# A Feasible Application of Constrained Optimization in the IMRT System

Juan M. Artacho Terrer*, Miguel A. Nasarre Benedé, Emiliano Bernués del Río, and Santiago Cruz Llanas

***Abstract*—The planning of intensity modulated radiation therapy cancer treatment poses an inverse problem and is usually solved by optimization methods. Treatment planning, in a majority of cases, requires restrictions to be imposed on the healthy organs, sensitive to the radiation, which justifies the use of constrained optimization. The application of these techniques in treatment planning usually involves serious complications and limitations due to the huge number of variables appearing in the planning process. This leads to large computation times and memory requirements. In this paper, strategies and algorithmic issues are proposed in order to cope with these limitations. The proposed methods have been applied and tested in real cases of prostate cancer, obtaining satisfactory results regarding computational limitations.**

***Index Terms*—Biomedical applications of radiation, dosimetry, optimization methods, radiation therapy.**

## I. Introduction

RADIATION therapy is the treatment of cancer with ionizing radiation, in such a way that radiation, when passing through the tissue, damages tumor cells slowing or reversing the growth of tumors. A relevant development in the last decade is the intensity modulated radiotherapy (IMRT) system, where the distribution of the dose can be controlled with spatial accuracy using multileaf collimators (MLCs) [1], [2]. At the surface of a collimator the radiation delivered by the corresponding beam can be seen as a two-dimensional discrete pattern of *beamlets*. Numerical values associated to these beamlets, corresponding to fluence or energy fluence, are called *beamlet weights*. A schematic diagram of the system is shown in Fig. 1.

Inverse planning of IMRT systems consists of automatically determining the beamlet weights, so that a prescribed dose can be attained at discrete volume locations (*voxels*). Other parameters (such as the number of beams and their orientations) could additionally be considered [3], [4], although this issue is not taken into account here.

Several mathematical frameworks have been proposed in order to solve this IMRT inverse planning. Most of them are particular ways of analyzing large ill-conditioned systems of equations, where the beamlet weights are linearly related to the data, which are the prescribed doses at voxels (see [2] or Section II). Because of the nature of the problem, different types of constraints should also be considered. *Physical constraints* arise because of the limitations of the IMRT system (nonnegativeness of the beamlet weights). *Medical constraints* are prescribed by the oncologist (limits on the received dose for different organs). The problem can be so hard that semi-automatic or interactive solutions are sometimes regarded as practical trade-offs [5]. Fully automatic frameworks include, among others: algebraic methods (constrained least squares [6]), stochastic optimization methods (such as simulated annealing [7] and genetic or evolutionary algorithms [4]) or backprojection-based methods [8].

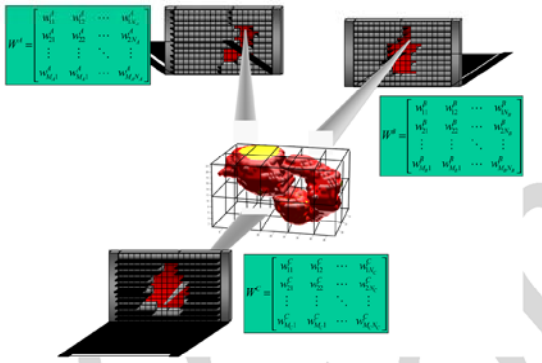

Fig. 1. A schematic diagram for IMRT (3 beams).

It is important to keep in mind that, despite the number of available theoretical frameworks mentioned so far, several versions of nonstochastic traditional gradient-search methods remain extensively applied. This basic idea can be expressed with different names in IMRT planning: constrained optimization [9], active set algorithms [10] or nonlinear programming [11]. For this kind of methods, mathematic definitions must be given for the specific objective function and for the constraints of the problem.

In IMRT planning, linear or quadratic objective functions are the most usual choices [11], [12]. When the constraints are also linear, these special objective functions can lead to well-known optimization problems such as linear programming (LP) [13]–[15] or quadratic programming (QP). LP and QP have good theoretic properties, such as well-known conditions for convexity. Moreover, an important advantage is that a great number of practical algorithmic solutions are available for these problems [16].

The election of quadratic objective functions, compared to linear ones, has certain advantages. Least-squares emerge as a natural solution for IMRT problems [6] and, moreover, quadratic objective functions can be thought as more flexible

Manuscript received July 22, 2005; revised July 29, 2006. This work was supported in part by the Government of Aragón (Spain) Investigation Program under Project PIP 061/2005 and in part by Técnicas Radiofísicas S.L (Spain). *Asterisk indicates corresponding author.*

*J. M. Artacho Terrer is with the Communications Technology Group, 13A, University of Zaragoza (Spain), C/ María de Luna 1, CP: 50018, Zaragoza, Spain (e-mail: jartacho@unizar.es).

M. A. Nasarre, E. Bernues, and S. Cruz are with the Communications Technology Group, 13A, University of Zaragoza, CP: 50018, Zaragoza, Spain (e-mail: 442269@unizar.es; ebr@unizar.es; cruzll@unizar.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2006.890487

models than linear ones. The price to be paid for this choice is the increased complexity of the problem. It has been argued that, for IMRT planning, this is specially critical, as it is a large-scale problem which includes a high number of constraints [11].

One of the main concerns of this work is the proposal of special strategies in order to reduce the computational cost when nonlinear objective functions are used (see Section III). The effectiveness of the proposed techniques is shown on real cases of prostate cancer.

Results in this paper confirm that a wise election of a particular algorithmic solution, exploiting all the available knowledge, can have a great impact regarding practical concerns such as the computation time. For the type of QP problem solved here, significant differences have been found between two algorithms (Lemke and Rosen Gradient Projection algorithms). Lemke's algorithm leads to reduced computational times, as it fully exploits the nature of the QP problem, whereas Rosen's algorithm is more general (the objective function is not necessarily quadratic).

The rest of the paper is organized as follows. First, in Section II, the inverse planning problem for IMRT is posed and the particular choices for its solution are described (QP framework, Lemke and Rosen Gradient Projection algorithms). In Section III, the new proposed techniques for reducing the dimension of the problem are explained. Finally, results and conclusions are given.

## II. Inverse Planning

The external beam radiation therapy with intensity modulation poses a problem which can be represented by a mathematical model to relate the dose delivered by the linear accelerator (linac) that generates the radiation beam and the dose absorbed in the anatomy of the patient. The area affected by the tumour is divided in small cubes called voxels, usually tens of thousands, in order to obtain an accurate representation of the volume.

Likewise, radiation beams are discretized in beamlets [6]. The radiation beam delivers a dose which is described in terms of a matrix of weights (or fluence matrices), where each weight determines the dose delivered by its corresponding beamlet (see Fig. 1).

The dose delivered to each voxel is given by

$$D_{oi} = \sum_{j=1}^{n} f_{ij} w_j \quad i = 1, \ldots, m \tag{1}$$

where $n$ is the number of beamlet weights, $m$ is the number of *voxels*, $w_j$ is the value of the $j$th weight, and $f_{ij}$ is the contribution of dose from the $j$th beamlet to the $i$th *voxel*.

The reordering of the weights $w_j$ of the fluence matrices in a vector $w$ allows the previous expression to be expressed as a matrix

$$D_o = F \cdot w \tag{2}$$

where $F$ is an $m \times n$ matrix whose elements are the coefficients $f_{ij}$ and $D_o$ is an $m$ dimensional vector where the doses corresponding to all the voxels are grouped. Consequently, once the matrix $F$ is calculated, the dose obtained for a specific intensity pattern can be found directly.

The inverse problem for radiotherapy treatment planning is based on the knowledge of matrix $F$. The goal is to find the weighting vector $w$ that approximates the dose $D_p$ prescribed by the oncologist in each organ or structure. Usually, a distinction is made between clinical target volume (CTV) (or affected organ), organs at risk (OARs) (or healthy organs), and unspecified tissues (UTs).

For the formulation of the problem, an objective error function $G(w)$ has to be defined. It has been built as the sum of quadratic terms, measuring the difference between prescribed and obtained doses. In the definition of the objective function, a set $S$ of organs or structures is considered (CTVs, OARs, and UT). The $k$th structure comprises $N_{T_k}$ voxels. $T_k$ is a set of indexes for the voxels belonging to this structure. $N_{T_k}$ enters a normalization term in the definition of the objective function preventing that big organs dominate its value. Moreover, a weighting coefficient $P_k$ is introduced, allowing the oncologist to attach more relevance to one organ or another in the treatment. Objective functions considering similar criteria than the ones described have been proposed elsewhere [17], [18]. The objective function used here is defined as follows:

$$\begin{aligned} G(w) &= \sum_{k \in S} \frac{P_k}{N_{T_k}} \sum_{i \in T_k} (D_o i - D_p i)^2 \\ &\sum_{k \in S} P_k = 1 \quad 0 < P_k < 1 \\ &\sum_{k \in S} N_{T_K} = m \\ &S = \{CTV, OAR1, OAR2, \ldots\}. \end{aligned} \tag{3}$$

This function can be written as a quadratic function of the weights as

$$G(w) = \frac{1}{2} w^T Q w + R^T w + c \tag{4}$$

being

$$\begin{aligned} Q &= \sum_{k \in S} \frac{2P_k}{N_{T_k}} F_k^T F_k \quad R = -\sum_{k \in S} \frac{2P_k}{N_{T_k}} F_k^T D_{pk} \\ c &= \sum_{k \in S} \frac{P_k}{N_{T_k}} D_{pk}^T D_{pk} \end{aligned} \tag{5}$$

where $F_k$ and $D_{pk}$ are, respectively, submatrix/subvector of $F$ and $D$ taking only the parts related to the $k$ structure.

As several structures of the volume (OARs) need to be particularly protected, when considering the rows of matrix $F$ which belong to the voxels or structures whose dose must be constrained by an upper bound, matrix $A_U$ is obtained, being $b_U$ a vector that contains the information of the maximum allowed dose. We apply dose constraints by means of imposing

$$A_U w \leq b_U$$

$m_U$ being the number of rows in $A_U$ and $n$ the number of beamlet weights $w_j$.

Furthermore, given that the elements $w_j$ represent radiation intensities, and these are positive quantities, to obtain a physically significant solution it is necessary to impose an additional restriction owing to the nonnegativity of the weightings $(w \geq 0)$. This is such that defining $-A_w w \leq b_w$, where $A_w$ is the identity matrix and $b_w$ is a $n$-sized vector with all its elements are null. Resulting in the matrices

$$A_c = \begin{bmatrix} A_U \\ -A_w \end{bmatrix} \quad b_c = \begin{bmatrix} b_U \\ -b_w \end{bmatrix}. \tag{6}$$

The whole set of restrictions can be expressed as

$$A_c w \leq b_c. \tag{7}$$

The number of dose constraints will be

$$m_c = m_U + n. \tag{8}$$

This allows the planning process to be described by a conditioned optimization problem

$$\begin{aligned} \text{Minimize}: \quad & \frac{1}{2} w^T Q w + R^T w + c \\ \text{Subject to}: \quad & A_c w \leq b_c. \end{aligned}$$

The existence of inequality equations on constrained optimization methods based on the Lagrange theory requires the use of new slack variables for the conversion of inequality equations into equality equations [19].

### A. Slack Variables, Lagrange Multipliers, Lagrangian, and KKT Conditions

The limitation of the dose in a voxel to a value $b_{U_{i_c}}$ corresponding to an organ at risk is represented by

$$\sum_{j=1}^{n} a_{i_c j} w_j \leq b_{U_{i_c}} \quad i_c = 1, \ldots, m_U \tag{9}$$

or

$$h_{i_c}(w) = \sum_{j=1}^{n} a_{i_c j} w_j - b_{U_{i_c}} \leq 0 \tag{10}$$

where the coefficients $a_{i_c j}$ are the elements of the matrix $A_c$. Therefore, the upper condition in the voxels of OAR involves the creation of these variables. This is such that introducing slack variables $\left(s_i^2\right)$ the previous equation is transformed into an equation of equality

$$\sum_{j=1}^{n} a_{i_c j} w_j - b_U i_c + s_{i_c}^2 = 0 \quad i_c = 1, \ldots, m_U. \tag{11}$$

As $s_i^2$ is a positive term, the dose received by the $i$th voxel does not surpass the limit $b_{U i_c}$ established.

In the same way, the condition of nonnegativity of the weightings is solved using slack variables to translate

$$\begin{aligned} h_{i_c}\left(w_{i_c - m_U}\right) + s_{i_c}^2 &= -w_{i_c - m_U} + s_{i_c}^2 = 0 \\ i_c &= m_U + 1, \ldots, m_c. \end{aligned} \tag{12}$$

The Lagrangian is formed as

$$L(w, \lambda, s) = G(w) + \sum_{i_c=1}^{m_c} \lambda_{i_c} \left(h_{i_c}(w) + s_{i_c}^2\right) \tag{13}$$

where $\lambda$ is the vector containing the Lagrange multipliers $\lambda_{i_c}$ for the constraints and $s$ is a vector with the slack variables $s_{i_c}$.

We finally express the Lagrangian function in matrix form for this problem as

$$\begin{aligned} L(w, \lambda, s) = & \frac{1}{2} w^T Q w + R^T w + c \\ & + \lambda_U^T \left(A_U w - b_U + s_U^2\right) + \lambda_w^T \left(-w + s_w^2\right) \end{aligned} \tag{14}$$

$\lambda_U$ and $s_U^2$ are column vectors formed with the first $m_U$ elements of $\lambda_{i_c}$ and $s_{i_c}^2$ respectively. Similarly, $\lambda_w$ and $s_w^2$ are column vectors formed with the last n elements of $\lambda_{i_c}$ and $s_{i_c}^2$.

In a constrained optimization problem, the necessary conditions for the existence of a relative minimum at a point $w^*$ means that the Karush-Kuhn and Tucker (KKT) conditions must be satisfied. Lemke and Rosen algorithms have been used in the present work for the resolution of this equation system.

### B. Lemke's Algorithm

Lemke's algorithm uses the simplex method as an effective solution to the linear complementary problem [20]. Next, we briefly show the Lemke's algorithm.

From the KKT conditions and (14), it is obtained

$$\begin{aligned} Qw + R + A_U^T \lambda_U - \lambda_w &= 0 \\ A_U w + s_U^2 &= b_U \\ w^T \lambda_w + \lambda_U^T s_U^2 &= 0 \\ w \geq 0, \quad \lambda_U \geq 0, & \\ \lambda_w \geq 0, \quad s_U^2 \geq 0. & \end{aligned} \tag{15}$$

Grouping the variables as

$$v = \begin{bmatrix} \lambda_w \\ s_U^2 \end{bmatrix} \quad z = \begin{bmatrix} w \\ \lambda_U \end{bmatrix} \quad q = \begin{bmatrix} R \\ b_U \end{bmatrix} \quad M = \begin{bmatrix} Q & A_U^T \\ -A_U & 0 \end{bmatrix} \tag{16}$$

we express the preceding conditions (15) as follows:

$$v - M \cdot z = q \tag{17}$$

$$v^T z = 0. \tag{18}$$

Lemke adds a new variable $z_0$ with the purpose of building the next equation system

$$Iw - Mz - ez_0 = q. \tag{19}$$

Being $I$ the identity matrix and $e = [1 \quad 1 \quad \dots \quad 1]$. This system is described by a table where the method performs operations in the columns and rows of the table until the optimal solution is achieved subject to (18).

### C. Rosen's Algorithm

The Rosen's method is based on the theory of feasible directions, such that the search for a solution is carried out within the feasible region in a direction that reduces the desired function and meets all the conditions [19]. This method obtains the solution with an iterative process in which the solution vector $w^k$ in the $k$th iteration is updated according to the expression

$$w^{k+1} = w^k + \alpha \bar{d}. \tag{20}$$

The searching direction $\bar{d}$ is the most similar direction to the opposite of the gradient vector of the objective function in $w^k$, which is expressed as

$$\nabla G(w^k) = \left(Qw^k + R\right). \tag{21}$$

The most similar direction is the one which minimizes the expression

$$\left(-G\left(w^k\right) - \overline{d}\right)^T \cdot \left(-G\left(w^k\right) - \overline{d}\right). \tag{22}$$

In addition, the direction $\bar{d}$ has to meet the descendant and feasibility conditions. For the direction to be descendant, so that the value of the desired function is reduced, the following equation should be met

$$\left.\frac{d}{d\alpha} G(w^k + \alpha \bar{d})\right|_{\alpha=0} = \bar{d}^T \nabla G(w_k) \leq 0. \tag{23}$$

Whilst the feasibility condition, assuming that the point $w^k$ is feasible, and since the constraints are linear, requires that for all the active conditions in $w^k$ (conditions that are within the limit of the equality of $w^k$) meet

$$\left.\frac{d}{d\alpha}\left(a^i(w^k + \alpha \bar{d}) - b_{U_i}\right)\right|_{\alpha=0} = \bar{d}^T a^i \leq 0 \tag{24}$$

where $a^i$ are the rows of $A_c$ that, in $w^k$, meet the condition $a^i w^k - b_{U_i} = 0$. The coefficients of these different rows form a vector with which the direction $\bar{d}$ has to observe the (24) to be a feasible direction; this means that a differential movement from a feasible point along this direction leads us to another feasible point.

The parameter $\alpha$ represents the distance traveled along the length of the normalized search direction $\bar{d}$, and its value is either the distance to the intersection of the closest active condition, or the distance $\alpha^*$ that meets the condition

$$\left.\frac{d}{d\alpha} G(w^k + \alpha \bar{d})\right|_{\alpha=\alpha^*} = \bar{d}^T \left(Q(w^k + \alpha^* \bar{d}) + R\right) = 0. \tag{25}$$

The iterative process is repeated until it becomes impossible to find a feasible descendant direction.

## III. Techniques for the Dimensional Reduction of the Problem

The constrained optimization methods are such that the execution times of the algorithms depend strongly on the number of restrictions on the dose and the number of weights. The number of these constraints for real cases can be so large that execution times become unpractical (in the order of hours).

In a typical case of cancer, the volume is divided into tens of thousands of voxels, of which several thousands make up the OARs. In this context, a number of beamlet weights ranging approximately from 200 to several thousands is obtained.

Three reduction techniques are next proposed in order to decrease the complexity and the number of variables in the IMRT inverse planning system.

### A. Dimensional Reduction of Constraints

Instead of applying dose constraints on the whole volume of the OAR structures, dose limitations are only applied to the boundary area of these OARs. Voxels belonging to the boundary are defined as those having at least one neighbour not belonging to the considered OAR. By considering only boundary voxels, the number of constraints is highly reduced.

There are several reasons which justify that voxels belonging to the boundary should be taken into account rather than voxels inside the OAR.

First, the simple and nonrealistic case where there is only one single beam is considered. It is known that the radiation is attenuated along the distance, and so, the most radiated voxels of the OAR would be the nearest ones to the source of the beam. These voxels will always be on the OAR boundary area.

Second, in IMRT planning more than one beam is used. Theoretically, the beams should converge on the CTV, so that the maximum dose would be delivered there. At OAR, critical voxels, under these assumptions, would be the closest ones to CTV, also belonging to the boundary as before.

When the region of convergence of the beams overlaps organs at risk and these organs are between the beam sources and the CTV, it could eventually happen that voxels inside the OAR absorb more dose than the limit imposed on boundaries. In this case, when two or more beams meet inside the OAR, the sum of their contributions is a bit higher inside the organ (shorter distance to the source beam) than on the boundary of the organ (longer distance to the source). However, our claim is that the dose inside the OAR hardly surpasses the limit that is imposed on the boundary. From an informal and intuitive point of view, this hypothesis can sound reasonable, because the dose on the

CTV should be homogeneous and the region of convergence includes CTV. However, this can be difficult to prove theoretically in a general case.

In this paper, in order to validate this assumption, the real case of an U-shaped prostate tumor is considered (see Fig. 2). For this case, the beam orientations have been intentionally chosen so that the convergence of the beams overlaps OARs (rectum and bladder). Figs. 3 and 4 show noncumulative histograms of the dose on OAR1 (rectum) for two regions: boundary (left) and internal (right). In Fig. 3, constraints are not applied. It can be seen that the rectum receives high levels of dose, especially on the boundary. When constraints are applied on the whole OAR1 volume [Fig. 4(top)] or only on the boundary [Fig. 4(bottom)] the obtained histograms are very similar. In the second case (constraints only on boundaries), there is an important reduction in the number of constraints, and so in the computer execution times. The final quality of the results is so comparable to the case when constraints are imposed to the whole OAR, although the limits of the dose can be slightly exceeded at a few internal voxels.

### B. Reduction of Voxels

The proposed beam tracing system automatically adjusts the aperture of the beams to the size of the tumor, even when the volume containing the tumor is very large. Because of this adjustment a significant reduction in the number of beamlet weights or variables can be made.

This fact is reflected on the value of matrix $F$. The $i$th row of matrix $F$ determines the contribution dose of the beamlet weights on the $i$th voxel

$$F(i,j) = f_{ij} \quad i = 1,\ldots,m \;\; j = 1,\ldots,n. \tag{26}$$

If there is a row $i$ whose elements are null (FN: Null row), it is obvious that the $i$th voxel will not be irradiated, so we can remove that row from the matrix and assign a null value to the $i$th voxel without modifying the results

$$i \in \mathrm{FN} \Leftrightarrow (f_{ij} = 0,\ \forall\, j \neg j = 1,\ldots,n) \Rightarrow D_{oi} = 0. \tag{27}$$

Likewise, a column $j$ whose elements are all nulls (CN: Null column), shows that the $j$th weight does not contribute to the radiation of the volume, so the $j$th column can be removed and a null value can be assigned to the $j$th weight

$$j \in \mathrm{CN} \Leftrightarrow (f_{ij} = 0,\ \forall\, i \neg i = 1,\ldots,m) \Rightarrow w_j = 0. \tag{28}$$

The reduced matrix $F$ obtained is

$$F = F(i \notin \mathrm{FN}, j \notin \mathrm{CN}).$$

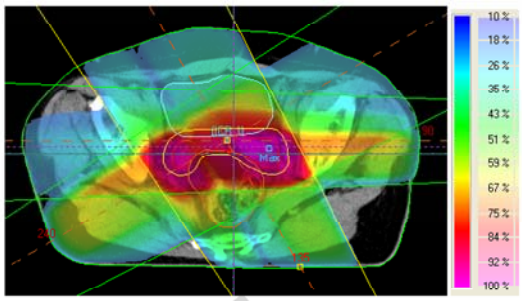


Fig. 2. Example for discussing the case where the region of convergence of the beams overlaps OAR's.

### C. Reduction of Weights

The aim of the radiation therapy is to irradiate the tumor, sparing the healthy tissue. For this reason, if there is a column $j$ on the $F$ matrix whose elements do not provide any radiation to the *voxels* of the tumor (CTV), this column may be removed and a null value is assigned to the corresponding weight. The physical meaning of that is that any positive value of the $j$th weight could distribute radiation on healthy tissues and not on the tumor. Taking this fact into account leads to a reduction in the number of variables or beamlet weights. Mathematically

$$\begin{aligned} i \in \mathrm{FN_{CTV}} &\Leftrightarrow (f_{ij} = 0,\ \forall\, j \neg j \notin \mathrm{CN_{CTV}}) \\ j \in \mathrm{CN_{CTV}} &\Leftrightarrow (f_{ij} = 0,\ \forall\, i \neg i \in T_{\mathrm{CTV}}) \end{aligned} \tag{29}$$

$F = F(i \notin \mathrm{FN_{CTV}}, j \notin \mathrm{CN_{CTV}}) w_j = 0 j \in \mathrm{CN_{CTV}}$ and $D_{oi} = 0 i \in \mathrm{FN_{CTV}}$ where $T_{\mathrm{CTV}}$ is the set of indexes of the voxels belonging to the CTV.

The matrix $F$ shows an example where in the $j$th column, although not all the elements are nulls, the elements correspondent to the CTV are nulls, so the $j$th weight does not matter, it does

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1(j-1)} & 0 & \cdots & f_{1n} \\ \vdots & \vdots & \vdots & 0 & \cdots & \vdots \\ f_{i1} & \cdots & f_{i(j-1)} & 0 & \cdots & f_{in} \\ 0 & 0 & 0 & f_{(i+1)j} & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ f_{m1} & \cdots & f_{m(j-1)} & f_{mj} & \cdots & f_{mn} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{voxels} \in \text{CTV} \\ \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{voxels} \notin \text{CTV} \end{matrix} .$$

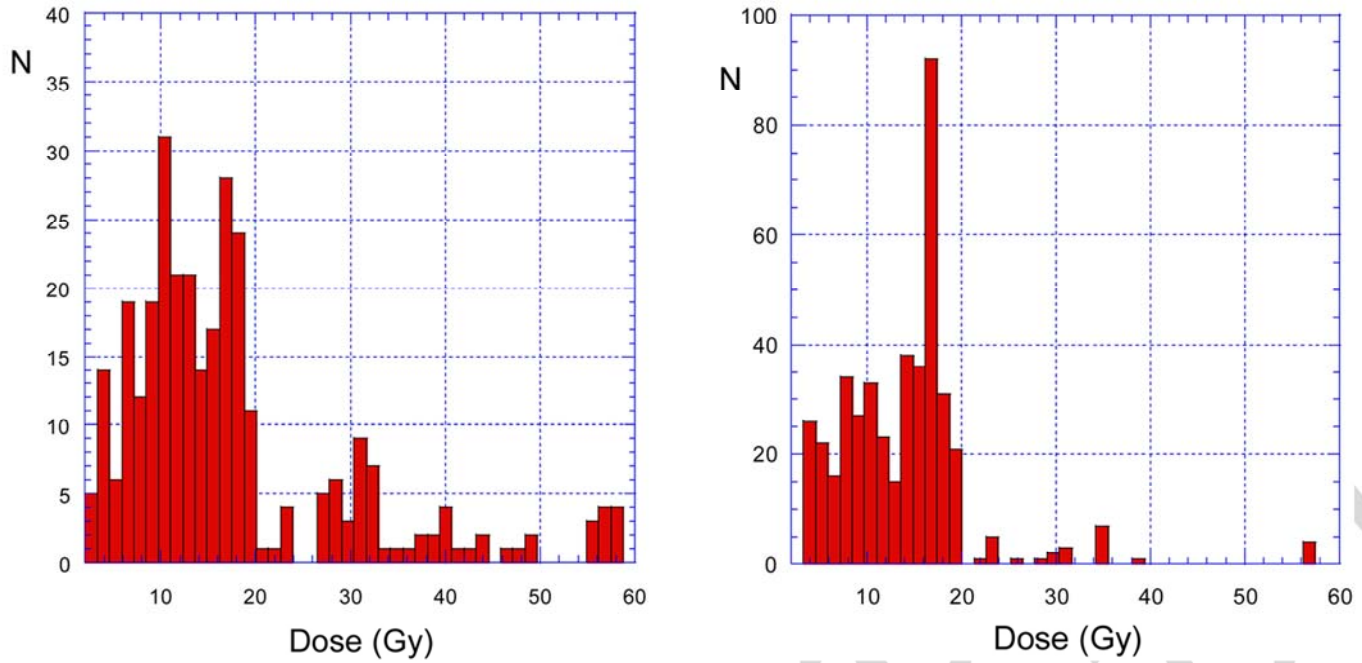


Fig. 3. Noncumulative dose histograms when constraints are not applied.

not provide radiation to the CTV, see equation at the bottom of the previous page.

Therefore, we could assign a null value to the $j$th weight and remove the $j$th column from the matrix $F$.

### D. Reduction of Weights and Voxels

It consists of the application of the reduction of weights and the reduction of voxels in a sequential way. By removing the $j$th column due to the reduction of weights, the $(i+1)$th row would have its elements nulls and could be eliminated from the matrix in application of the preceding dimensional reduction of voxels, see the equation at the bottom of the page.

## IV. Results

The proposed techniques and algorithms have been applied to a case of study, consisting of the IMRT planning for a real prostate tumor. The considered region is composed by prostate (CTV), rectum (OAR1), bladder (OAR2), and unspecified healthy tissue (UT).

As it has been previously said, the number of imposed constraints has a great impact on the cost of the optimization problem. Table I shows numbers of constraints when the proposed reduction techniques are applied to the case of study. Two cases are considered regarding the discretization of the beams: beamlets of size $(10 \times 10)$ mm$^2$ and $(10 \times 5)$ mm$^2$. Next, the meaning of the columns of Table I is explained.

- $n$ is the number of beamlet weights. Obviously, it is also the number of constraints to be imposed because of their nonnegativity.
- $m$ is the whole number of voxels in the $G$ objective function.
- $N_{\mathrm{OAR1}}$ and $N_{\mathrm{OAR2}}$ are, respectively, the whole number of considered voxels of rectum and bladder. When restrictions on the dose are applied only on OARs, the number of constraints due to this fact are $N_{\mathrm{OARs}} = N_{\mathrm{OAR1}} + N_{\mathrm{OAR2}}$.
- $m_c$ is the whole number of constraints (nonnegativity + dosage constraints).
- $m_{\mathrm{OAR1B}}$, $m_{\mathrm{OAR2B}}$, $m_{\mathrm{OARsB}}$, and $m_{\mathrm{cB}}$ have the same meaning that the variables described previously. The difference is that, for this new case, only boundary voxels for OARs are taken into account for the constraints due to dose restrictions.

Note that the value of all these variables can generally change depending on the reductions explained in Sections III-B and III-D.

The most significant values in Table I are these of variables $m_c$ and $m_{\mathrm{cB}}$, because they correspond to the number of constraints and, so, they have an important influence on the computational cost. For discussing the results, the case of beamlets of size $(10 \times 10)$ mm$^2$ can be taken as a reference. For this case, when none of the reduction techniques proposed in

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1(j-1)} & 0 & \cdots & f_{1n} \\ \vdots & \vdots & \vdots & 0 & \cdots & \vdots \\ f_{i1} & \cdots & f_{i(j-1)} & 0 & \cdots & f_{in} \\ 0 & 0 & 0 & f_{(i+1)j} & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ f_{m1} & \cdots & f_{m(j-1)} & f_{mj} & \cdots & f_{mn} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \text{voxels} \in \text{CTV} \\ \\ \\ \text{voxels} \notin \text{CTV} \end{matrix}$$

$$\xrightarrow{\text{Weights}} F = \begin{bmatrix} f_{11} & \cdots & f_{1(j-1)} & \cdots & f_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f_{i1} & \cdots & f_{i(j-1)} & \cdots & f_{in} \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{m1} & \cdots & f_{m(j-1)} & \cdots & f_{mn} \end{bmatrix} \xrightarrow{\text{Voxels}} F = \begin{bmatrix} f_{11} & \cdots & f_{1(j-1)} & \cdots & f_{1n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f_{i1} & \cdots & f_{i(j-1)} & \cdots & f_{in} \\ \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ f_{m1} & \cdots & f_{m(j-1)} & \cdots & f_{mn} \end{bmatrix}$$

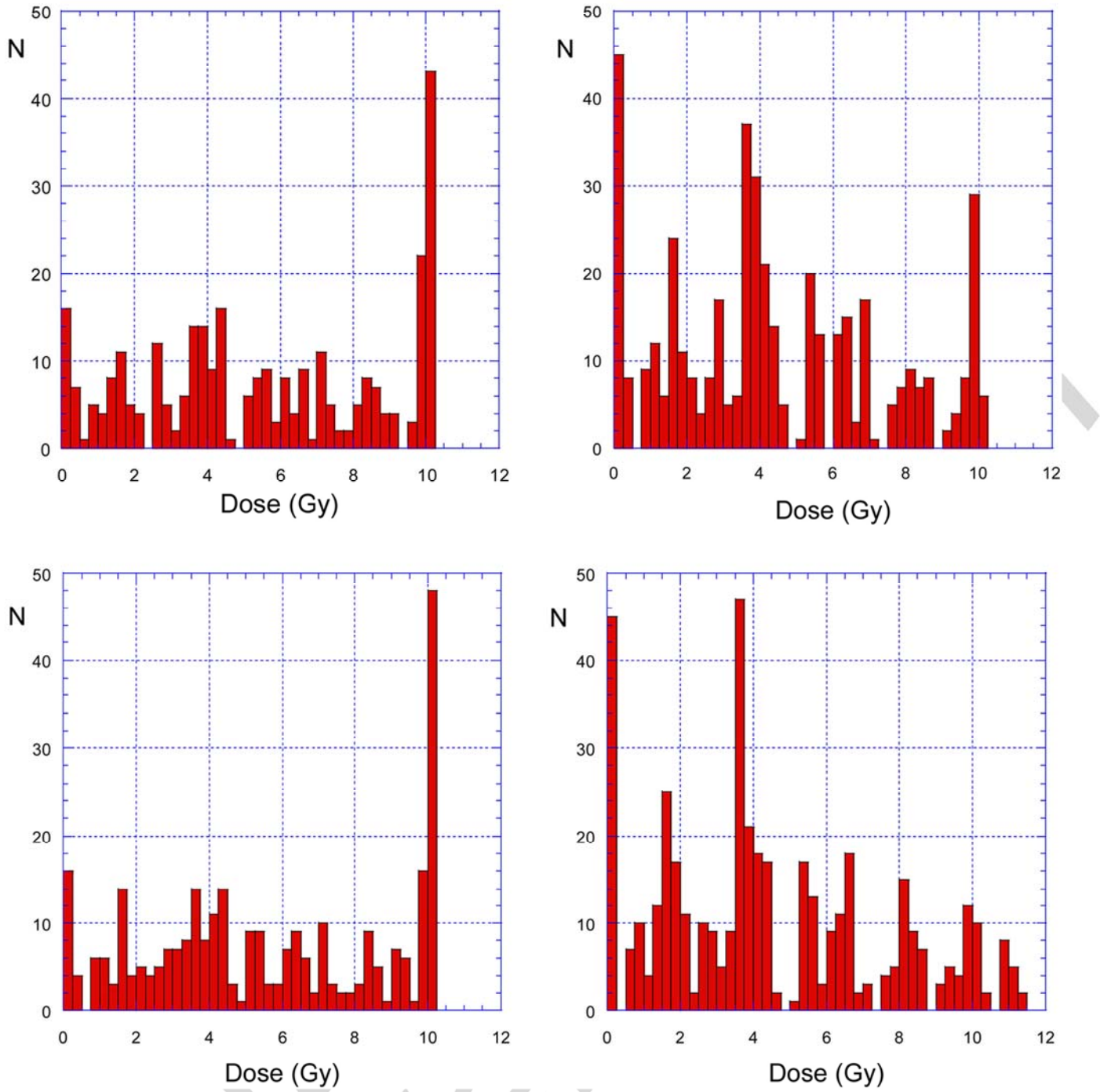


Fig. 4. Noncumulative dose histograms when constraints are applied to the whole volume or only to the boundary region.

the paper are used, 8469 constraints have to be applied. When only boundary voxels from OARs are considered for dosage restrictions, the whole number of constraints decreases to 3013. As noted in Section III-A, when constraints are applied only on boundaries, the limits can occasionally be exceeded at internal voxels. But, with the rest of the proposed reduction techniques (Sections III-B–D), and keeping all the OAR voxels for the dosage constraints, the original 8469 can still be reduced to 961. If all the reduction strategies are applied simultaneously (constraints only at boundary voxels or Sections III-A–D) the number of constraints is further reduced down to 467. From the previous discussion and, in general, from a visual inspection of the results in Table I, it can be argued that with proposed reduction methods the number of constraints to be applied decreases significantly.

In order to quantify the influence of the reduction of the number of constraints in computation times, results are shown in Table II. Computation times refer to a computer with Pentium IV processor running at 2.4 GHz and with RAM of 1 GB. The optimization algorithms were implemented in Matlab 5.3. It is reminded that two optimization algorithms (Lemke and Rosen, Sections II-B, C) had been chosen. All the computation times in

TABLE I
NUMBER OF CONSTRAINTS IN THE CASE OF STUDY

| Cell size | Reduction | n | m | $N_{OAR1}$ | $N_{OAR2}$ | $N_{OAR's}$ | $m_c$ | $m_{OAR1B}$ | $m_{OAR2B}$ | $m_{OARsB}$ | $m_{cB}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (10x10) $mm^2$ | None | 266 | 39744 | 2323 | 5880 | 8203 | 8469 | 982 | 1765 | 2747 | 3013 |
| | Voxels | 266 | 23578 | 1155 | 946 | 2101 | 2367 | 455 | 537 | 992 | 1258 |
| | Weights and Voxels | 126 | 17483 | 835 | 0 | 835 | 961 | 341 | 0 | 341 | 467 |
| (10x5) $mm^2$ | None | 671 | 39744 | 2323 | 5880 | 8203 | 8874 | 982 | 1765 | 2747 | 3418 |
| | Voxels | 671 | 20416 | 943 | 230 | 1173 | 1844 | 377 | 230 | 607 | 1278 |
| | Weights and Voxels | 382 | 16021 | 786 | 0 | 786 | 1168 | 324 | 0 | 324 | 706 |

TABLE II
COMPUTATION TIMES WITH THE PROPOSED REDUCTION STRATEGIES

| Dose limitations in OARs | Reduction of voxels and weights | Constraints on the boundary | (10x10) mm2 | (10x5) mm2 |
|---|---|---|---|---|
| ☐ Yes ☑ No | ☐ Yes ☑ No | | 3 sec | 25 sec |
| ☐ Yes ☑ No | ☑ Yes ☐ No | | 3 sec | 25 sec |
| ☑ Yes ☐ No | ☐ Yes ☑ No | ☐ Yes ☑ No | Up to 10 hours | Up to 25 hours |
| ☑ Yes ☐ No | ☐ Yes ☑ No | ☑ Yes ☐ No | 4 hours | Up to 10 hours |
| ☑ Yes ☐ No | ☑ Yes ☐ No | ☐ Yes ☑ No | 3 min | 11 min |
| ☑ Yes ☐ No | ☑ Yes ☐ No | ☑ Yes ☐ No | 40 sec | 3 min |

Table II refer to Lemke algorithm. As expected, comparing in this table values from rows 1 and 2 with respect to the values of the third row, it is confirmed that it is the use of constraints what makes the optimization problem time costly. This justifies the focus of this paper on reducing the number of constraints. The last three rows must be compared to the third one in order to study the improvement in computation times with the proposed strategies. Taking as a reference the case of $(10 \times 10)$ $mm^2$ cells of radiation, it is seen that, combining all the strategies, execution times in the order of 10 h are reduced down to 40 s. Computation times with Rosen Algorithm are not shown because they were always longer. For instance, the mentioned 40-s case took 140 s with Rosen's Algorithm. Moreover, it has to be said that the times shown in the table are only optimization times (the time needed to compute matrix $F$ is not included).

Finally, it is important to check that, with all the reductions considered, the obtained results can still match clinical specifications such as dose-volume histograms (DVH). Fig. 5 shows the obtained DVH for a prostate tumor in several cases. For all the cases the dose prescribed to the prostate (CTV) is 60 Gy. Depending on the case, the upper bound limit on the dose for the rectum (OAR1) is 10 Gy (16,6%), 20 Gy (33,3%), and 30 Gy (50%). The bladder is not considered because for this particular case it receives a very little amount of radiation. It can be seen again that the imposed upper dose limits can be exceptionally exceeded, as a consequence of the reduction technique from Section III-A (constraints only on boundaries).

Until now, the general idea of this paper regarding the constrained optimization problem has been to put constraints (upper bounds) only on OARs (on the boundary or the whole volume). However, the ideas and the mathematical formulation are general and other situations could be envisaged. As an example, it is possible to constrain only with lower limit on CTV instead of upper limits on OARs. Fig. 6 shows a case where constraints have been applied only on the boundary of CTV for a dose higher than 60 Gy. It is observed that 100% of the CTV reaches 90% of its prescribed dose, and 60% is above 100%. Not all the voxels of the CTV exceed 60 Gy because the conditions are only applied on the boundary area of the CTV, however the results are satisfying.

## V. CONCLUSION

The main focus of this paper has been to propose techniques for reducing the computational requirements (memory and

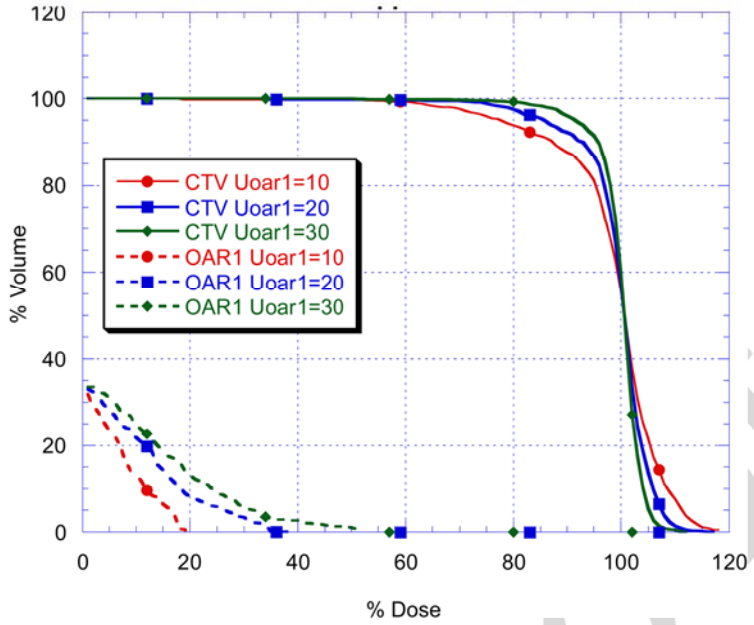


Fig. 5. DVH for different upper limits on OAR1 dose.

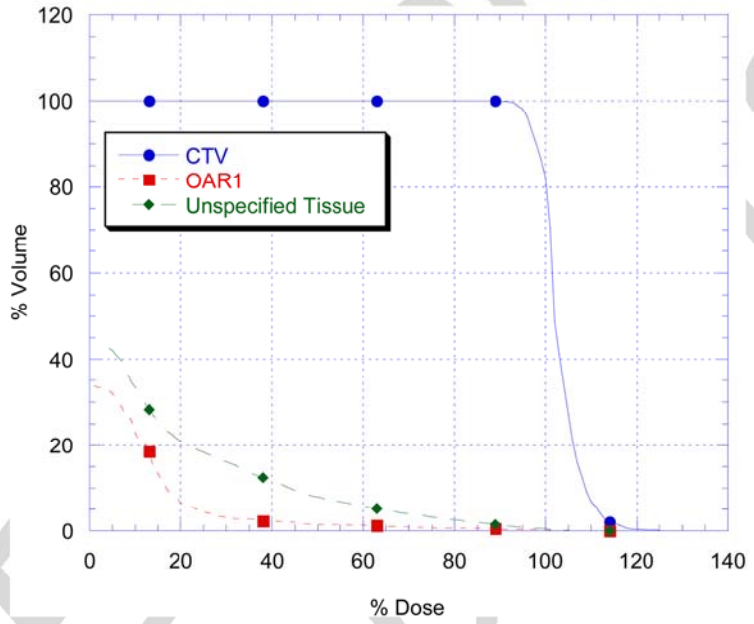


Fig. 6. DVH for lower bound on the boundary area of CTV.

time) for solving the constrained optimization problem which arises in the IMRT planning. The constrained optimization problem has been cast in a QP context and two particular algorithms (Lemke and Rosen) have been explored. The cost of these algorithms depends strongly on the number of imposed constraints. In this paper, several strategies have been proposed in order to decrease this number. Broadly, two main ideas are behind these reductions. First, because of physical considerations, the number of variables in the optimization problem (weights and voxels) can be reduced without any decrease in the quality of the results. Second, imposing constraints only on voxels on boundaries reduces greatly the cost with a little decrease in the obtained performances. The effectiveness of the proposed strategies has been studied taking into account the value of several important parameters (number of constraints, execution times, DVHs) in real cases of prostate cancer.

## ACKNOWLEDGMENT

The authors would to thank **<provide company name?>** for allowing the use of the PCRT-3D system for checking the obtained results.

## REFERENCES

[1] A. Boyer *et al.*, "Intensity-modulated radiotherapy: Current status and issues of interest," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 51, no. 4, pp. 880–914, 2001.

[2] S. Webb, "The physical basis of IMRT and inverse planning," *Br. J. Radiol.*, vol. 76, pp. 678–689, 2003.

[3] A. Pugachev and L. Xing, "Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy," *Phys. Med. Biol.*, vol. 46, no. 9, pp. 2467–2476, 2001.
[4] E. Schreibmann, M. Lahanas, L. Xing, and D. Baltas, "Multiobjective evolutionary optimization of number of beams, their orientations and weights for IMRT," *Phys. Med. Biol.*, vol. 49, no. 5, pp. 747–770, 2004.
[5] I. Rosen, H. Liu, N. Chikdress, and Z. Liao, "Interactively exploring optimized treatment plans," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 61, no. 2, pp. 570–582, 2005.
[6] S. Crooks and L. Xing, "Application of constrained least-squares techniques to IMRT treatment planning," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 54, no. 4, pp. 1217–1224, 2002.
[7] S. Morrill, K. Lam, R. Lane, M. Langer, and I. Rosen, "Very fast simulated reannealing in radiation therapy treatment plan optimization," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 31, no. 1, pp. 179–188, 1995.
[8] T. Bortfeld, J. Burkelbach, R. Boesecke, and W. Schlegel, "Methods of image reconstruction from projections applied to conformation radiotherapy," *Phys. Med. Biol.*, vol. 35, no. 10, pp. 1423–1434, 1990.
[9] O. Sauer, D. Shepard, and T. Mackie, "Application of constrained optimization to radiotherapy planning," *Med. Phys.*, vol. 26, no. 11, pp. 2359–2366, 1999.
[10] D. Hristov and B. Fallone, "An active set algorithm for treatment planning optimization," *Med. Phys.*, vol. 24, no. 9, pp. 1455–1464, 1997.
[11] D. Shepard, M. Ferris, G. Olivera, and T. Mackie, "Optimizing the delivery of radiation therapy to cancer patients," *SIAM Rev.*, vol. 41, no. 4, pp. 721–744, 1999.
[12] A. Holder and B. Salter, "A tutorial on radiation oncology and optimization," in *Tutorials on Emerging Methodologies and Applications in Operations Research*, H. Greenberg, Ed. Boston, MA: Kluwer Academic, 2004.
[13] I. Rosen, R. Lane, S. Morrill, and J. Belli, "Treatment plan optimization using linear programming," *Med. Phys.*, vol. 18, no. 2, pp. 141–152, 1991.
[14] X. Wu, Y. Zhu, and L. Luo, "Linear programming based on neural networks for radiotherapy treatment planning," *Phys. Med. Biol.*, vol. 45, no. 3, pp. 719–728, 2000.
[15] H. Romeijn, R. Ahuja, J. Dempsey, A. Kumar, and J. Li, "A novel linear programming approach to fluence map optimization for IMRT treatment planning," *Phys. Med. Biol.*, vol. 48, pp. 3521–3542, 2003.
[16] R. Fletcher, *Practical Methods of Optimization*. : Wiley, 2001, pp. 150-159–229-245.
[17] Q. Wu and R. Mohan, "Algorithms and functionality of an intensity modulated radiotherapy optimization system," *Med. Phys.*, vol. 27, no. 4, pp. 701–711, 2000.
[18] L. Xing, C. Pelizzari, E. Kuchnir, and G. Chen, "Optimization of relative weights and wedge angles in treatment planning," *Med. Phys.*, vol. 24, no. 2, pp. 215–221, 1997.
[19] S. Rao, *Optimization theory and applications*. New York: Wiley, 1977, pp. 74–74.
[20] A. Belegundu and T. Chandrupatla, *Optimization Concepts and Applications in Engineering*. Upper Saddle River, NJ: Prentice-Hall, 1999, pp. 125–128.

**Juan M. Artacho Terrer** was born in Zaragoza, Spain, in 1965. He received the B.S., M.S. and the Ph.D. degrees from the University of Zaragoza.

Since 1995, he has been an Assistant Professor at the Department of Electronics Engineering and Communications, University of Zaragoza. He is a member of the Communications Technology Group and a member of the I3A engineering research institute. His research fields involve signal proccesing, modelling of systems, inverse problems, and biomedical applications.

**Miguel A. Nasarre Benedé** was born in Huesca, Spain, in 1979. He received the B.S. degree from the Polytechnic University of Zaragoza, Zaragoza, Spain. He is currently working towards the Ph.D. degree in the Department of Electronics Engineering and Communications (DEEC), University of Zaragoza.

His research fields involve image processing, inverse problems and biomedical applications.

**Emiliano Bernués del Río** was born in Zaragoza, Spain, in 1968. He received the B.S. degree from the Polytechnic University of Madrid, Madrid, Spain, and the Ph.D. degree from the University of Zaragoza.

Since 1993, he has been an Assistant Professor at the Department of Electronics Engineering and Communications, University of Zaragoza. He is a member of the Communications Technology Group and a member of the I3A engineering research institute. His research fields involve image processing, inverse problems and biomedical applications. He is technology director of the Aragon television.

**Santiago Cruz Llanas** was born in Madrid, Spain, in 1973. From 1991 to 1996 he studied Telecommunication Engineering at Universidad Politecnica de Madrid. In 1997, he was at Ecole Nationale Superieure des Telecommunication, where he finished with undergraduate studies and received the M.S. degree in image processing and television systems.

He is an Associate Professor with Universidad de Zaragoza, Zaragoza, Spain. His main topics of reseach are biometrics (subject of Ph.D. dissertation, defended at Universidad Politécnica de Madrid, 2005), statistical pattern recogntion, and image and speech processing. His previous activities regarding biometrics were developed at Biometrics research Lab ATVS (Universidad Autonoma de Madrid).

To: jartacho@unizar.es
Subject: TBME-00363-2005.R1 - Review complete
From: principe@bme.cnel.ufl.edu
Cc: julie@cnel.ufl.edu, gao@ece.ufl.edu
Date: Sat, 29 Jul 2006 22:22:03 -0400 (EDT)
X-Mail-Scanned: Criba 2.0 + Clamd & Spamassassin

Jul 29, 2006

Re: A feasible application of constrained optimization in the IMRT system
id: TBME-00363-2005.R1

Dear Dr. Artacho Terrer

Congratulations. Your manuscript, cited above in abbreviated form, has been reviewed by expert referees, and the Associate Editor coordinating the review suggests ACCEPTANCE. I have enclosed their comments so that you may fine tune the manuscript

even further. If there is no comments, you can consider the paper accepted as is.

The next step is to prepare your paper for final submission. I would like to receive at my office in Gainesville, as soon as possible but no longer than 30 days from today, the following material from you:

-3 paper copies of manuscript in double-space, single-column format
-1 “publication-ready” copy of manuscript (single-space, double-column format).
-electronic version of manuscript (CD or disk). Go to http://www.ieee.org/organization/pubs/transactions/information.htm and click on “Guidelines for Author-Supplied Electronic Text and Graphics” for guidelines on the format of the graphics and text.
-1 author photo/biography kit (if accepted as regular paper)
-Page Charges form (http://bme.cnel.ufl.edu/forms.htm)
-Color Consent form (http://bme.cnel.ufl.edu/forms.htm)
-Author Consent form (if more than one author)
(http://bme.cnel.ufl.edu/forms.htm)
-IEEE Copyright form (http://bme.cnel.ufl.edu/forms.htm)

ONLINE PREPRINT
TBME is has started to put preprint versions of accepted papers on the IEEE Xplore website. These are author formatted preprints that will be uploaded very quickly (within two weeks of receipt at IEEE) to Xplore and which will be automatically replaced by the final version once formatting is completed and the issue is published. Hence, your paper can be accessed by the community much earlier, while the paper issue is being prepared.

If you want your paper to appear as a preprint, please create a fully formatted (2 column, IEEE Transactions style) version in PDF format. For a template go to:
http://www.ieee.org/organizations/pubs/transactions/information.htm
The following copyright notice must be present on the first page (bottom left footnote area):
"Copyright (c) 2006 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org."

Please name the file: TBME-number-preprint.pdf (example: TBME-00001-2005-R3-preprint.pdf).

Send this file along with the materials listed above.

The sooner you send these materials to me the sooner your paper

will be published. The publication ready version of the paper indicates your willingness to pay page charges if the paper is longer than 7 pages (3 pages if a Communications), and lets us (and you) estimate the length of the final manuscript. If you have any questions please write to me at tbme@bme.cnel.ufl.edu.

We value your feedback on our review process. Please take a moment and fill out the survey located at the following web site: http://bme.cnel.ufl.edu/survey/authors.html

Once again congratulations. I appreciate your selection of the Transactions on Biomedical Engineering to publish your work, and

I look forward to receiving the final version of your manuscript.

Sincerely yours,

Jose C. Principe
Editor-in-Chief
IEEE Transactions on Biomedical Engineering

Mailing address
==========
Dr. Jose C. Principe
Editor in Chief - TBME
Electrical & Computer Engineering
PO Box 116130
University of Florida
Gainesville, FL 32611
USA

Note: if any of the reviews appear to be missing, please login to Manuscript Central to check. If a reviewer submitted the

review as an attachment, you will have to retrieve it there.

Associate Editor comments
==========

Dear Authors,

I'm pleased to inform you that the reviewers are
very happy with the revisions you have made, and have
suggested acceptance of your paper in TBME.

Sincerely,

Jianbo

Reviewer's Comments
==========
Reviewer 1 Comments:
The paper has been significantly improved. I ackowledge the work that has been achieved.

Detailed comments :
1. Karush is still miswritten
2. Ref [12] is incomplete (book title is missing)