in this domain. To compute the incident photon fluence modulation under dose-volume requirements, the following constraint sets are used.

1. Normal tissue dose–volume constraint set

First, we consider the organ dose constraint without volume restrictions and assume that the organ dose is simply required to lie below a certain maximum. It is easy to show that in this case the set of doses that lie between 0 and the upper bound is convex. Proof is made by applying the convexity condition stated above. According to the definition of convex set, any linear combination of (positive) doses less than E is also less than E. For instance, if $E = 40$ Gy, $D_1 = 39$ Gy, and $D_2 = 38$ Gy, then $\alpha D_1 + (1-\alpha)D_2 < 40$ Gy for all $0 \leqslant \alpha \leqslant 1$. This is true for any other $D_1 < E$ and $D_2 < E$. Therefore, the organ dose minimization using a single upper bound is a convex problem.

Now, the dose–volume constraints require that the organ be subject to multiple dose boundaries. In addition to the upper limit E_1, a portion of the volume must not receive doses greater than a secondary upper limit E_2 where $0 < E_2 < E_1$. As shown above, the doses that are bound between 0 and E_1 do form a convex set. However, in order for a certain fraction of these doses to remain below E_2 so as to satisfy the volume demand, the following must hold true: $\alpha D_1 + (1-\alpha)D_2 < E_2 < E_1$. This condition is guaranteed only when $E_2 = E_1$, in which case, the constraint is reduced to the simple case of a single upper limit with no partial volume requirement. Therefore, the dose–volume constraints cannot be expressed in a single convex set.

We circumvent the problem by using two separate convex sets. The first set constrains the maximum dose to the whole organ while the second limits the integral dose. The first pushes the dose–volume histograms toward the origin along the dose axis and the other along the volume axis. When applied to the same volume, these two constraints provide

the ability to manipulate the dose–volume relationship. The constraint which determines the maximum whole organ dose is convex and is given by

$$C_{O1}=\left\{ \vec{d} \middle| 0 \leqslant \sum_{k=1}^{Q} d_k(m) \leqslant E_1 \right\}, \quad m \in S, \tag{10}$$

where S denotes a set of indices corresponding to the sample points located within the organ. The set dictates that each dose sample point remains below E_1. The constraint which limits the integral organ dose is also convex:

$$C_{O2}=\left\{ \vec{d} \middle| 0 \leqslant \sum_{\text{organ}} \sum_{k=1}^{Q} \vec{d}_k \leqslant I(i) \right\}, \tag{11}$$

where the outer summation integrates all the dose vectors within the organ volume. Note that the integral dose limiting parameter I is varied with iteration i. Suppose the fractional volume, which is required to remain below the dose limit E_2, is not compliant, and instead, its upper limit is found to be E_2' where $E_2' > E_2$. Then I is reduced in proportion to the difference between E_2' and E_2. Otherwise, I is unchanged. E_2' is updated at each iteration. Projection operations onto C_{01} and C_{02} force the dose vector to conform to these sets. Details are given in Sec. II B 5.

2. Target dose prescription set

This constraint set requires that the dose in the target volume be the same as the prescription. The set of dose vectors within the target volume whose sum forms the prescribed dose TD is convex:

$$C_T=\left\{ \vec{d} \middle| \sum_{k=1}^{Q} \vec{d}_k = TD \right\}. \tag{12}$$

Projection onto the target dose constraint set has the effect of minimizing the deviation from the prescription. The projection operator for the kth component is

$$P_T(\vec{d}_k)=\vec{d}_k+\frac{1}{Q}\left(TD-\sum_{k=1}^{Q} \vec{d}_k \right). \tag{13}$$

3. Beam-dose relationship set

The set of dose distributions that can be generated from different combinations of beam intensities is convex and is given by

$$C_B=\{\vec{d} | \vec{d}_k = \mathbf{A}_k \vec{b}_k\}, \quad 1 \leqslant k \leqslant Q. \tag{14}$$

Any dose vector estimates must be a member of this set. This is accomplished by projecting $\vec{d}_k$ onto C_B. As a rule, projection must occur between sets existing in the same domain, in this case, the dose vector domain. This necessitates that the beam vector $\vec{b}_k$ be expressed in terms of dose vector quantity. However, the beam vector cannot be obtained through a simple inversion of $\mathbf{A}_k$. The difficulty arises from the fact that, generally, the dimension of the dose vector exceeds that of the beam elements. In other words, the matrix is not square and, therefore, not invertible. The problem is circumvented by using the pseudo-inverse method[30] to find $\vec{b}_k$,

$$\vec{b}_k=(\mathbf{A}_k^T\mathbf{A}_k)^{-1}\mathbf{A}_k^T\vec{d}_k. \tag{15}$$

Then the beam-dose transformation is established through the projection operator,

$$P_B(\vec{d}_k)=\mathbf{A}_k(\mathbf{A}_k^T\mathbf{A}_k)^{-1}\mathbf{A}_k^T\vec{d}_k. \tag{16}$$

4. Non-negativity constraint set

Negative beam fluence is not physically possible. The non-negativity constraint set requires beam weight to be non-negative. The corresponding convex set is defined as

$$C_+=\{\vec{d} | \vec{d}_k = \mathbf{A}_k \vec{b}_k, \quad \text{each element of } \vec{b}_k \geqslant 0\}, \quad 1 \leqslant k \leqslant Q. \tag{17}$$

The projection onto C_+ is

$$P_+(\vec{d}_k)=\begin{cases} \mathbf{A}_k'(\mathbf{A}_k'^T\mathbf{A}_k')^{-1}\mathbf{A}_k'^T\vec{d}_k, & b_{kn}<0, \\ d_k, \quad \text{otherwise}, \end{cases} \tag{18}$$

where the matrix $\mathbf{A}_k'$ is formed for negative beam weights. For instance, if the nth element of the kth beam array b_{kn} is negative, the matrix $\mathbf{A}_k'$ is formed by removing the nth column of $\mathbf{A}_k$. Thus the projection effectively sets negative beam weights to zero.

5. Iterative projections

By iteratively projecting among these five convex sets, the beam weights that satisfy the dose–volume constraints and target objectives are determined. For a given dose vector $\vec{d}$, which does not belong to a constraint set C, the projection onto C is given by the unique vector $\vec{p} \in C$ such that the distance between $\vec{d}$ and $\vec{p}$ is minimum. For instance, the projection for the integral organ dose limit is formed by computing the distance J,

$$\begin{aligned} J &= \frac{1}{2}\sum_{k=1}^{Q} \|\vec{p}_k-\vec{d}_k\|^2+\lambda\left(\sum_{k=1}^{Q} \vec{r}^T\vec{p}_k - I \right) \\ &= \frac{1}{2}\sum_{k=1}^{Q} \|\mathbf{A}_k\vec{b}_k-\vec{d}_k\|^2+\lambda\left(\sum_{k=1}^{Q} \vec{r}^T\mathbf{A}_k\vec{b}_k - I \right), \end{aligned} \tag{19}$$

and minimizing it.

$$\frac{\partial J}{\partial \vec{b}_k}=\mathbf{A}_k^T(\mathbf{A}_k\vec{b}_k-\vec{d}_k)+\lambda\mathbf{A}_k^T\vec{r}=0, \tag{20}$$

where $\vec{r}$ is a structure discriminator whose elements are unity if inside the organ and zero if outside. After solving and substituting for λ, Eq. (20) is solved for $\vec{b}_k$. The projection vector $\vec{p}$ is found by transforming $\vec{b}_k$ back to the dose space:

$$\vec{p}_k=P_{O2}(\vec{d}_k)=\begin{cases} \mathbf{T}_k\left(\vec{d}_k+\dfrac{E_2-u}{\nu}\vec{r} \right), & u>I, \\ \mathbf{T}_k\left(\vec{d}_k+\dfrac{u}{\nu}\vec{r} \right), & u<0, \\ \mathbf{T}_k\vec{d}_k, \quad \text{otherwise}, \end{cases} \tag{21}$$

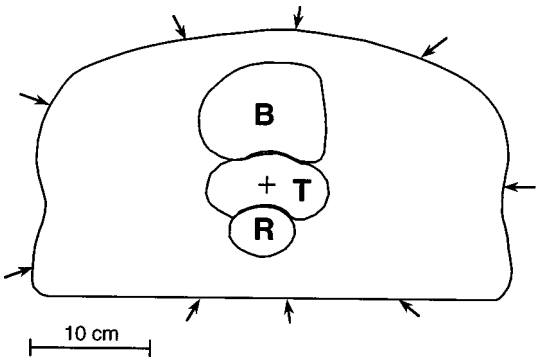

FIG. 1. External and organ contours for bladder (B), rectum (R), and prostate (T) used to demonstrate the algorithms. The arrows indicate the central axis of the beams directed at the isocenter (+).

where

$$\mathbf{T}_k = \mathbf{A}_k(\mathbf{A}_k^T\mathbf{A}_k)^{-1}\mathbf{A}_k^T,$$

$$u = \sum_{k=1}^{Q} \vec{r}^T\mathbf{T}_k\vec{d}_k, \tag{22}$$

$$\nu = \sum_{k=1}^{Q} \vec{r}^T\mathbf{T}_k\vec{r}.$$

Likewise, the projection for the other constraints are formed. The POCS iteration proceeds as follows:

$$\vec{d}^{\,i+1} = P_+(P_B(P_T(P_{O2}(P_{O1}(\vec{d}^{\,i}))))), \tag{23}$$

where i is the iteration counter. The iterative process is terminated when improvement in the mean-square distance between projections becomes sufficiently small.

C. Application to optimization of beam intensity modulation

The proposed algorithms were applied to intensity modulated beam optimization of a prostate plane. Dose–volume restrictions were imposed on the bladder and the rectum. Upper dose limits were set to 49 Gy and 46 Gy, respectively. In addition, 60% of the bladder was required to receive a dose less than 25 Gy while the fractional rectum volume receiving a dose less than 22 Gy was varied in three different trials to attain 65% (Trial 1), 75% (Trial 2), and 85% (Trial 3) volume. Target prescription TD in Eqs. (4) and (13) was set to 73 Gy. The organ dose constraints for the Trial 1 were chosen so that further reduction in the upper dose limits would cause the target volume treated to 95% of the prescription (V95%) to be less than 100%. Therefore, the algorithms were required to operate in the regions where it was sufficiently challenging to maintain a good tumor coverage. The solution goals were to satisfy the dose–volume constraints for the organs while minimizing the deviation from the target prescription. The target and organ contours extracted from a CT slice (Fig. 1) were discretized into 0.6 $\times 0.6\ \mathrm{cm}^2$ pixels, yielding 105 target points, 201 bladder points, and 40 rectum points. Each of the nine equiangular beams encompassing the target was quantized to 19 rays also at the 0.6-cm intervals. The matched dose-ray sampling resolution permits the use of the same dose points for both the optimization and the computation of dose–volume histograms. The dose contributions to the dose points from the adjacent rays were precomputed by a simplified dose computation model using the tissue-phantom ratios from an 18-MV machine. Effects of the scatter and beam penumbra were not considered. The model is adequate, however, for the purpose of demonstrating the algorithms.

III. RESULTS

Table II summarizes the dose–volume constraints and solutions obtained by the two algorithms. Both algorithms were able to control the dose-volume relationships and satisfy all of the constraints in the process of intensity modulated beam optimization. In both methods, the cumulative dose–volume histograms for the bladder remained essentially unchanged throughout the trials in agreement with the prescription (Fig. 2). The rectum curves changed as the partial volume not permitted to exceed 22 Gy was increased (Fig. 3). Histograms for the rectum reflect the difference in the two algorithms. POCS appears to reduce the overall organ dose more than CFM. This is understood by recalling that POCS controls the dose–volume relationship primarily by reducing the integral organ dose until the constraints are met, while the cost function used in this study tries to suppress the dose exceeding E_2, which corresponds to the fractional volume prescription. This also explains why in Trial 3, where the fractional volume requirement was most stringent, the upper limit for the rectum was overconstrained in the case of the cost function method, causing the decrease in the V95% value (Fig. 4).

In CFM the constraint parameters, γ, E_1, and E_2, were set to the prescribed values and it was only necessary to vary the Lagrange multiplier λ to attain the constraint goals. However, several trial runs were required to determine the best λ values for each trial. Solutions were obtained in about 50 000 iterations. In POCS, the adjustment of the allowable integral dose parameter I was performed automatically, thus eliminating the need for multiple runs. Convergence was achieved in less than 400 iterations.

The algorithms were coded in C and compiled with the GNU-gcc compiler version 2.7.2 (Free Software Foundation, Cambridge, MA) using the $-O2$ optimization flag. The programs were executed on a DEC Alpha Station 500 (Digital Equipment Corporation, Maynard, MA). The total solution times for each method were measured using the system called ''gettimeofday'' from within the program. It took an average of 481.9 s for the CFM method and 5.2 s for the POCS. Although the examples shown in this study were 2D, the algorithms should remain valid when extended to 3D, as there is no conceptual difference between 2D and 3D applications except for the computational complexity. The dose computation matrix **A** was sparse because only the primary beam contributions were considered. Inclusion of the scattered radiation will fill the matrix **A** with nonzero elements. For 3D computation, the size of the **A** matrix will increase

TABLE II. Dose-volume constraints and solutions obtained by CFM and POCS methods.

	Critical organ dose (Gy)		Tumor dose (Gy)				
Trial	Bladder	Rectum	Min	Max	Mean	σ	V95%[a]
1. Constraints	100%≤49	100%≤47					
CFM	100%≤47	100%≤47					
POCS	100%≤49	100%≤47					
Constraints	60%≤25	65%≤22					
CFM	60%≤25	65%≤22					
POCS	60%≤25	65%≤22					
Objective				minimize deviation from 73 Gy			
CFM			70.8	73.7	73.0	0.5	100%
POCS			71.2	75.9	73.3	1.3	100%
2. Constraints	100%≤49	100%≤47					
CFM	100%≤48	100%≤47					
POCS	100%≤49	100%≤47					
Constraints	60%≤25	75%≤22					
CFM	62%≤25	75%≤22					
POCS	61%≤25	75%≤22					
Objective				minimize deviation from 73 Gy			
CFM			71.5	74.5	73.0	0.6	100%
POCS			71.5	75.9	73.4	1.3	100%
3. Constraints	100%≤49	100%≤47					
CFM	100%≤46	100%≤40					
POCS	100%≤48	100%≤46					
Constraints	60%≤25	85%≤22					
CFM	63%≤25	85%≤22					
POCS	61%≤25	85%≤22					
Objective				minimize deviation from 73 Gy			
CFM			68.6	76.7	72.9	1.6	97%
POCS			69.1	76.9	73.7	1.9	100%

[a]Volume treated to 95% of the prescribed dose.

significantly. In order to evaluate the effect of increased computational demand on POCS method when a more sophisticated dose model is employed, the **A** matrix was fully populated with nonzero numbers; also, the treatment geometry was extended to 3D by replicating the 2D slice 19 times at 6-mm spacing. Each beam now consisted of 2D array of 19×19 rays yielding the total of 3249 beam weights to be optimized. The number of 3D dose points was 6574 and the corresponding **A** matrix consisted of 21 358 926 elements. The POCS method requires pseudo-inversion of the **A** matrix [Eq. (15)]. But this needs to be performed only once and the results are saved for repeated usage. The matrix inverse operation took 0.5 s and the subsequent projection operation 8.4 s per iteration. As shown in Fig. 5, the pseudo-inverse computation time scales linearly with the size of the **A** matrix which is proportional to the number of floating point calculations. As for the effect of including the lateral scatter, a recent study by Mohan *et al.* reports that ''taking lateral transport into consideration would lead to dose distributions that are significantly closer to the desired dose distributions.''[31] Lateral transport can provide additional fluence that can be deposited to increase the target dose uniformity. It can also provide protection for normal tissues by allowing the reduction of margins and by generating steeper dose gradients in the overlap regions.

IV. DISCUSSION

In general, the nature of an optimization problem is determined by how the constraints are formulated. In this study, we have constructed the dose–volume constraints using two substantially different methods. In the cost function method the constraints were designed to penalize solutions that increased the fractional volume permitted to exceed a dose limit while maintaining the entire organ to below a maximum bound. The penalty function is a product of the dose penalty and the volume penalty terms. The dose penalty function, which is a sum of a quadratic and a linear function, is an increasing function with respect to dose. Also, the volume penalty term as a function of the volume parameter γ' is increasing for $\gamma'>\gamma$, or when the volume penalty is active. Therefore, the product given by Eq. (1) is an increasing function with no apparent local minima. However, it can be argued that it is possible for different segments of the organ volume to meet the dose–volume constraints, leading to multiple minima in dose–volume penalty score. For instance, assuming that no part of an organ is permitted to exceed 50 Gy and that one-third of the volume must remain below 40 Gy, multiple minima will occur if different parts of the organ can comprise the required fractional volume while maintaining the whole organ dose to below 50 Gy. These multiple

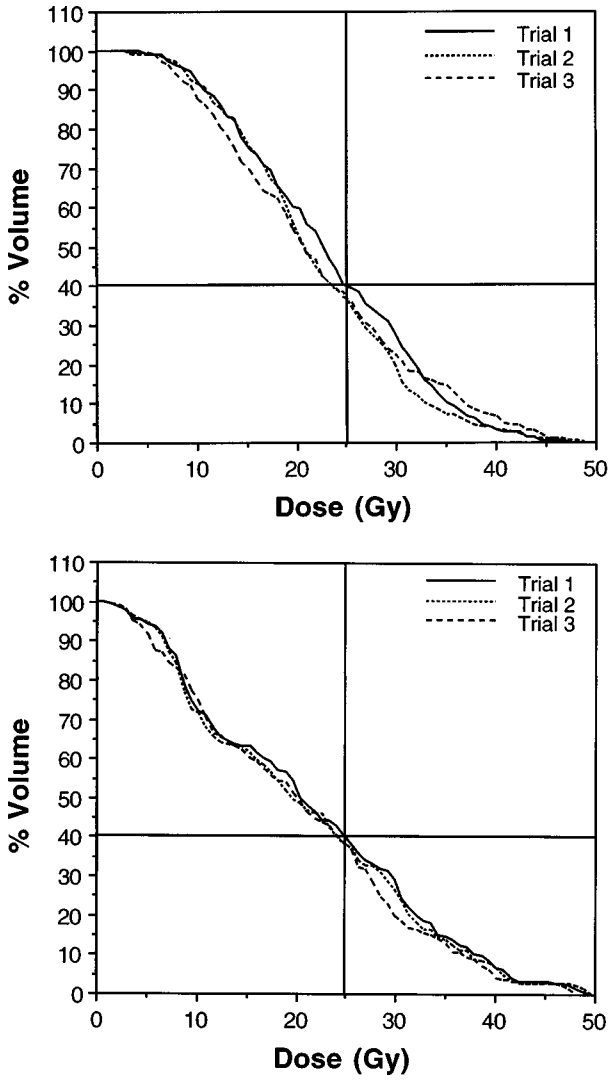

FIG. 2. Dose volume histograms for the bladder obtained with (a) the cost function minimization method and with (b) the method of projections onto convex sets. The partial volume limits for rectum were varied for each trial while for bladder the dose–volume constraints were kept constant. The entire volume was required to remain less than or equal to 49 Gy. In addition, 60% volume was restricted to doses less than or equal to 25 Gy, i.e., only 40% volume was allowed to exceed 25 Gy as indicated by the vertical line.

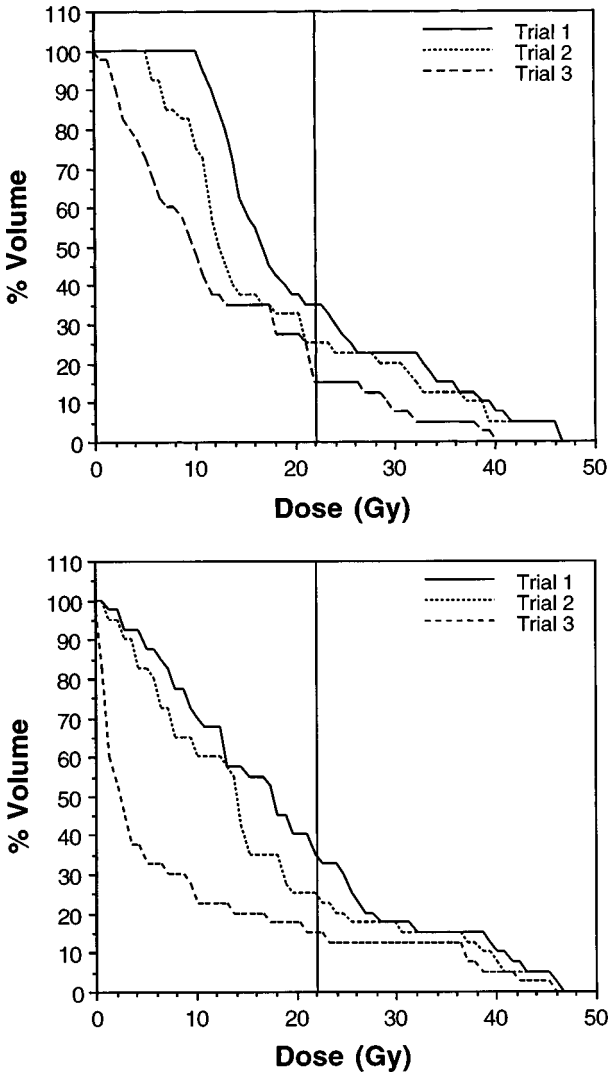

FIG. 3. Dose–volume histograms for the rectum obtained with (a) the cost function minimization method and with (b) the method of projections onto convex sets. The dose–volume constraints were varied for each trial as follows: (Trial 1) $100\% \leqslant 47$ Gy and $65\% \leqslant 22$ Gy; (Trial 2) $100\% \leqslant 47$ Gy and $75\% \leqslant 22$ Gy; (Trial 3) $100\% \leqslant 47$ Gy and $85\% \leqslant 22$ Gy. The vertical line indicates the 22-Gy partial dose–volume limit E_2.

solutions all satisfy the organ dose requirements, but it is likely that they differ in terms of the total cost if other constraints and objectives are present. For instance, out of few possible solutions that satisfy the organ dose–volume constraints (local minima), only one may give the best tumor coverage (global minimum). The presence of multiple local minima in dose–volume optimization has been noted by others.[32,33]

In POCS, there is no concept of local minima. Rather, the nature of optimization problem is determined by whether convex formulation of the constraints exists. If a convex formulation is found, then a solution can be obtained deterministically. Our strategy involved redefining the nonconvex dose–volume problem in terms of a limit on integral dose. This permitted the use of the convex paradigm. The dose–volume manipulation was accomplished by reducing the integral dose parameter I while at the same time limiting the maximum organ dose. In radiotherapy it is generally not possible for all the convex sets to intersect at a single point since the target objectives and organ constraints are by nature mutually exclusive. If convex sets do not have a common intersection, the projection operation will eventually reach a limit cycle at which point no further improvement in convergence can be made. This condition represents no improvement in the mean-square distance between projections and is detectable by the termination condition associated with Eq. (23).

This study focused on the partial volume constraints to normal tissues. Ideally, the dose–volume constraint should also be imposed on the target such that in addition to the minimum ($E_{\min}$) and maximum dose ($E_{\max}$) limits, a fraction of the target volume permitted to receive doses below prescription dose TD can be specified where $E_{\min} < TD < E_{\max}$. In the CFM method, an additional penalty function can be

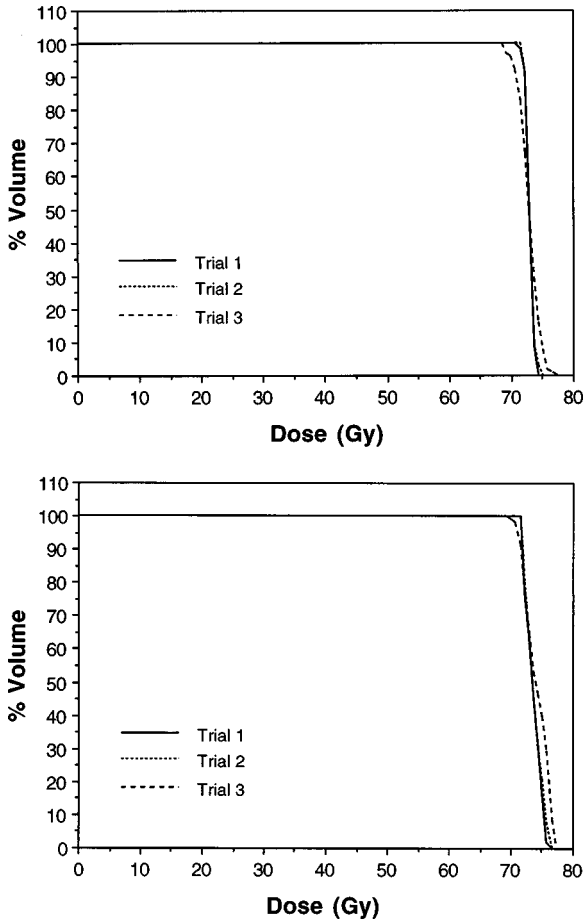

FIG. 4. Dose–volume histograms for the target obtained with (a) the cost function minimization method and with (b) the method of projections onto convex sets. The target dose prescription was 73 Gy.

formulated that penalizes doses below $E_{\min}$ and those above $E_{\max}$ with a cost amplifying factor similar to Eq. (3). In the POCS method the bounded dose vectors will form a convex set:

$$C_T=\left\{\vec{d}\,\middle|\,E_{\min}\leqslant\sum_{k=1}^{Q}\vec{d}_k\leqslant E_{\max}\right\}. \tag{24}$$

Imposition of volume objective can be implemented again using the concept of integral dose limit. The corresponding convex set is

$$C_{TI}=\left\{\vec{d}\,\middle|\,I_T(i)\leqslant\sum_{\text{target}}\sum_{k=1}^{Q}\vec{d}_k\leqslant I_{T\,\max}\right\}, \tag{25}$$

where $I_{T\,\max}=$(number of target voxels)$\times E_{\max}$. The target integral dose parameter I_T is varied at each iteration until the target dose–volume prescription is met.

It is often accepted that the inverse techniques[26–29] (including POCS) are not suitable for lung tumors[32] in which, unlike the prostate cancer, the division of the normal tissues into high and low dose regions is not clearly defined. However, a recent follow-up study reports that the reason for previous failure of the inverse method as applied to lungs was precisely that the dose–volume factor had been

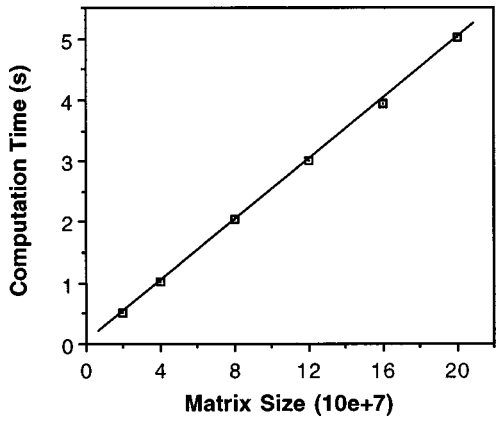

FIG. 5. Computation time for the pseudo-inverse operation scales linearly with the size of the matrix.

neglected.[9] The study also compared the dose–volume-dependent dosimetric optimization with the biological optimization. The results suggest that even without the use of biological parameters, clinically relevant optimization is possible if the dose-volume factor is considered in the dosimetric model. These findings indicate that the applicability of the inverse methods can, in fact, be extended to more general use if a dose–volume control such as the one described here is incorporated.

V. CONCLUSIONS

Two optimization techniques for intensity beam modulation with dose–volume constraints were presented. The methods offer the ability to incorporate the volume effect of the normal tissue tolerance. In the cost function minimization method, the constraints were designed to penalize solutions that increased the fractional volume permitted to exceed a certain tolerance while maintaining the entire organ to below a maximum dose. The volume-sensitive penalty function described here can easily be adopted by existing optimization programs such as simulated annealing. In the convex projection method, the nonconvex problem of the dose–volume constraint was reformulated in terms of the integral dose limit, which permitted the use of convex constructs. The convex projection method can find solutions in much shorter time with minimal user interaction.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for the insightful comments and constructive suggestions.

[a]Electronic-mail: cho@radonc.washington.edu

[1]A. Niemierko, ''Selection of the objective function for optimization of intensity modulated beams,'' Med. Phys. **23**, 1172 (1996) (abstract).
[2]T. E. Schultheiss, C. G. Orton, and R. A. Peck, ''Models in radiotherapy: Volume effects,'' Med. Phys. **10**, 410–415 (1983).
[3]P. Rubin, L. S. Constine III, D. F. Nelson, and G. W. Casarett, ''Late effects of cancer treatment: radiation and drug toxicity,'' in *Clinical Oncology: A Multidisciplinary Approach for Physicians and Students*, 7th ed., edited by P. Rubin (Saunders, Philadelphia, 1993), pp. 735–765.
[4]R. Paterson, *The Treatment of Malignancy by Radium and X-Rays* (Williams and Wilkins, Baltimore, 1948).

[5]B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson, ''Tolerance of normal tissue to therapeutic irradiation,'' Int. J. Radiat. Oncol., Biol., Phys. **21**, 109–122 (1991).
[6]See, for example, the bibliography in chapter two of S. Webb, *The Physics of Three-Dimensional Radiation Therapy* (IOP, Bristol, 1993), pp. 127–134.
[7]M. Langer, R. Brown, M. Urie, J. Leong, M. Stracher, and J. Shapiro, ''Large scale optimization of beam weights under dose-volume restrictions,'' Int. J. Radiat. Oncol., Biol., Phys. **18**, 887–893 (1990).
[8]M. Langer, S. Morrill, R. Brown, O. Lee, and R. Lane, ''A comparison of mixed integer programming and fast simulated annealing for optimizing beam weights in radiation therapy,'' Med. Phys. **23**, 957–964 (1996).
[9]T. Bortfeld, J. Stein, and K. Preiser, ''Clinically relevant intensity modulation optimization using physical criteria,'' XIIth International Conference on the Use of Computers in Radiation Therapy, Salt Lake City, May 27–30 (1997), pp. 1–4.
[10]M. P. Carol, R. V. Nash, R. C. Campbell, R. Huber, and E. Sternick, ''The development of a clinically intuitive approach to inverse treatment planning: partial volume prescription and area cost function,'' XIIth International Conference on the Use of Computers in Radiation Therapy, Salt Lake City, May 27–30 (1997), pp. 317–319.
[11]M. Oldham and S. Webb, ''The optimization and inherent limitations of 3D conformal radiotherapy treatment plans for the prostate,'' Br. J. Radiol. **68**, 882–893 (1995).
[12]S. Webb, ''Optimization of conformal radiotherapy dose distributions by simulated annealing,'' Phys. Med. Biol. **34**, 1349–1370 (1989).
[13]S. Geman and D. Geman, ''Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images,'' IEEE Trans. Pattern. Anal. Mach. Intell. **6**, 721–41 (1984).
[14]H. Szu and R. Hartley, ''Fast simulated annealing,'' Phys. Lett. B **122**, 157–162 (1987).
[15]G. S. Mageras and R. Mohan, ''Application of fast simulated annealing to optimization of conformal radiation treatments,'' Med. Phys. **20**, 639–647 (1993).
[16]L. M. Bregman, ''Finding the common point of convex sets by the method of successive projections,'' Dokl. Akad. Nauk. USSR **162**, 487–490 (1965).
[17]L. G. Gubin, B. T. Polyak, and E. V. Raik, ''The method of projections for finding the common point of convex sets,'' USSR Comput. Math. Math. Phys. **7**, 1–24 (1967).
[18]S. J. Yen and H. Stark, ''Iterative and one-step reconstruction from nonuniform samples by convex projections,'' J. Opt. Soc. Am. **7**, 491–499 (1990).
[19]H. Peng and H. Stark, ''Signal recovery with similarity constraints,'' J. Opt. Soc. Am. **6**, 844–851 (1989).
[20]R. J. Marks II, ''A class of continuous level associative memory neural nets,'' Appl. Opt. **26**, 2005–2009 (1987).
[21]M. K. Ozkan, A. M. Tekalp, and M. I. Sezan, ''POCS-based restoration of space-varying blurred images,'' IEEE Trans. Image Process. **3**, 450–454 (1994).
[22]D. C. Youla and H. Webb, ''Image restoration by the method of convex set projections: Part I-Theory,'' IEEE Trans. Med. Imaging **MI-1**, 81–94 (1982).
[23]P. Oskoui and H. Stark, ''A comparative study of three reconstruction methods for a limited-view computer tomography problem,'' IEEE Trans. Med. Imaging **8**, 45–58 (1989).
[24]S. Lee, P. S. Cho, R. J. Marks II, and S. Oh, ''Conformal radiotherapy computation by the method of alternating projections onto convex sets,'' Phys. Med. Biol. **42**, 1065–1086 (1997).
[25]S. Kaczmarz, ''Angenaherte Auflosung von Systemen linearer Gleichungen,'' Bull. Acad. Sci. Pologne, A **35**, 335–357 (1937).
[26]B. K. Lind, ''Properties of an algorithm for solving the inverse problem in radiation therapy,'' Inverse Probl. **6**, 415–426 (1990).
[27]T. Bortfeld, J. Burkelbach, R. Boesecke, and W. Schlegel, ''Methods of image reconstruction from projections applied to conformation therapy,'' Phys. Med. Biol. **35**, 1423–1434 (1990).
[28]T. W. Holmes, T. R. Mackie, D. Simpkin, and P. Reckwerdt, ''A unified approach to the optimization of brachytherapy and external beam dosimetry,'' Int. J. Radiat. Oncol., Biol., Phys. **20**, 859–873 (1991).
[29]D. H. Hristov and B. G. Fallone, ''An active set algorithm for treatment planning optimization,'' Med. Phys. **24**, 1455–1464 (1997).
[30]P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic New York, 1981), p. 41.
[31]R. Mohan, Q. Wu, X. Wang, and I. Stein, ''Intensity modulation optimization, lateral transport of radiation, and margins,'' Med. Phys. **23**, 2011–2021 (1996).
[32]R. Mohan, X. Wang, A. Jackson, T. Bortfeld, A. L. Boyer, G. J. Kutcher, S. A. Leibel, Z. Fuks, and C. C. Ling, ''The potential and limitations of the inverse radiotherapy technique,'' Radiother. Oncol. **32**, 232–248 (1994).
[33]J. O. Deasy, ''Multiple local minima in radiotherapy optimization problems with dose-volume constraints,'' Med. Phys. **24**, 1157–1161 (1997).