INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **51** (2006) 2353–2365 doi:10.1088/0031-9155/51/10/001

A unified approach for inversion problems in intensity-modulated radiation therapy

Yair Censor[1], Thomas Bortfeld[2], Benjamin Martin[2] and Alexei Trofimov[2]

[1] Department of Mathematics, University of Haifa, Mt Carmel, Haifa 31905, Israel
[2] Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

E-mail: yair@math.haifa.ac.il

Received 20 April 2005, in final form 16 February 2006
Published 27 April 2006
Online at stacks.iop.org/PMB/51/2353

Abstract

We propose and study a unified model for handling dose constraints (physical dose, equivalent uniform dose (EUD), etc) and radiation source constraints in a single mathematical framework based on the split feasibility problem. The model does not impose on the constraints an exogenous objective (merit) function. The optimization algorithm minimizes a weighted proximity function that measures the sum of the squares of the distances to the constraint sets. This guarantees convergence to a feasible solution point if the split feasibility problem is consistent (i.e., has a solution), or, otherwise, convergence to a solution that minimally violates the physical dose constraints and EUD constraints. We present computational results that demonstrate the validity of the model and the power of the proposed algorithmic scheme.

1. Introduction

In intensity-modulated radiation therapy (IMRT) (see, e.g., Palta and Mackie (2003)), beams of penetrating radiation are directed at the tumour lesion from external sources. A multileaf collimator (MLC) is used to split each beam into many beamlets with individually controllable intensities. There are two principal aspects of radiation teletherapy that call for computational modelling. The first is the calculation of the radiation dose absorbed in the irradiated tissue based on a given distribution of beamlet intensities. This dose calculation is a *forward problem*. The second aspect is the *inverse problem* of the first: to find a distribution of radiation intensities (radiation intensity map) deliverable by all beamlets, which would result in a clinically-acceptable dose distribution (i.e., such that the dose to each tissue would be within the desired *upper* and *lower bounds*, which are prescribed on the basis of medical diagnosis, knowledge and experience). To be of practical value, however, this radiation intensity map must be *implementable*, in a clinically acceptable form, on the available treatment machine. Therefore, in addition to the physical and biological parameters of the irradiated object,

 © 2006 IOP Publishing Ltd Printed in the UK

the relevant information about the capabilities and specifications of the available treatment machine (i.e., radiation source) should be taken into account.

The concept of *equivalent uniform dose* (EUD) was introduced in recent studies to describe dose distributions with a greater clinical relevance. EUD is defined for tumours as the biological equivalent dose that, if given uniformly, will lead to the same cell-kill in the tumour volume as the actual non-uniform dose distribution. EUD can also be defined so as to be applicable for normal tissues. The use of EUD had been proposed for treatment planning in Brahme (1984), Kwa *et al* (1998), Mohan *et al* (1992), Niemierko (1997). In this paper, we use the concept of generalized EUD (Niemierko 1999).

EUD-based IMRT optimization was described by Wu *et al* (2002) and Thieke *et al* (2003). The latter proposed to formulate EUD as constraints instead of optimizing the EUD functions. Motivated by their work we develop and study here a unified theory that enables treatment of EUD constraints and physical dose constraints in a combined manner. The model can be extended to include other kinds of dose constraints, such as dose–volume constraints (Bortfeld *et al* 1997, Spirou and Chui 1998).

A *split feasibility problem* was first proposed and analysed by Censor and Elfving (1994), in the formulation:

$$\text{find } x \in C \subseteq R^m \qquad \text{such that} \quad Ax \in Q \subseteq R^m \tag{1}$$

where C and Q are two given closed convex sets in the m-dimensional vector space R^m, and $A : R^m \to R^m$ is a linear bijection (i.e., a one-to-one and onto mapping) represented by a full-rank matrix. The algorithm proposed there relied on inversion of the matrix A and was, therefore, unrealistic for large matrix dimensions. Later, Byrne (2002) developed the 'CQ-algorithm' for this problem, which does not call for any matrix inversion.

In this paper, we present a framework that enables a unified approach for handling dose constraints and radiation source constraints in intensity-modulated radiation therapy. The main feature of the approach is the rigorous formulation and solution of the feasibility model in a situation where the constraint sets are split between two Euclidean vector spaces: the space of beamlet intensity vectors and the space of dose vectors. Our split feasibility problem is more complicated than that given by (1) and calls for a generalization that will allow finite families of sets in each of the two relevant spaces. Therefore, we call this framework the *multiple-sets split feasibility* approach.

The mathematical analysis of the results presented here is included in our companion paper (Censor *et al* 2005).

2. Formulating the unified model

Most clinical constraints are naturally described as constraints on the dose delivered to the patient. For example, upper and lower dose bounds on each tissue can be described as minimum and maximum dose constraints on the dose delivered to each voxel. Similarly, EUD constraints or dose–volume constraints are most naturally described in the dose space. In contrast, constraints on the deliverable radiation intensities are best described in the intensity space. The universal example of this is the requirement of non-negative beamlet intensity. Additional constraints can be described such as limits on the complexity of the intensity map or limits on the acceptable delivery time.

It is possible to translate constraints formulated in the dose space into constraints in the intensity space. Unfortunately, the resulting constraints tend to be more complicated than the original ones. Thus, it is more convenient to leave the constraints in their respective spaces and develop a framework for combining them. Each constraint is described by a set of either

dose vectors or intensity vectors that fulfil that constraint. A satisfactory treatment plan would be that which is in the intersection of all such sets.

Projection algorithms are frequently used to find a point that belongs to the intersection of closed convex sets. A projection onto a set is the point within that set that is closest to the current point. In projection algorithms, a point is repeatedly projected onto each set, according to some algorithmic scheme, until a point in all of the sets is found.

Unfortunately, for most IMRT inverse problems, there are no possible treatment plans that satisfy all of the constraints. Thus, we need to find plans that are as close to satisfying the constraints as possible. To judge this, we introduce proximity functions that measure the distance of a point to the set. We can still use projection algorithms for this case, with some modification. We will seek to minimize a weighted sum of the proximity functions.

Let us first define the notation. We divide the entire volume of the patient into I voxels, enumerated by $i = 1, 2, \ldots, I$. Assume that T anatomical structures have been outlined, including planning target volumes (PTVs) and the organs at risk (OAR). We denote the set of voxels indices in structure t by S_t. Individual voxels i may belong to several sets S_t, i.e., different structures may overlap. We will further assume that the radiation is delivered independently from each of the J beamlets, which are arranged in a certain geometry and indexed by $j = 1, 2, \ldots, J$. The intensities x_j of the beamlets are arranged in a J-dimensional vector $x = (x_j)_{j=1}^{J} \in R^J$, in the J-dimensional Euclidean space R^J—the radiation intensity space.

The quantities $d_{ij} \geqslant 0$, which represent the dose absorbed in voxel i due to radiation of unit intensity from the jth beamlet (Censor *et al* 1988a, Bortfeld *et al* 1993), are calculable by any forward calculation program. Let h_i denote the total dose absorbed in voxel i and let $h = (h_i)_{i=1}^{I}$ be the vector of doses absorbed in all voxels. We call the space R^I the dose space. We can now calculate h_i as

$$h_i = \sum_{j=1}^{J} d_{ij} x_j. \tag{2}$$

The *dose influence matrix* $D = (d_{ij})$ is the $I \times J$ matrix whose elements are the d_{ij}'s mentioned above. Thus, (2) can be rewritten as the vector equation

$$h = Dx. \tag{3}$$

Finally, let us assume that we have M constraints in the dose space and N constraints in the intensity space. Let H_m be the set of dose vectors that fulfil the mth dose constraint, and let X_n be the set of beamlet intensity vectors that fulfil the nth intensity constraint. Let us consider some concrete examples of the sets H_m and X_n. Each of the constraint sets H_m and X_n can be one of the specific H and X sets, respectively, described below.

In the dose space, a typical constraint is that, in a given critical structure S_t, the dose should not exceed an upper bound u_t. The corresponding set $H_{\max,t}$ is

$$H_{\max,t} = \{h \in R^I \mid h_i \leqslant u_t, \text{ for all } i \in S_t\}. \tag{4}$$

Similarly, in the target volumes (TVs), the dose should not fall below a lower bound l_t. The set $H_{\min,t}$ of dose vectors that fulfil this constraint is

$$H_{\min,t} = \{h \in R^I \mid l_t \leqslant h_i, \text{ for all } i \in S_t\}. \tag{5}$$

To handle EUD constraints for each volume of interest S_t, consisting of N_t voxels, a real-valued function $E_t : R^I \to R$, called the EUD function, is defined as

$$E_t(h) = \left(\frac{1}{N_t} \sum_{i \in S_t} (h_i)^{\alpha_t} \right)^{1/\alpha_t}. \tag{6}$$

The parameter α_t is a tissue-specific number which is negative for target volumes and positive for OAR. For $\alpha_t = 1$, the EUD function is the mean dose of the organ for which it is calculated. On the other hand, letting $\alpha_t \to \infty$ makes the EUD function approach the maximal value: $\max\{h_i \mid i \in S_t\}$. The EUD constraint for an upper EUD bound e_t for a structure S_t can be described by the set

$$H_{\mathrm{EUD},t} = \{h \in R^I \mid E_t(h) \leqslant e_t\}. \tag{7}$$

Lower EUD bounds can be described similarly.

Choi and Deasy (2002) (theorem 1) showed that, due to non-negativity of the dose, $h \geqslant 0$, the EUD function of (6) is convex for all $\alpha_t \geqslant 1$ and concave for all $\alpha_t \leqslant 1$. Therefore, the constraint sets $H_{\mathrm{EUD},t}$ are always convex sets in the dose vector space, since they are level sets (i.e., sets on which the function values are smaller or equal to some fixed real constant) of the convex functions $E_t(h)$ for OAR (with $\alpha_t \geqslant 1$), or of the convex functions $-E_t(h)$ for targets (with $\alpha_t < 0$).

In the radiation intensity space, the most prominent constraint is the non-negativity of the intensities, described by the set

$$X_+ = \{x \in R^J \mid x_j \geqslant 0, \text{ for all } j = 1, 2, \ldots, J\}. \tag{8}$$

Thus, we have a multiple-sets split feasibility problem, where some constraints (the non-negativity of radiation intensities) are defined in the radiation intensity space R^J and other constraints (the upper and lower bounds on dose and the EUD constraints) are defined in the dose space R^I, and the two spaces are related by a linear transformation D (according to (3)). The *unified problem* can be formulated as follows:

$$\text{find } x^* \in X_+ \cap \left(\cap_{n=1}^N X_n\right) \qquad \text{such that} \quad h^* = Dx^* \quad \text{and} \quad h^* \in \left(\cap_{m=1}^M H_m\right). \tag{9}$$

3. Solving the split feasibility problem

3.1. Projections

Next, we describe the operators that are used to project a point to the nearest point that fulfils the constraint, i.e., belongs to the set H_m or X_n. We refer to projection operators in the dose and intensity spaces as P_{H_m} and P_{X_n}, respectively. (Projection algorithms and general projection operators are discussed in detail in the appendix.)

Let us look at some examples of projection operators in the dose space first. In critical structures, it is often required that no voxel should receive a dose above a certain tolerance bound, i.e., a constraint H_t of the form (4) is satisfied. In this case the corresponding projection operator $P_{H_{\max,t}}(h)$ simply cuts off doses beyond u_t in a given structure S_t. The ith component of the projection, i.e., the dose at the ith voxel, is given for all $i = 1, \ldots, I$ by

$$P_{H_{\max,t}}(h)_i = \begin{cases} u_t & \text{if } i \in S_t \text{ and } u_t < h_i \\ h_i & \text{otherwise.} \end{cases} \tag{10}$$

In a target volume, on the other hand, it is requested that the dose should be above the prescription dose l_t, i.e., for all i:

$$P_{H_{\min,t}}(h)_i = \begin{cases} l_t & \text{if } i \in S_t \text{ and } l_t > h_i \\ h_i & \text{otherwise.} \end{cases} \tag{11}$$

Approximate EUD projectors $P_{H_{\mathrm{EUD},t}}(h)$ are derived as in Thieke *et al* (2003), for the convex constraints $H_{\mathrm{EUD},t}$ (given by equation (7)).

In the intensity space, for the non-negativity constraint X_+ (equation (8)), the corresponding projection operator cuts off negative intensities as follows:

$$P_{X_+}(x)_j = \begin{cases} 0 & \text{if } x_j < 0 \\ x_j & \text{otherwise.} \end{cases} \tag{12}$$

Another example of a projector in the intensity space is the intensity gradient projector, which is relevant, e.g., in the optimization of particle therapy delivery with continuous beam scanning (Trofimov and Bortfeld 2003). Assuming that the beamlets are numbered in the order of scanning, from the first to the Jth, if the change in intensity between two consecutive beamlets, $(j-1)$ and j, is constrained to a maximum value Δ, then the corresponding gradient projector is

$$P_{X_\nabla}(x)_j = \begin{cases} x_{j-1} + \Delta & \text{if } x_j > x_{j-1} + \Delta \\ x_{j-1} - \Delta & \text{if } x_j < x_{j-1} - \Delta \\ x_j & \text{otherwise.} \end{cases} \tag{13}$$

Multileaf collimator constraints for IMRT delivery can be enforced in a similar manner.

In practice, it is often impossible to fulfil all constraints simultaneously. In this infeasible case, it makes sense to find a solution that is as close as possible to a feasible solution. Using the square Euclidean norm as a closeness measure, a suitable objective function, called proximity function, is

$$F(x) = \frac{1}{2}\sum_{m=1}^{M} w_{H_m} \left\| P_{H_m}(Dx) - Dx \right\|^2 + \frac{1}{2}\sum_{n=1}^{N} w_{X_n} \left\| P_{X_n}(x) - x \right\|^2 \tag{14}$$

where each of the projectors P can be one of the sample projection operators mentioned above, and $\{w_{H_m}\}_{m=1}^{M}$, $\{w_{X_n}\}_{n=1}^{N}$ are their corresponding weight factors. Note that there may be multiple constraints, i.e., multiple projection operators, per anatomical structure. For example, in the target volumes one usually aims to keep the dose above a lower bound l_t, but at the same time below an upper bound u_t, often with various degrees of priority (expressed by weights w_{H_m}, i.e., penalties for under- and overdose).

3.2. Algorithm

We seek an x^* that minimizes the proximity function (equation (14)). We do that by applying the gradient ∇_x to $F(x)$ and using a result from Aubin and Cellina (1984), which states that

$$\nabla_x \left\| P_{X_n}(x) - x \right\|^2 = 2\big(P_{X_n}(x) - x\big). \tag{15}$$

Using the chain rule we calculate the gradient with respect to the dose projectors:

$$\nabla_x \left\| P_{H_m}(Dx) - Dx \right\|^2 = 2D^{\mathrm{T}}\big(P_{H_m}(Dx) - Dx\big) \tag{16}$$

where D^{T} is the transposed matrix.

The general iterative gradient projection scheme, with the stepsize s, designed to find a minimum of $F(x)$ subject to $x \in \Omega$, where $\Omega \subseteq R^J$ is some constraint set, whose projector is P_Ω, is

$$x^{(k+1)} = P_\Omega(x^{(k)} - s\nabla_x F(x^{(k)})). \tag{17}$$

Inserting equations (15) and (16), and using $P_\Omega = P_{X_+}$, yields the algorithm:

$$x^{(k+1)} = P_{X_+}\left[x^{(k)} - s\left(D^{\mathrm{T}}\sum_{m=1}^{M} w_{H_m}\big(P_{H_m}(Dx^{(k)}) - Dx^{(k)}\big) + \sum_{n=1}^{N} w_{X_n}\big(P_{X_n}(x^{(k)}) - x^{(k)}\big)\right)\right]. \tag{18}$$

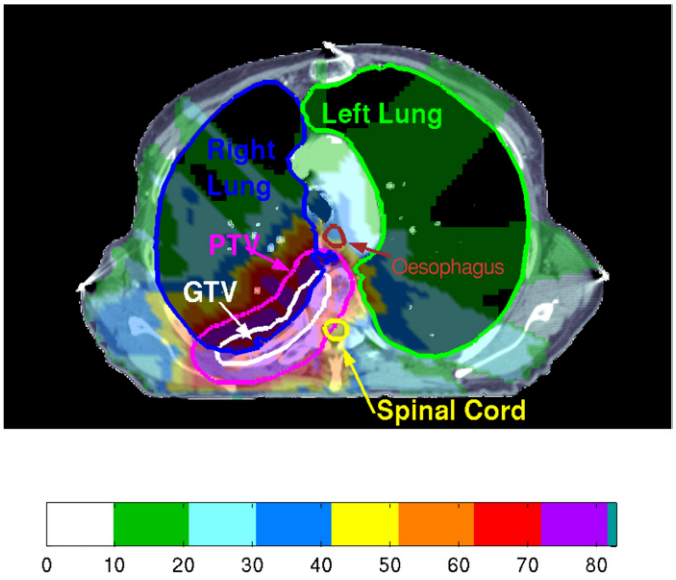

Figure 1. The optimized dose distribution on a transversal CT slice.
(This figure is in colour only in the electronic version)

The projection P_{X_+} must be applied at each iteration step, *after* all other projections, to eliminate the unphysical negative intensities. As shown in Censor *et al* (2005), the algorithm of equation (18) generates a sequence $\{x^{(k)}\}$ of intensity vectors that will converge towards x^* if the stepsize s is chosen appropriately.

The algorithm (18) is similar to that of Thieke *et al* (2003). However it does not involve scaling of the gradient with second derivatives, and is formulated for a more general case.

3.3. Test case

To demonstrate the practical usefulness of our algorithm, we applied it to a clinical case of a tumour in the thorax. Figure 1 shows the geometry of the gross tumour volume (GTV), the planning target volume and the surrounding critical structures in a transversal CT slice. A thorax case was chosen because it involves serially organized organs such as the spinal cord and oesophagus, and parallel organs such as the lung. Dose prescriptions and constraints were given as lower and/or upper dose bounds for the target volume and the serially organized critical structures. These upper and lower dose bounds, as well as the weights of importance w_H, are summarized in table 1. The weights of importance were normalized by the number of voxels in the respective organ. Clearly, because the target lower dose bound was set equal to the upper dose bound, there is no feasible solution in this case. For the lung, an EUD bound was used. The lung constraints are displayed in table 2.

The irradiation geometry was a coplanar beam arrangement with gantry angles of 110°, 180°, 220°, 260° and 320° (clockwise). The angles were taken from the clinical treatment plan with which the patient was actually treated. The beam energy was 6 MV. There was a total of $125 \times 75 \times 82 = 768,750$ voxels, measuring $2.92 \times 2.5 \times 2.92$ mm^3 each. The number of beamlets was between 225 and 336 per beam (1460 in total) and their size was 5×5 mm^2. The dose operator D was pre-calculated with the KonRad program (Preiser *et al* 1997). Specifically, the d_{ij} matrix elements were calculated with a pencil beam algorithm (Bortfeld

Table 1. Physical dose prescriptions (lower and upper dose bounds l_t and u_t) and their corresponding weights of importance (w_H) used for the example case.

Volume	Lower (Gy)	Weight	Upper (Gy)	Weight
PTV	72	1000	72	200
Oesophagus	–	–	60	30
Spinal cord	–	–	50	1000
Heart	–	–	60	30

Table 2. Equivalent uniform dose bounds with their α_t parameters (see the EUD definition in (6)), and the corresponding weights of importance (w_H).

Volume	EUD (Gy)	α	Weight
Right lung	15	1	50
Left lung	15	1	50

et al 1993, Thieke *et al* 2003) and stored in a file. Our algorithm, given by equation (18), was implemented in a dedicated temporo-spatial radiotherapy in-house optimization tool, called 'opt4D' (Trofimov *et al* 2005). For the present application, 'opt4D' was used in the static (three-dimensional) mode. The projection operator $P_{H_{\mathrm{EUD}}}$ onto each EUD constraint set was implemented using the method of Thieke *et al* (2003).

The stepsize s was calculated as follows: the iteration was initialized with $x^{(0)} = 0$ and the first iteration was performed with a stepsize $s = 1$. This resulted in a vector of beamlet intensities $x^{(1)}$ and its corresponding dose vector $Dx^{(1)}$. Then a normalization factor κ was defined such that the average target dose of $\kappa Dx^{(1)}$ was identical to the prescribed target dose. The purpose of this exercise was to put the dose range in the right order of magnitude. The vector of intensities $x^{(1)}$ was then multiplied by κ and all subsequent iterations were performed with a stepsize of $s = \kappa$. To test the convergence behaviour with different stepsizes s, we performed iterative algorithmic runs with $s = 2\kappa$, $s = 4\kappa$ and so forth.

The heuristic stopping criterion was that the relative difference between the proximity functions of two consecutive iteration steps should be below 0.2%. The convergence behaviour of the proximity function (14), for different stepsizes s, is plotted in figure 2. For $s = \kappa$ it took 85 iteration steps to reach the stopping criterion. In this specific case, it turned out that $s = 2\kappa$ resulted in the fastest convergence, in 65 steps, without leading to oscillations. The calculation time per iteration step was 7.2 s on a 2.4 GHz Pentium 4 PC running Linux.

The resulting dose distribution is shown superimposed on a transversal CT slice in figure 1, and in the form of a dose–volume histogram in figure 3.

4. Discussion

We used physical dose constraints, non-negativity of beamlets' intensities and equivalent uniform dose constraints to describe the proposed framework and demonstrate its computational viability. These constraints are formulated in two different Euclidean vector spaces. Some constraints, such as the physical dose constraints, can be formulated in either the space of radiation intensity vectors or of dose vectors (as in equations (4), (5) and (A.7)). However, the EUD constraints are formulated exclusively in the space of dose vectors. Similarly, the non-negativity constraints on the radiation intensities are formulated exclusively in the radiation intensity space. Thus, there is no escape from addressing constraints that reside in both of these different spaces. Additionally, these constraints have a *different mathematical*

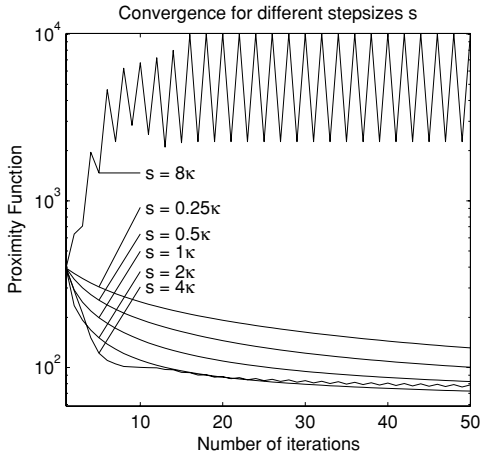

Figure 2. The value of the proximity function (14) versus the iteration number for different stepsizes s. The definition of κ is given in subsection 3.3. For $s = \kappa$ it took 85 iterations to reach the 0.2% relative difference stopping criterion.

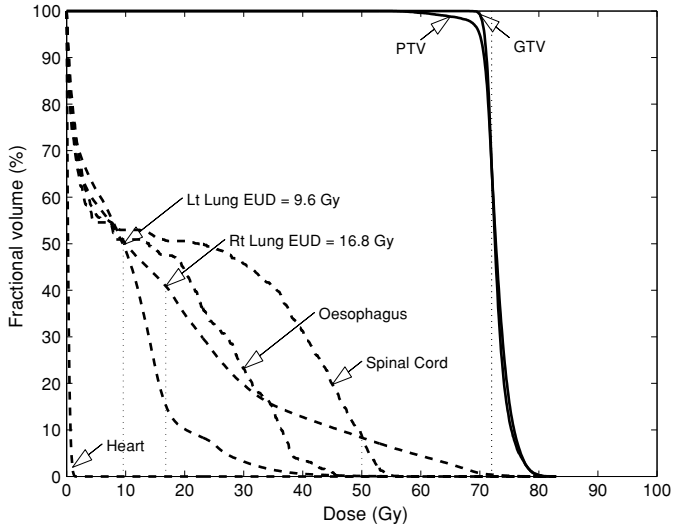

Figure 3. The dose–volume histograms for the test case described in subsection 3.3.

nature, namely, physical dose constraints are linear while EUD constraints are convex but nonlinear. To accommodate such constraints we developed a logical framework for performing a search that iterates in each of the two spaces and correctly passes back and forth between the spaces, without resorting to the inverse matrix of D.

More generally, our framework fits any situation in which an inversion problem in IMRT can be presented by constraint sets that are split between the beamlet intensity space and the dose space. For example, the multileaf collimator (MLC) hardware constraints that are integrated into the IMRT inversion problem by Cho and Marks (2000) could be included. The unified new model relies on the multiple-sets split feasibility problem formulation, that we proposed in Censor *et al* (2005), which is further developed here to accommodate the specific

IMRT situation and to establish a valid framework for handling both physical dose constraints and EUD constraints in a unified manner.

Our unified framework and algorithm also cover the situation in which the constraints are *inconsistent*, i.e., it is impossible to satisfy all of them. In such a case, our simultaneous iterative projections algorithm minimizes a proximity function that measures how close the constraints are being satisfied. The original split feasibility problem of Censor and Elfving (1994) addressed only the consistent case and the iterative algorithm proposed there employed the inverse D^{-1} of the dose matrix D that maps the radiation intensity space to the dose space. This is a practical disadvantage since the inversion of a large matrix is a computationally difficult task. However, recently Byrne (2002) refined the algorithms of Censor and Elfving (1994) so that only the transposed matrix D^{T} is needed, and not D^{-1}. The existing algorithms for the split feasibility problem, which currently handle only a single constraint set in each of the spaces, had to be properly modified to handle multiple constraints.

A further useful modification can be implemented here by replacing, in (14) and (18), the *orthogonal projections* onto the EUD constraint sets by *subgradient projections* (see, e.g., Censor and Lent (1982) or Censor and Zenios (1997) (subsection 5.3)). These projections do not require the (iterative) minimization of distance between the point and the set but are rather given by closed-form analytical expressions. In the vector space of radiation intensities, this would make no difference because the physical dose constraints there are linear, but in the dose vector space the EUD constraints are nonlinear and, thus, the algorithm benefits from replacing orthogonal projections by subgradient projections because the latter are easier to compute. Subgradient projections are discussed in more detail in the appendix, where examples of relevant operators are also given.

Treating physical dose and EUD as constraints, instead of optimizing respective functions, has been previously proposed, thus creating what we now call a 'split feasibility problem'. In our approach, this idea is cast into a unified framework based on a rigorous foundation. The multiple-sets split feasibility approach applies to any inversion problem, in its fully-discretized formulation, where the constraints are split between two different vector spaces that are related by a linear transformation.

Acknowledgments

We thank Dr John Wolfgang and Dr Noah Choi of Massachusetts General Hospital for providing the clinical example case. This research is supported by grant no 2003275 from the United States–Israel Binational Science Foundation (BSF). The work of Y Censor on this research was also supported by a National Institutes of Health (NIH) grant no HL70472. Part of this work was done at the Center for Computational Mathematics and Scientific Computation (CCMSC) at the University of Haifa and supported by research grant no 522/04 from the Israel Science Foundation (ISF). T Bortfeld and A Trofimov acknowledge support from the National Cancer Institute programme project grant no 5-P01-CA21239-25 and grant no 1-R01-CA103904-01. This work was supported in part by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821).

Appendix. Projection algorithms

Projection algorithms employ projections onto convex sets in various ways. They may use different kinds of projections and, sometimes, even use different projections within the same

algorithm. They serve to solve a variety of problems which are either of the feasibility or the optimization types. They have different algorithmic structures, of which some are particularly suitable for parallel computing, and they demonstrate nice convergence properties and/or good initial behaviour patterns. This class of algorithms has witnessed great progress in recent years and its member algorithms have been applied with success to fully-discretized models of problems in image reconstruction and image processing (Bauschke and Borwein 1996, Censor and Zenios 1997, Stark and Yang 1998).

The convex feasibility problem is a fundamental problem in many areas of mathematics and the physical sciences (see, e.g., Combettes (1994) and references therein). It has been used to model significant real-world problems in image reconstruction from projections (Herman 1980), in radiation therapy treatment planning (Censor *et al* 1988b, Censor 2003), and in crystallography (Marks *et al* 1999), to name but a few, and has been used under additional names such as *set theoretic estimation* or the *feasible set approach*. A common approach to such problems is to use projection algorithms (see, e.g., Bauschke and Borwein (1996)), which employ *orthogonal projections* (i.e., nearest point mappings) onto the individual closed convex constraint sets C_i in an n-dimensional vector space R^n. The orthogonal projection $P_\Omega(\hat{x})$ of a point $\hat{x} \in R^n$ onto a closed convex set $\Omega \subseteq R^n$ is defined by

$$P_\Omega(\hat{x}) = \operatorname{argmin}\{\|\hat{x} - x\| \mid x \in \Omega\} \tag{A.1}$$

where $\|\cdot\|$ is the Euclidean norm in R^n. Frequently a *relaxation parameter* is introduced so that

$$P_\Omega(\hat{x}, \lambda) = (1 - \lambda)\hat{x} + \lambda P_\Omega(\hat{x}) \tag{A.2}$$

is the *relaxed projection* of $\hat{x}$ onto Ω with relaxation λ. Orthogonal projections are easy to calculate for linear sets such as hyperplanes, half-spaces, hyperslabs, or for balls, etc. But for a general (nonlinear) convex set, finding the orthogonal projection of a given point requires the solution of a subsidiary constrained optimization problem of equation (A.1) to minimize the distance of the given point to the set over all points in the set. Therefore, it is useful to employ *subgradient projections* instead. Let the convex set be given as a level-set of a convex function $f : R^n \to R$, i.e.,

$$\Omega = \{x \in R^n \mid f(x) \leqslant 0\} \tag{A.3}$$

and let $\hat{x} \in R^n$ be a given point. Then the *subgradient projection of* $\hat{x}$ *onto* the set Ω of (A.3) is the point $\Pi_\Omega(\hat{x})$ given by

$$\Pi_\Omega(\hat{x}) = \begin{cases} \hat{x} - \dfrac{f(\hat{x})}{\|g(\hat{x})\|^2} \cdot g(\hat{x}) & \text{if } \hat{x} \notin \Omega \\ \hat{x} & \text{if } \hat{x} \in \Omega \end{cases} \tag{A.4}$$

where $g(\hat{x})$ is a subgradient vector of the function f calculated at $\hat{x}$. Note that, according to the traditional definition, this is not a projection, because $\Pi_\Omega(\hat{x})$ does not even need to belong to Ω. Nevertheless, this is a very useful type of 'projection', because it has a closed-form formula and does not require one to solve an optimization problem as required in equation (A.1). For the notion of subgradients of convex function see, e.g., Hiriart-Urruty and Lemaréchal (2001), Boyd and Vandenberghe (2003). For subgradient projection algorithms for feasibility problems see, e.g., Censor and Zenios (1997) (section 5.3) or Bauschke and Borwein (1996).

It is well known that if the convex function f is differentiable at the point $\hat{x}$ then it has a unique subgradient there which is equal to the gradient $\nabla f(\hat{x})$. The relaxed subgradient projection with relaxation parameter λ may also be introduced by

$$\Pi_\Omega(\hat{x}, \lambda) = (1 - \lambda)\hat{x} + \lambda \Pi_\Omega(\hat{x}). \tag{A.5}$$

Projection algorithmic schemes for the convex feasibility problem are, in general, either *sequential* or *simultaneous* or *block-iterative* (see, e.g., Censor and Zenios (1997) for a classification of projection algorithms into such classes, and the review paper of Bauschke and Borwein (1996) for a variety of specific algorithms of these kinds). Recently, Censor and Tom (2003) have investigated yet another class of projection algorithms, called *string-averaging* algorithms.

A.1. Projections onto a generalized box

The projection $P_H(h)$, in our algorithm, is easy to perform since H is a generalized box (i.e., is the product of all individual H_m). For all $i = 1, 2, \ldots, I$,

$$P_H(h)_i = \begin{cases} l_i & \text{if } h_i < l_i \\ h_i & \text{if } l_i \leqslant h_i \leqslant u_i \\ u_i & \text{if } u_i < h_i. \end{cases} \tag{A.6}$$

Note that if the physical dose constraints H_m (e.g., equations (4) and (5)) were defined in the intensity space R^J by the sets

$$X_i = \{x \in R^J \mid l_i \leqslant (Dx)_i \leqslant u_i\} \tag{A.7}$$

then the projections $P_{X_i}(x)$ would have to be calculated as

$$P_{X_i}(x) = \begin{cases} x + \dfrac{u_i - (Dx)_i}{\|d^i\|^2} \cdot d^i & \text{if } u_i < (Dx)_i \\ x & \text{if } l_i \leqslant (Dx)_i \leqslant u_i \\ x + \dfrac{l_i - (Dx)_i}{\|d^i\|^2} \cdot d^i & \text{if } (Dx)_i < l_i \end{cases} \tag{A.8}$$

where $d^i = (d_{ij})_{j=1}^J$ is the J-dimensional vector that forms the ith column of the dose matrix D.

A.2. Subgradient projections for the EUD constraints

The subgradient projections $\Pi_H(h)$ for the EUD constraints are calculated from (6), (7) and (A.4) as follows. The EUD function of (6) is differentiable and the ith component of its gradient $\nabla E_t(h)$ is calculated, for all $i \in S_t$, by

$$\begin{aligned} (\nabla E_t(h))_i &= \frac{1}{\alpha_t} \left(\left(\frac{1}{N_t} \right) \sum_{i \in S_t} (h_i)^{\alpha_t} \right)^{(1/\alpha_t)-1} \cdot \frac{1}{N_t} \alpha_t (h_i)^{\alpha_t - 1} \\ &= \left(\frac{1}{N_t} \right)^{1/\alpha_t} \left(\sum_{i \in S_t} (h_i)^{\alpha_t} \right)^{(1/\alpha_t)-1} \cdot (h_i)^{\alpha_t - 1}. \end{aligned} \tag{A.9}$$

To obtain $\Pi_H(h)$ for a target volume (TV), with $\alpha_t < 0$, we rewrite (7) as

$$H_{\mathrm{EUD},t}^{(\mathrm{TV})} = \{h \in R^I \mid -E_t(h) \leqslant -e_t\} \tag{A.10}$$

which makes it a level set of the convex function $-E_t(h)$. Now we calculate $\Pi_{H_{\mathrm{EUD},t}}^{(\mathrm{TV})}(h^{(t)})$ for individual target(s), and this is done by using (A.4) with $x = h^{(t)}$, $f(h^{(t)}) = -E_t(h^{(t)}) + e_t$, and with $g(h^{(t)}) = \nabla(-E_t(h^{(t)}))$, so that

$$\Pi_{H_{\mathrm{EUD},t}}^{(\mathrm{TV})}(h^{(t)}) = \begin{cases} h^{(t)} + \dfrac{E_t(h^{(t)}) - e_t}{\|\nabla(-E_t(h^{(t)}))\|^2} \cdot \nabla(-E_t(h^{(t)})) & \text{if } h^{(t)} \notin H_{\mathrm{EUD},t}^{(\mathrm{TV})} \\ h^{(t)} & \text{if } h^{(t)} \in H_{\mathrm{EUD},t}^{(\mathrm{TV})}. \end{cases} \tag{A.11}$$

Similarly, for an organ at risk (OAR, with $\alpha_t \geqslant 1$):

$$H_{\mathrm{EUD},t}^{(\mathrm{OAR})} = \{h \in R^I \mid E_t(h) \leqslant e_t\} \tag{A.12}$$

and $\Pi_{H_{\mathrm{EUD},t}}^{(\mathrm{OAR})}(h^{(t)})$ is obtained by plugging $x = h^{(t)}$, $f(h^{(t)}) = E_t(h^{(t)}) - e_t$, and $g(h^{(t)}) = \nabla E_t(h^{(t)})$, into (A.4):

$$\Pi_{H_{\mathrm{EUD},t}}^{(\mathrm{OAR})}(h^{(t)}) = \begin{cases} h^{(t)} - \dfrac{E_t(h^{(t)}) - e_t}{\|\nabla E_t(h^{(t)})\|^2} \cdot \nabla E_t(h^{(t)}) & \text{if } h^{(t)} \notin H_{\mathrm{EUD},t}^{(\mathrm{OAR})} \\ h^{(t)} & \text{if } h^{(t)} \in H_{\mathrm{EUD},t}^{(\mathrm{OAR})} \end{cases} \tag{A.13}$$

which turns out equivalent to (A.11).

References

Aubin J P and Cellina A 1984 *Differential Inclusions: Set-Valued Maps and Viability Theory* (Berlin: Springer) p 24

Bauschke H H and Borwein J M 1996 On projection algorithms for solving convex feasibility problems *SIAM Rev.* **38** 367–426

Bortfeld T, Schlegel W and Rhein B 1993 Decomposition of pencil beam kernels for fast dose calculations in three-dimensional treatment planning *Med. Phys.* **20** 311–8

Bortfeld T, Stein J and Preiser K 1997 Clinically relevant intensity modulation optimization using physical criteria *Proc. 12th ICCR* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 1–4

Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)

Brahme A 1984 Dosimetric precision requirements in radiation therapy *Acta Radiol. Oncol.* **23** 379–91

Byrne C 2002 Iterative oblique projection onto convex sets and the split feasibility problem *Inverse Problems* **18** 441–53

Censor Y 2003 Mathematical optimization for the inverse problem of intensity modulated radiation therapy *Intensity-Modulated Radiation Therapy: The State of The Art* ed J R Palta and T R Mackie (Madison, WI: Medical Physics Publishing) pp 25–49

Censor Y, Altschuler M D and Powlis W D 1988a A computational solution of the inverse problem in radiation therapy treatment planning *Appl. Math. Comput.* **25** 57–87

Censor Y, Altschuler M D and Powlis W D 1988b On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning *Inverse Problems* **4** 607–23

Censor Y and Elfving T 1994 A multiprojections algorithm using Bregman projections in a product space *Numer. Algorithms* **8** 221–39

Censor Y, Elfving T, Kopf N and Bortfeld T 2005 The multiple-sets split feasibility problem and its applications for inverse problems *Inverse Problems* **21** 2071–84

Censor Y and Lent A 1982 Cyclic subgradient projections *Math. Program.* **24** 233–35

Censor Y and Tom E 2003 Convergence of string-averaging projection schemes for inconsistent convex feasibility problems *Optim. Methods Softw.* **18** 543–54

Censor Y and Zenios S A 1997 *Parallel Optimization: Theory, Algorithms, and Applications* (New York, NY: Oxford University Press)

Cho P S and Marks R J II 2000 Hardware-sensitive optimization for intensity modulated radiotherapy *Phys. Med. Biol.* **45** 429–40

Choi B and Deasy J O 2002 The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning *Phys. Med. Biol.* **47** 3579–89

Combettes P L 1994 Inconsistent signal feasibility problems: least-squares solutions in a product space *IEEE Trans. Signal Process.* **42** 2955–66

Herman G T 1980 *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography* (New York, NY: Academic)

Hiriart-Urruty J-B and Lemaréchal C 2001 *Fundamentals of Convex Analysis* (Berlin: Springer)

Kwa S L, Theuws J C, Wagenaar A, Damen E M, Boersma L J, Baas P, Muller S H and Lebesque J V 1998 Evaluation of two dose–volume histogram reduction models for the prediction of radiation pneumonitis *Radiother. Oncol.* **48** 61–9

Marks L D, Sinkler W and Landree E 1999 A feasible set approach to the crystallographic phase problem *Acta Crystallogr.* A **55** 601–12

Mohan R, Mageras G S, Baldwin B, Brewster L J, Kutcher G J, Leibel S, Burman C M, Ling C C and Fuks Z 1992 Clinically relevant optimization of 3-D conformal treatments *Med. Phys.* **19** 933–44

Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10

Niemierko A 1999 A generalized concept of equivalent uniform dose *Med. Phys.* **26** 1100

Palta J R and Mackie T R (ed) 2003 *Intensity-Modulated Radiation Therapy: The State of The Art* (Madison, WI: Medical Physics Publishing)

Preiser K, Bortfeld T, Hartwig K, Schlegel W and Stein J 1997 A new program for inverse radiotherapy planning *Proc. 12th ICCR* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 425–8

Spirou S V and Chui C-S 1998 A gradient inverse planning algorithm with dose–volume constraints *Med. Phys.* **25** 321–33

Stark H and Yang Y 1998 *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics* (New York, NY: Wiley)

Thieke C, Bortfeld T, Niemierko A and Nill S 2003 From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning *Med. Phys.* **30** 2332–9

Trofimov A and Bortfeld T 2003 Optimization of beam parameters and treatment planning for intensity modulated proton therapy *Technol. Cancer Res. Treat.* **2** 437–44

Trofimov A, Rietzel E, Lu H-M, Martin B, Jiang S B, Chen G T Y and Bortfeld T 2005 Temporo-spatial IMRT optimization: concepts, implementation and initial results *Phys. Med. Biol.* **50** 2779–98

Wu Q, Mohan R, Niemierko A and Schmidt-Ullrich R 2002 Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose *Int. J. Radiat. Oncol. Biol. Phys.* **52** 224–35