The Dose-Volume Constraint Satisfaction Problem for Inverse Treatment Planning with Field Segments

Darek Michalski†, Ying Xiao†, Yair Censor‡ and James M. Galvin†

† Department of Radiation Oncology, Kimmel Cancer Center, Jefferson Medical College of Thomas Jefferson University, 111 South 11^{th} Street, Philadelphia, PA, 19107, USA

E-mail: darek.michalski@mail.tju.edu
ying.xiao@mail.tju.edu
james.galvin@mail.tju.edu

‡ Department of Mathematics, University of Haifa, Mt. Carmel, Haifa 31905, Israel

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

Abstract. The prescribed goals of radiation treatment planning are often expressed in terms of dose–volume constraints. We present a novel formulation of a dose–volume constraint satisfaction search for the discretized radiation therapy model. This approach does not rely on any explicit cost function. The inverse treatment planning uses the aperture based approach with predefined, according to geometric rules, segmental fields. The solver utilizes the simultaneous version of the cyclic subgradient projection algorithm. This is a deterministic iterative method designed for solving the convex feasibility problems. A prescription is expressed with the set of inequalities imposed on the dose at the voxel resolution. Additional constraint functions control the compliance with selected points of the expected cumulative dose–volume histograms. The performance of this method is tested on prostate and head–and–neck cases. The relationships with other models and algorithms of similar conceptual origin are discussed. The demonstrated advantages of the method are: the equivalence of the algorithmic and prescription parameters, the intuitive setup of free parameters, the improved speed of the method as compared to similar iterative as well as other techniques. The technique reported here will deliver an approximate solutions for inconsistent prescriptions.

PACS numbers: 87.53.Tf 87.10.+e

Submitted to: *Phys. Med. Biol.*

1. Introduction

The process of treatment planning can be viewed as a sequence of logically and/or functionally separate phases determined by the modeling methodology and medical insight as well as technological restrictions. The main goal is to determine the essential set of parameters that will provide the required spatial dose distribution. This might be quantified with dose–based or biological criteria. In practice the former are usually employed to articulate the goals of the therapeutic radiation. (Deasy, 1997; Shepard et al., 1999; Brahme, 2000; Glatstein, 2001; IMRTCWG, 2001; Deasy et al., 2002). The dose–based criteria also appear to be clinically relevant, at the end of the process, as a main gauge of plan quality expressed in terms of the cumulative dose–volume histograms (DVH) and isodose distributions.

A dose–volume prescription is a conceptual starting point for *inverse* treatment planning (ITP). A useful analogy (Bortfeld et al., 1990; Censor and Zenios, 1997; Kak and Slaney, 2001) with the discretized model of computed tomography (CT) shows another characteristic of ITP. The crucial difference between ITP and the CT is that the latter is a solvable problem of constraint sets that model *a posteriori* a physical phenomenon. The *a priori* character of the ITP nullifies this certainty. Mathematically, ITP belongs to *inverse problems*. They are usually ill–posed and concerned with the determination of the causes of desired effects. Current approaches for ITP usually rely on various optimization techniques that utilize objective functions.

Ideally these functions should reflect treatment goals in an unequivocal manner so that the determination of the relevant parameters optimize the plan quality. However in practice it is difficult to quantify and consolidate all pertinent components for all possible clinical scenarios. Our understanding of the radiation response of many tissues is limited so that formulating meaningful objective function is not straightforward. (Deasy, 1997; Shepard et al., 1999; Brahme, 2000; Glatstein, 2001; IMRTCWG, 2001; Chen et al., 2002). It is also important to consider other aspect like the smoothness of the intensity maps or the optimization of gantry angles (Spirou et al., 2001; Pugachev et al., 2001). Possible non–convexity of the cost functions may give rise to the problem of local minima that might compromise specific clinical objectives. Execution times are also important. Rigorous modeling and implementation of the dose–volume constraints (DVC) in terms of the mixed integer programming (MIP) is a NP–complete problem that might not solve plans in clinically acceptable time (Garey and Johnson, 1979; Deasy, 1997; Bednarz et al., 2000). This points out another issue that the cost functions might be predetermined by a given optimization scheme, *e.g.* the MIP must use only linear cost functions. Thus the very nature of these conditions render the problem a complex multi–objective decision and multi–criteria optimization process with mutually competing and/or incommensurable objectives, and inevitable trade–offs.

As a viable alternative we consider the inverse planning as a constraint satisfaction problem. The determination of the beam intensities quantified by the beam weight vector within this computational context has been examined since the very outset of the development of the fully discretized feasibility approach to the inverse treatment planning (Altschuler and Censor, 1984; Censor et al., 1988b; Censor et al., 1988a).

A novel enhancement of this class of modeling and weight vector computation is examined here. Instead of relying on heuristics, the DVCs are incorporated into the model for a constraint satisfaction search. This computational approach determines the weight vector beyond the paradigm based on the notion of the measure of plan

quality mirrored by an objective function or that of the MIP. In contrast with those approaches, we present a new model that uses an iterative algorithm based on the method of cyclic subgradient projections (CSP) which solves the convex feasibility problem. It is a natural extension of our previous work, currently implemented in our clinic, that uses only dose constraints expressed in terms of interval inequalities. We report the performance study of this new method for aperture–based inverse planning (ABIP). However the model and algorithmic part of the paper are independent of this particular modality which is based on predefined segments. The method can be easily applied to the beamlet–based inverse treatment planning(Xiao et al., 2003b).

2. Materials and Methods

2.1. Aperture–Based Inverse Planning

The methodology of the aperture–based inverse planning has been already described in detail elsewhere (Xiao et al., 2000; Galvin et al., 2000; Bednarz et al., 2000; Xiao et al., 2003a). The first step is the setting–up of apertures to define all relevant beams. For all chosen gantry angles the process starts with conformal apertures that cover combined targets. Additional apertures address boost regions. Further apertures provide maximal coverage of the targets while sparing OARs. Subsequently beam weights are determined.

The APIB method was applied to six prostate and four oropharyngeal cancer cases. The aperture definition was carried out with commercial CMS FOCUS treatment planning system (Computerized Medical Systems, Inc., 2001). Two sets of coplanar gantry angles were used for the prostate cases. Three cases used set $\mathcal{A} = \{45, 90, 135, 225, 270 \text{ and } 315\,\text{deg}\}$, the others used set $\mathcal{B} = \{35, 90, 125, 180, 235, 270 \text{ and } 325\,\text{deg}\}$. As described later, the prostate cases used different dose–volume prescriptions. Nine equally spaced coplanar gantry angles starting from 0 deg were used for head–and–neck cases. For the prostate cases, a 5mm margin surrounded the clinical target volume (CTV) to define the planning target volume (PTV) (IMRTCWG, 2001). An additional 8mm margin was added to accommodate the beam penumbra. Radiation Therapy Oncology Group protocol H–0022 (RTOG–H–0022) was used for the oropharyngeal cancer cases (Eisbruch et al., 2001). The number of apertures depended on the geometry and the topology of a particular site as well as the complexity of the prescription, *e.g.*, the number of boost regions. The prostate cases yielded between 12 and 49 apertures. The head–and–neck cases, a much more anatomically intricate disease site with more demanding dose goals, yielded between 51 and 99 apertures. Subsequently, the patient volume was automatically discretized into a three–dimensional grid of voxels with a user–assigned resolution of 3mm. For a given set of N fields with a weight vector $\mathbf{x} = [x_1, \cdots, x_j, \cdots, x_N]^\mathsf{T}$ (T denotes transpose) and for every structure s the CMS FOCUS dose engine using the scatter integration with heterogeneity correction calculated the unit weight dose matrix, $d_{ij}^{(s)}$, that allows us to compute the dose, $D_i^{(s)}$, delivered to voxel i of a given structure s,

$$D_i^{(s)} = \sum_j^N d_{ij}^{(s)} x_j. \tag{1}$$

These steps concluded preparation for the weight determination. Having been granted access to the source code of the pertinent modules of the CMS Focus treatment

STRUCTURE		underdose	lower reference limits	upper reference limits	overdose
Table A		Volume %	Gy	Gy	Volume %
Bladder			0.00	60.00	20.00
Rectum			0.00	60.00	20.00
Left Femur			0.00	45.00	15.00
Right Femur			0.00	45.00	15.00
PTV	(a)	5.00	73.80	78.00	5.00
	(b)	10.00	75.00		
Table B		Volume %	Gy	Gy	Volume %
Bladder			0.00	65.00	20.00
Rectum			0.00	65.00	20.00
Left Femur			0.00	50.00	15.00
Right Femur			0.00	50.00	15.00
PTV		5.00	75.60	83.16	

Table 1. The DVCs for prostate cases. The cases that used gantry angles assigned from set $\mathcal{A}$ as described in § 2.1 utilized constraints marked by A, and those that used values from set $\mathcal{B}$ utilized constraints B.

planning system, the re–engineered software used our weight determination program.

2.2. Dose–Volume Constraints: Prescriptions

Two disease sites selected for the study differ in anatomical and dose prescription complexity. The oropharyngeal cases are more challenging than the prostate cases in both of these aspects. The latter are interesting because they allow for an intuitive understanding of the dependencies between the parameters during the computation.

The DVCs for prostate cases are listed in Table 1 and for head–and–neck cases in Table 2. Two oropharyngeal cancer cases had four different target regions with three different dose prescription levels (PTV 54, PTV 60 and PTV 66). The other two head–and–neck cases had two boost regions (PTV 54 and PTV 66). We imposed a few types of constraints for the prostate cases denoted by A and B in Table 1. Three cases used constraints A(a) and A(b) to demonstrate their effect on the cumulative DVH of the PTV. The other three cases used constraints B. Lower and upper reference limits mark the desired dose intervals for listed structures. The pertinent structures may be either uderdosed or overdosed with respect to these limits as the prescriptions stipulate with the percentages of organs's volumes listed in the table.

2.3. Dose–Volume Constraints: Model

In a first approximation, dose constraints are based on the voxel dose limits given in terms of the inequalities

$$D_i^{(s)} \leq R^{(s)} \text{ for OARs and} \tag{2}$$

$$r^{(s)} \leq \ D_i^{(s)} \leq \ R^{(s)} \text{ for targets} \tag{3}$$

STRUCTURE	underdose	lower reference limits	upper reference limits	overdose
	Volume %	*Gy*	*Gy*	Volume %
Brainstem		0.00	54.00	
Spinal Cord		0.00	45.00	
Right Parotid		0.00	30.00	50.00
Mandible		0.00	70.00	
Primary PTV	5.00	66.00	72.60	20.00
(PTV 66)	1.00	61.38		
Subclinical Disease	5.00	60.00	66.00	20.00
(PTV 60)	1.00	55.80		
Bilateral Lymph Nodes	5.00	54.00	59.40	20.00
(PTV 54)	1.00	50.22		

Table 2. The DVCs for oropharyngeal cases as given by the RTOG–H–0022 protocol. The DVC for salivary glands used for computation listed in table is one the three alternate stipulations; the others are: i) mean dose to either parotid below 26 Gy or ii) at least 20 cc of the combined volume below 20 Gy. The cases had two targets at the subclinical disease level.

where $r^{(s)}$ and $R^{(s)}$ define the permissible dose range for a given structure. The lower limit for OARs is not necessary since dose $D_i^{(s)}$ is nonnegative. If these are the only directives for a given structure, we denote them for short as the dose limit constraints (DL). However, the problem may be inconsistent and these limits will not determine the actual delivered dose.

Thus, to augment the model to reflect the dose–volume based prescription we use a new set of contraints. We only give the description of the dose–volume constraint (DVC) of the overdose since the equivalent formulation for the underdose can be easily inferred. The permissible overdose for a given structure s assumes the values given by a set

$$(U^{(s)}, M, F^{(s,1)}, u^{(s,1)}, \ldots, F^{(s,M)}, u^{(s,M)}) \tag{4}$$

where $U^{(s)}$ is a redefined value corresponding to a new limit at the voxel resolution, *cf.* the value $R^{(s)}$ in Equation (2), M is the number of the DVCs, and for $1 \leq l \leq M$, value $F^{(s,l)}$ is the fraction of structure s allowed to obtain the dose between a new reference limit $u^{(s,l)}$ and $U^{(s)}$ ($u^{(s,l)} < U^{(s)}$). If no DVCs are imposed, $U^{(s)}$ corresponds to $R^{(s)}$. For example, the overdose of the right parotid is described by the set $(72.60Gy, 1, 0.50, 30.00Gy)$ (the values are from TABLE (2)).

For every structure s we rewrite constraints as a set of inequalities,

$$g_t(\mathbf{x}) \;\leq\; 0, \quad t \in T^{(s)} \tag{5}$$

where $T^{(s)}$ is a set of indices enumerating the constraints for the given structure. At the voxel resolution we use

$$g_t(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad D_i^{(s)} \leq U^{(s)} \quad (L^{(s)} \leq D_i^{(s)} \leq U^{(s)} \text{ for targets}) \,, \\ L^{(s)} - D_i^{(s)} & \text{if} \quad D_i^{(s)} < L^{(s)} \quad \text{for targets}, \\ D_i^{(s)} - U^{(s)} & \text{if} \quad U^{(s)} < D_i^{(s)} \end{cases} \tag{6}$$

where $L^{(s)}$ is the lower limit equivalent of $U^{(s)}$. These constraints gauge the discrepancy of the voxel dose with the prescription. The cumulative DVCs, *i.e.*, those controlling the compliance with selected points of the required DVH, are

$$g_t(\mathbf{x}) = \sum_{i}^{V^{(s)}} H_i^{(s,l)}(\mathbf{x}) \; - F^{(s,l)} V^{(s)} \; (U^{(s)} - u^{(s,l)}) \tag{7}$$

where $V^{(s)}$ is the volume of the s^{th} structure in terms of the number of voxels, l tags the given overdose DVC and

$$H_i^{(s,l)}(\mathbf{x}) = \begin{cases} 0 & \text{if } \; D_i^{(s)} \; \leq \; u^{(s,l)}, \\ (D_i^{(s)} - u^{(s,l)}) + (U^{(s)} - u^{(s,l)}) & \text{if } \; u^{(s,l)} \; < \; D_i^{(s)} \; \leq \; U^{(s)}, \\ (D_i^{(s)} - u^{(s,l)}) & \text{if } \; U^{(s)} \; < \; D_i^{(s)}. \end{cases} \tag{8}$$

These constraints measure the compliance of the organ's total dose with the prescription. It is instructive to view definitions given by Equation (6) and Equation (7) as functions concerning real and virtual voxels, respectively. If a weight vector complies with all the constraints the plan is *guaranteed* to meet the DVC–based prescription. The proof is in Appendix. Thus this set of DVCs allows us to consider the weight determination as a constraint satisfaction problem.

2.4. The Simultaneous Subgradient Projection Method

Constraint optimization problems are expressed generically as a set of problem functions of the decision parameters. One of these functions is the objective and the rest are the constraints. Thus the vector that satisfies the constraints and optimizes the cost function is a solution of the optimization problem. However any vector satisfying only the constraints is a solution of the *constraint satisfaction problem*. The model of the dose–volume constraints and the algorithm we use here belong to this class. Thus the weight determination assumes the minimalistic approach addressing the dose prescription without any underlying policy. Different solutions can be obtained by changing a starting point or adjustable parameters, or the termination criterion.

We use the simultaneous version of the CSP method (SSP) (Iusem and Moledo, 1986; Censor and Zenios, 1997). The algorithm finds a vector $\mathbf{x}$ that belongs to the set of feasible solutions defined as

$$Q = \{\mathbf{x} \in \mathbb{R}^N | \; \forall_{t \in T} \; g_t(\mathbf{x}) \leq 0 \;\; \wedge \;\; g_t(\mathbf{x}) \text{ is convex}\}, \tag{9}$$

where T is the set containing indices enumerating constraints; in our case $T = \bigcup_s T^{(s)}$.

The SSP method is an iterative algorithm that starts from a weight vector of an arbitrary initial value. The k-th iteration commences with a weight vector, $\mathbf{x}^k$. For every constraint set, t, the following intermediary vector, $\mathbf{X}_t^k$, is computed

$$\mathbf{X}_t^k = \mathbf{x}^k - \; \frac{\max(0, g_t(\mathbf{x}^k))}{\|\partial g_t(\mathbf{x}^k)\|^2} \partial g_t(\mathbf{x}^k), \tag{10}$$

where $\partial g_t(\mathbf{x}^k)$ is a relevant subgradient (or the gradient for the differentiable case) of the function, $g_t(\mathbf{x}^k)$, and $\| \cdot \|$ is the Euclidean norm. The iterative update of the weight vector, $\mathbf{x}^k$, takes the form,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda_k \; (\sum_{t \in T} \omega_t \; \mathbf{X}_t^k - \mathbf{x}^k), \tag{11}$$

where $0 < \lambda_k < 2$ is a relaxation parameter, and $0 < \omega_t$ is the importance of the t-th constraint such that $\sum_{t \in T} \omega_t = 1$. Since the weight vector is nonnegative, the final step is the projection onto the nonnegative orthant, $\mathbb{R}^N_+$.

The computation of the vectors, $\mathbf{X}^k_t$, is the step which determines a parallel character of this algorithm (Andrews, 2000; Censor and Zenios, 1997). For a given iteration all vectors, $\mathbf{X}^k_t$, at the voxel resolution and relevant components in Equation (8) are independent and can be computed simultaneously which we took advantage of in our implementation using a multithreaded Posix technology on a dual Pentium Linux box. In practice this resulted in nearly halving the execution time with comparison to the sequential control flow of the program.

The CSP algorithm is designed to solve the convex feasibility problem. Yet the cumulative DVC functions are quasiconvex. However the use of non–convex constraint functions is not without a precedence for similar algorithms (Fienup, 1982). It seems reasonable to compromise the type of a few constraint functions to maintain the veracity of the model. Moreover all constraint functions are differentiable that allows us to use an analytic form of gradients in the implementation.

2.5. Algorithm Parameters

The importances are user–assigned and systematized in our implementation as follows. First, an importance, ϖ_s, is assigned to each structure, s. Then, the normalized importances are split into a user-assigned fraction for a DVC importance of Equation 7 and the organ's importance of Equation 6. It is worth pointing out that in our implementation the importance of the given constraint can be viewed as an attribute of the given voxel, either real or virtual.

Our previous experiments with the DL-SSP algorithm applied to feasible cases showed that this projection method with such an initial weight vector delivers solutions that approach the vector with smallest norm (Xiao et al., 2003a). This feature is beneficial in that smooth intensity patterns result in reduced total beam–on time. Due to affinity of the DL–SSP and DVC–SSP algorithms, the initial weight vector was set to 0.

The relaxation parameter is also user–assigned, $\lambda^k = \lambda = 1.999$. This decision is dictated by the accelerated performance of the algorithm with the over–relaxation.

Any formal modeling associates the relevant parameters with expressions and operations of mathematical constructs. It is desirable to apply directly the dose–volume contraints used in the physician's prescription as input for the inverse planning. If a computational method resorts to some exogenous structures, parameter values or computational processing it may loose its rigour. However practical situations sometime necessitate these empirical adaptations and modifications to re–gain computational efficacy. They stem from the knowledge of the algorithm's behavior or characteristics of the plan. For example, the loss of robustness of the conjugate gradient method with a quadratic cost function for the beamlet based ITP prompted the authors to differentiate explicitly between computational values and those from the prescription (Spirou and Chui, 1998; Hunt and Burman, 2003). The actual computational parameters can be based on the dosimetrist's experience, *e.g.*, applying stricter interval dose limits for a target or boost levels, which obviously make multiple point DVCs defunct. However the lowering of the upper dose limit for OARs does not have to be viewed as a deviation from the prescription since ideally for any case we would like this dose to be minimal.

For example, in all of the solved cases and for the DL–SSP method we set the spinal cord upper limit at 40.00 Gy. For the DVC-SSP method we imposed a corresponding DVC with the upper limit set at 45.00 Gy. This change appeared to be crucial for obtaining solutions otherwise these methods did not perform robustly enough. The same tactics was applied to the upper limit for the prostate PTV for cases that used constraints given in Table 1B. The DL–SSP used 80.00 Gy as the upper limit in the computation. The DVC–SSP method used interval $(80.00, 83.16]$ Gy as an allowed overdose. The same was for head–and–neck cases with two boost regions. The head–and–neck cases also required more adjustments of the importance parameters. This is our usual practice with complex cases (Bednarz et al., 2000).

In principle, all voxels affect the determination of the weight vector. Yet one must remember that the issue is more subtle since the geometry and topology of the disease site, and the given prescription interplay in a way difficult to envisage. For example, for some cases a significant difference in the volumes of organs might hinder solving plans. This was our rationale for deacreasing by ten–folds the importance of the bladder as compared to other OARs for some of the prostate cases.

2.6. Algorithm Termination

Computations were terminated either when the prescription or an arbitrary number of iterations was reached (in most cases 30000). This was dictated by an emphasis in this report on testing its efficiency in solving a given dose–volume based prescription rather than approaching the prescription or achieving some other particular goals. The same termination condition and values of pertinent adjustable parameters allows for the comparison of these methods.

2.7. Similar Computational Methods

It is instructive to place this particular model and the algorithm in a context of related computational methods (Xing and Chen, 1996; Shepard et al., 2000; Xiao et al., 2000; Starkschall et al., 2001; Xiao et al., 2003a). The SSP algorithm applied to the model without DVC functions, *i.e.*, only with the dose limit constraints (DL-SSP), coincides with the well behaved Cimmino method of the simultaneous projections onto linear constraint sets (Censor and Zenios, 1997; Stark and Yang, 1998). Its salient feature is its inherent convergence which can be viewed in a least-square sense. It is afforded by the existence of the monotonic function decreasing along iterations. This function called proximity is a sum of weighted square distances of the solution from constraints sets (Cimmino, 1938; Censor and Zenios, 1997; Stark and Yang, 1998). The CSP algorithm is also convergent but only for feasible problems of the continuous convex constraint functions and with the uniform boundedness of subgradients (Censor and Zenios, 1997). Thus it is more convenient for us to look at the DVC–SSP from the perspective of the Cimmino method. A sequential version of the latter, known as the Kaczmarz algorithm (Kaczmarz, 1937; Censor and Zenios, 1997), lacks what appears to be a crucial performance advantage of the simultaneous counterpart because in an infeasible case, at some point, iterations restricted exclusively to constraint sets start cycling through values. This is also the reason we used the simultaneous instead of sequential version of the CSP method.

The performance of the DL–SSP method can be accelerated with over–relaxation, *e.g.*, (Höffner et al., 1996). For the relaxation parameter, $\lambda \in (0, 2)$, the iterative

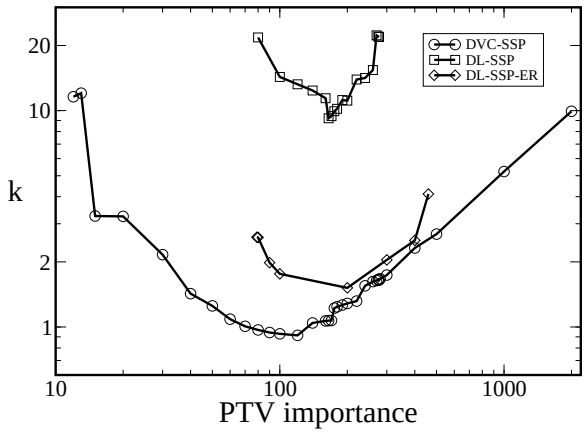

Figure 1. Number of iterations, $k \times 10^3$, as a function of the PTV importance for prostate case P–A–1 solved with the DL–SSP, DL–SSP–ER and DVC–SSP methods. The OARS' importances were set to 1.

update is a contraction mapping that guaranties the algorithm's convergence (Censor and Zenios, 1997; Stark and Yang, 1998; Luenberger, 1997). This is equivalent to preserving a non–increasing proximity function. The values below 1 decelerate the convergence, while parameter values above 1 accelerate it. Thus it is easy to control the over–relaxation along the iterations increasing its value as long as the proximity is non–increasing. If the proximity function increases, the relaxation parameter is set to 1.999. This is our default value for λ in our implementation of the DL–SSP method. It might be even used in a less formal mathematical sense, *i.e.*, with the function that is some other measure of the algorithm's performance. This adaptive adjustment of λ for the DL–SSP method we call elastic relaxation (ER). The value of λ was set to 1.999 for the DVC–SSP method.

3. Results and Discussion

3.1. The Importances and the Convergence Rate

For prostate cases we started experiments with equi–important structures, *e.g.*, $\forall_s \varpi_s = 1$. The DL–SSP and DVC–SSP methods with this initial set–up yielded solutions only for some cases. Next, the OARs' importances were kept fixed while solving with increased values of the PTV importance. The DVC–SSP was able to solve the plan with lower values of the PTV importance than the DL–SSP method . All figures show typical behavior of the algorithms. The comparison of the algorithms for the head–and–neck cases were not be achieved. The DL–SSP algorithm used modified dose limits to get a solution. The following notation is used in the figures. The prefix P or HN are for prostate or head–and–neck cases, respectively. Next, letters A and B are for the prostate cases with constraints given in Table 1A and 1B, respectively. For the head–and–neck cases, A and B denotes cases with four and two boost regions, respectively. The suffix denotes the case number.

Figure 1 demonstrates the number of iterations as a function of the PTV importance for the P–A–1 case. For the DVC–SSP we used the PTV DVCs given in Table 1A(a). The OARs' importances were set to 1. The fractional underdose

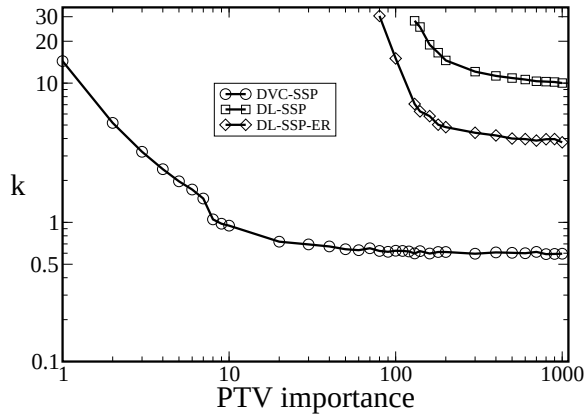

Figure 2. Number of iterations, $k \times 10^3$, as a function of the PTV importance for prostate case P–B–1 solved with the DL–SSP, DL–SSP–ER and DVC–SSP methods. The OARS' importances were set to 1.

(dvc) and overdose (DVC) importance for the PTV were $\varpi_{dvc}^{(PTV)} = \varpi_{DVC}^{(PTV)} = 0.275$. The first solution obtained with the DL–SSP method that met the prescription was for $\varpi_{PTV} = 80$. The DVC-SSP managed to solve the case for $\varpi_{PTV} = 12$. The algorithm performs better while increasing the PTV importance until again the efficiency is slowly lost. Another characteristic behavior is shown in Figure 2 for case P–B–1. The geometry of the anatomical site for this prescription led only to acceleration of the algorithm's convergence while increasing the PTV importance. In this case the DVC–SSP solved the plan with all equi–important structures. The DL–SSP method delivered solution with $\varpi_{PTV} = 130$. These and other figures also include the plots of the DL–SSP method with an adaptively changed relaxation parameter (DL–SSP–ER).

All head–and–neck cases solved with the DVC–SSP exhibited behavior similar to that of the P–A–1 case solved with the DVC–SSP as shown in Figure 1. However all these cases were solved with equi–important structures. Despite more complex anatomy and the dose prescription the lowest number of iterations were comparable to the solutions of the prostate cases.

The DL–SSP method requires more adjustments of the importance parameters, ϖ_s, to proceed efficiently towards a weight vector satisfying the prescription. Our experiments indicate that setting up these parameters is not an intuitive process. It is difficult to establish a template with values for same disease sites. The shape,the size and the relative location of structures affect the computation. Thus providing the algorithm with the additional knowledge of the cumulative dose requirements appears to alleviate this problem.

The DVC–SSP appears to take advantage of the interplay between the numerically well–behaved convergent DL–SSP as applied to constraint sets at the voxel resolution and sets based on cumulative dose–volume conditions. The control over the effects of these parts is *via* the distribution of the importance between constraint sets ascribed to voxels and the DVCs. This behavior for the P–A–1 case is demonstrated in Figure 3. It shows the number of iterations k as a function of the virtual voxels' importance. At first there is a dramatic decrease with best results occurring for the DVC importance about 0.55. In this case further increase of ω_{dvc} slightly reduced the number of iterations to solve the plan.

The use of too small or too large values for the importance of the DVCs may preclude effective convergence to the plan requirements. For example, if the DVC importance is too small, the algorithm neglects the particular DVC and drifts toward behavior that starts to resemble that of the DL–SSP method with a higher number of iterations to meet the plan. The increase of the importance of virtual voxels changes this behavior. The computation acquires the ability to fine–tune the dose to real and virtual voxels. This type of experiment was not carried out for the head–and-neck cases because there were too many combinations of the DVC importances.

3.2. Multipoint DVCs

The multi–point DVCs of the RTOG–H–0022 protocol for the oropharyngeal cancer are not suitable for simple testing. Therefore we used two DVCs restricting underdose of the PTV in prostate cases as marked by (a) and (b) in Table 1A. This type of stipulations can be used to mirror available prescriptions or to design custom made DVHs as shown in Figure 4. This simple experiment with the P–A–1 case shows how the dual dose–volume stipulation can affect the DVH.

For the HN–A–1 case the DVC–SSP algorithm terminated the execution close to 50% of the right parotid volume receiving 30.00 Gy. For other cases the algorithm met the prescription with a better sparing of this organ. Figure 5 demonstrates a simple experiment that improves the sparing of the right parotid for the HN–A–1 case. Instead of using stipulated by the protocol 50% of volume we used 40% of the volume receiving no more than 30.00 Gy. Other parameters remained the same.

Previously, the method of projections onto convex sets was used in a two dimensional feasibility study of the prostate case with the DVCs modeled through an integral dose controlled adaptively in the course of iteration by a limit adjusted proportionally to the violation of the prescription (Lee et al., 1997; Cho et al., 1998). The change of the limit along iterates is to help to guide the computation toward the solution. However such constraints do not mirror a prescription like the DVC–SSP method does.

The use of the CSP method with the cumulative dose–volume histograms as

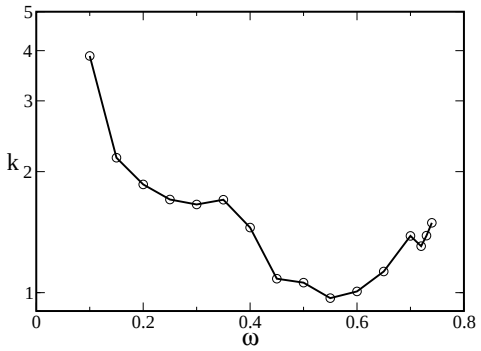

Figure 3. Number of iterations for prostate case P–A–1, $k \times 10^3$, in terms of the DVC importance, ω_{dvc} (fraction of the PTV importance). The number of iteration to reach the prescription with the DL-SSP for the same setup of structural importances was 21781, *cf.* plot DL–SSP in Figure 1. The lowest number of iterations obtained with the DVC-SSP was 969.

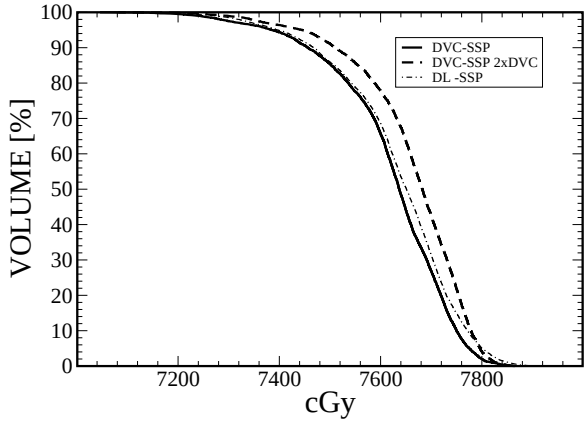

Figure 4. The cumulative DVH for the PTV of the P–A–1 case obtained with different dose(-volume) stipulations. The structural importances were set to match the most efficient performance of the DL–SSP algorithm. We could not achieve the 95% and 90% coverage of the PTV with the DL–SSP method within a reasonable time.

constraint functions has already been reported (Starkschall et al., 2001). The subgradients were approximated with numerical values of derivatives of the DVHs with respect to beam weights. However the discretized nature of the volume of interest may preclude obtaining non–zero gradients. The number of voxels in relevant intervals might remain the same yielding zero–valued gradient despite the change of the dose map. This was our experience with this method applied to prostate cases using 3mm voxel resolution for plan evaluation and starting with a zero weight vector. This separates our implementation from that reported (Starkschall et al., 2001) which used 5mm voxels and 500 randomly spaced points for PTV and 250 points for critical structures. It also used a different initial weight vector.

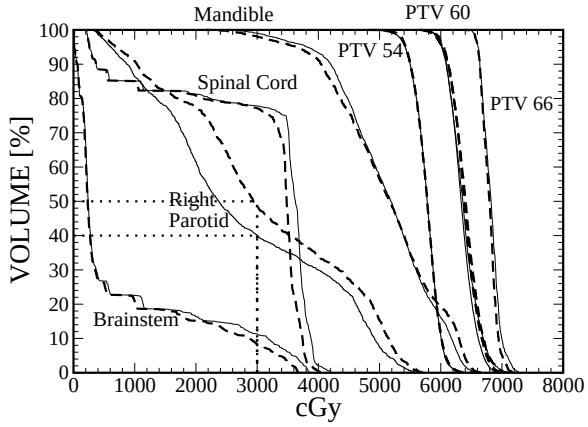

Figure 5. The cumulative DVH for case HN–A–1. The broken lines show the solution obtained for the RTOG–H–0022 protocol. The solid lines show the solution with the DVCs changed from 50% to 40% of the parotid overdose.

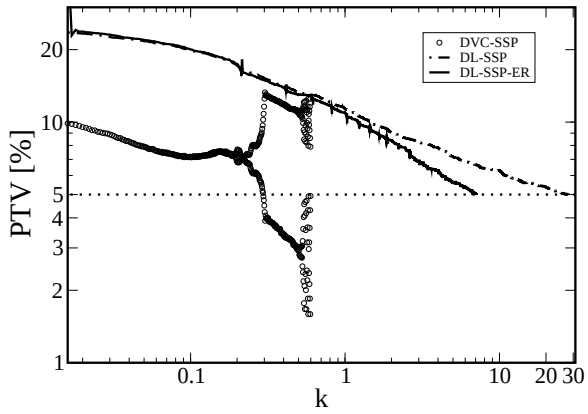

Figure 6. The evolution of the underdose of the PTV for the P–B–1 case along the iterations, $k \times 10^3$, solved with the DL–SSP, DL–SSP–ER and DVC–SSP methods.

3.3. The DVC and the Inconsistency of DL Constraints

Cases which manifest infeasible dose limit constraints can be compared to frustrated systems. Phase I of the simplex method revealed that all investigated for this report cases with the prescription without the DVCs turned out to be inconsistent, *i.e.*, $Q = \emptyset$ as defined in Equation (9). Adding DVCs removes this infeasibility. The DL–SSP method addresses all voxels concurrently. The DVC–SSP method tries to satisfy the voxel doses the same way, however concomitant to that is the effect on the cumulative dose–volume statistics. Figure 6 shows a typical evolution along iterations of an organ DVH. This is the PTV underdose for the P-B-1 case solved with the same structural importances as those that yielded first solution with the DL-SSP method (*cf.* Figure 2). In the case of the DL-SSP method it took a lot of iterations to meet the prescription. When the DVCs is switched on the system reaches compliance with the prescription much faster. The difference is one order of magnitude. Figure 6 is typical for the DVC–SSP method in that it exhibits alternating jumps until the underdose drops below the required value, in this case 5%, concurrently with other organs meeting the prescription. The evolution of the underdose for the DL–SSP method is smooth.

3.4. The DL–SSP Elastic Relaxation and the DVC–SSP

Two experiments were carried out with over–relaxation. First, a fixed value of the over–relaxation parameter above 2 was used to examine the susceptibility of this technique to the loss of the convergence. The DL–SSP method was less forgiving than the DVC–SSP algorithm. The former method was already vitiated with values of λ about 2.3. The behaviour of the DVC–SSP approach was practically not affected by values below 5. Next, the DL–SSP method was modified to adaptively over–relax λ. The increment of λ by a value of $\delta_\lambda = 5.0$ for the non–increasing proximity function was scheduled to occur at arbitrary intervals of 250 iterations. Setting $\delta_\lambda = 5.0$ was partly dictated by the experiments with a fixed over–relaxed λ. As soon as the proximity function deteriorated λ was decreased by δ_λ but not below 1.999. These adaptive

adjustments turned out to be a practical method of accelerating the performance of the DL–SSP algorithm as shown in Figures 2, 3, and 6. However the DVC–SSP method achieved required dose prescriptions in fewer iterations. In addition it also required fewer adjustments of the importances compared to the DL-SSP-ER method.

4. Conclusion

The DVC satisfaction problem was formulated to model prescriptions for ABIP. An iterative algorithm of the simultaneous subgradient projections was used to obtain solutions. The algorithm is easy to implement and has minimal memory requirements. This inherently parallel numerical approach allows for taking advantage of modern computational technologies such as multi-threading on multiple-chip boxes or computer clusters. The model and its solver might be viewed as a specific version of the Cimmino method with non–linear constraints. This algorithmic similarity implies practical salutary numerical features. The diversity of the examined cases provided a representative sample of the algorithm's behavior for two classes of disease sites and prescriptions.

In case of loss of robustness when the DVC–SSP algorithm cannot converge to a prescription compliant weight vector the method is amenable enough to be modified to save the best, in some metrics, weight vector. We want to point out that failure to arrive at the solution does not necessarily equate with inconsistent constraints.

The execution times of a single iteration of these three algorithms are mostly determined by the dose recalculation for a given weight vector. For example, the worst and best times for the DVC–SSP, DC–SSP–ER, and DC–SSP methods for the case in Figure 1 were $(8.55, 109.04\text{s})$, $(38.38, 109.24\text{s})$,$(82.05, 190.05\text{s})$, respectively (sequential implementation on Pentium III 800 MHz Linux box). The time to reach the compliance for the head–and–neck case in Figure 5 was 31.88s (514 iterations) and 324.03s (5370 iterations) for 50% and 40% of the parotid overdose, respectively. The set up of the DVC-SSP method is more intuitive and the performance is more robust than the DL–SSP and DL–SPP–ER method. This improvement is due to the synergy of dose–limit constraints at the voxel level and cumulative DVCs at the organ level. Albeit the elastic relaxation parameter for the DL–SSP method speeds up the execution significantly this approach must still seek proper sets of the importances and use *ad hoc* adjusted dose limits that differ from those given in the prescriptions. The DVC–SSP can also be favorably contrasted with the aforementioned method of the projections onto convex sets with an adaptively adjusted integral dose constraint (Cho et al., 1998) since this approach only partially relies on the prescription.

Acknowledgments

We gratefully acknowledge Computerized Medical Systems, St. Louis, Missouri, USA, that granted us the access to the source code of their FOCUS treatment planning system. We also thank Chris Houser and Nora Walker for their assistance with setting up apertures. The work of Yair Censor was partially supported by the grant 592/00 of the Israel Science Foundation founded by the Israel Academy of Science and Humanities.

Appendix

Let us consider the set of constraint functions as previously defined in § 2.3. The feasible case is equivalent with the certificate of solvability *i.e.* $\exists \mathbf{x}\ \forall_{t \in T^{(s)}}\ \ g_t(\mathbf{x}) \leq 0.$

We show the proof for the overdose since for the underdose is similar. For structure, s, with M overdose DVCs tagged by l, let us split the set of indices, $I^{(s)}$, that enumerate the real voxels into subsets, $C^{(s,l)}$ and $E^{(s,l)}$ defined as

$$C^{(s,l)} = \{t : t \in T^{(s)} \ \wedge D_i^{(s)} - u^{(s,l)} \ \leq \ 0\} \text{ and}$$

$$E^{(s,l)} = \{t : t \in T^{(s)} \ \wedge D_i^{(s)} - u^{(s,l)} \ > \ 0\}.$$

Thus, $E^{(s,l)} \cup C^{(s,l)} = I^{(s)}$, and $E^{(s,l)} \cap C^{(s,l)} = \emptyset$, and $|E^{(s,l)}| + |C^{(s,l)}| = |I^{(s)}| = V^{(s)}$, where symbol, $|\cdot|$, denotes the number elements in a set. Since all function inequalities are satisfied at the voxel resolution it is also true for this very case that

$$E^{(s,l)} = \{t : t \in T^{(s)} \ \wedge \ D_i^{(s)} - u^{(s,l)} \ > \ 0 \ \wedge \ D_i^{(s)} - U^{(s)} \ \leq \ 0\}.$$

The constraint function, $g_t(\mathbf{x})$, for a virtual voxel corresponding to the l^{th} DVC can be rewritten as

$$\begin{aligned} g_t(\mathbf{x}) &= \sum_{i \in E^{(s,l)}} H_i^{(s,l)} + \sum_{i \in C^{(s,l)}} H_i^{(s,l)} - F^{(s,l)} V_s \ (U^{(s)} - u^{(s,l)}) \\ &= \sum_{i \in E^{(s,l)}} \ (\ (D_i^{(s)} - u^{(s,l)}) \ + \ (U^{(s)} - u^{(s,l)})) + \\ &\quad + |C^{(s,l)}| \times 0 \ - F^{(s,l)} V_s \ (U^{(s)} - u^{(s,l)}) \end{aligned}$$

Since, $g_t(\mathbf{x}) \leq 0$, we have

$$\sum_{i \in E^{(s,l)}} \ (\ (D_i^{(s)} - u^{(s,l)}) \ + \ (U^{(s)} - u^{(s,l)})) \leq \ F^{(s,l)} V_s \ (U^{(s)} - u^{(s,l)})$$

For voxel $i \in E^{(s,l)}$ we have $0 \leq \ (D_i^{(s)} - u^{(s,l)})$ and

$$\sum_{i \in E_l^s} (U^{(s)} - u^{(s,l)}) \leq \ F^{(s,l)} V^{(s)} \ (U^{(s)} - u^{(s,l)}).$$

This is equivalent to

$$|E^{(s,l)}| \leq F^{(s,l)} V^{(s)}.$$

Thus for the weight vector, $\mathbf{x}$, complying with all the constraints, the dose map for a given structure complies with the dose–volume based prescription. The number of voxels in set $E^{(s,l)}$ does not exceed the stipulated fraction of the organ volume.

References

Altschuler, M. D. and Censor, Y. (1984). Feasibility solutions in radiation therapy treatment planning. In *Proceedings of the Eighth International Conference on the Use of Computers in Radiation Therapy*, pages 220–224. IEEE Computer Society Press, Silver Spring, Maryland, USA.

Andrews, G. R. (2000). *Foundations of Multithreaded, Parallel, and Distributed Programming*. Addison-Wesley.

Bednarz, G., Michalski, D., Houser, C., Huq, M. S., Xiao, Y., Anne, P. R., and Galvin, J. M. (2000). The use of mixed-integer programming for inverse treatment planning with predefined field segments. *Phys. Med. Biol.*, 47:2235–2245.

Bortfeld, T., Bürkelbach, J., Boesecke, R., and Schlegel, W. (1990). Methods of image reconstruction from projections applied to conformation radiotherapy. *Phys. Med. Biol.*, 35 No 10:1423–1434.

Brahme, A. (2000). Development of Radiation Therapy Optimization. *Acta Oncologica*, 39:579–598.

Censor, Y., Altschuler, M. D., and Powlis, W. D. (1988a). A computational solution of the inverse problem in radiation therapy treatment planning. *Applied Mathematics and Computation*, 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*, 4:607–623.

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

Chen, Y., Michalski, D., Houser, C., and Galvin, J. M. (2002). A deterministic iterataive least-square algorithm for beam weight optimization in conformal therapy. *Phys. Med. Biol.*, 47:1647–1658.

Cho, P., Lee, S., II, R. M., Oh, S., Sutlief, S., and Phillips, M. (1998). Optimization of intensity modulated beams with volume constraints using two methods: Cost function minimization and projection onto convex sets. *Med. Phys.*, 25:435–443.

Cimmino, G. (1938). Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. *La Ricerca Scientifica XVI, Series II, Anno IX,*, 1:326–333.

Computerized Medical Systems, Inc. (2001). Radiation Treatment Planning Software FOCUS. St. Louis, MO, USA.

Deasy, J. O. (1997). Multiple local minima in radiotherapy optimization problems with dose–volume constraints. *Medical Physics*, 24:1157.

Deasy, J. O., Niemierko, A., Herbert, D., Yan, D., Jackson, A., Haken, R. K. T., Langer, M., and Sapareto, S. (2002). Methodological issues in radiation dose–volume outcome analysis: Summary of a joint AAPM/NIH workshop. *Med. Phys.*, 29 (9):2109–2127.

Eisbruch, A., Chao, K. C., and Garden, A. (2001). Phase I/II study of conformal and intensity modulated irradiation for oropharyngeal cancer. www.rtog.com/members/protocols/h0022/index.html. Radiation Therapy Oncology Group protocol H–0022. Current Edition: January 15, 2002.

Fienup, J. R. (1982). Phase retrieval algorithms: a comparison. *Applied Optics*, 21:2758–2769.

Galvin, J. M., Bednarz, G., and Croce, R. (2000). Advanced forward planning techniques or forward planning is alive and well in IMRT world. In *General Practice of Radiation Oncology Physics in the 21st Century*, volume Medical Physcis Monograph No. 26, pages 73–100. Medical Physics Publishing.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of Np-Completeness*. W H Freeman & Co.

Glatstein, E. (2001). Personal Thoughts on Normal Tissue Tolerance, or, What the Textbooks don't Tell You. *Int. J. Radiat. Oncol. Biol. Phys.*, 51:1185–1189.

Höffner, J., Decker, P., Schmidt, E., Herbig, W., Rittler, J., and Wiss, P. (1996). Development of a Fast Optimization Preview in Radiation Treatment Planning. *Strahlentherapie und Onkologie*, 172:384–394.

Hunt, M. A. and Burman, C. M. (2003). Treatment planning considerations usingIMRT. In *A Practical Guide to Intensity-Modulated Radiation Therapy*, pages 103–121. Medical Physics Publishing.

IMRTCWG (2001). Intensity–Modulated Radiotherapy: Current Status and Issues of Interest. *Int. J. Radiat. Oncol. Biol. Phys.*, 51:880–914. (Intensity Modulated Radiation Therapy Collaborative Working Group).

Iusem, A. N. and Moledo, L. (1986). A finitely convergent method of simultaneously subgradient projections for the convex feasibility problem. *Mat. Aplic. Comp.*, 5:169–184.

Kaczmarz, S. (1937). Angenaherte auflosung von systemen linearer gleichungen. *Bulletin de l'Academie Polonaise des Sciences et Lettres*, A35:355–357.

Kak, A. C. and Slaney, M. (2001). *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, USA. Available also from the author webpage http://www.slaney.org/pct/.

Lee, S., Cho, P., II, R. M., and Oh, S. (1997). Conformal radiotherapy computation by the method of alternating projections onto convex sets. *Phys. Med. Biol.*, 42:1065–1086.

Luenberger, D. G. (1997). *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, NY, USA.

Pugachev, A., Li, J. G., Boyer, A. L., Hancock, S. L., Le, Q.-T., Donaldson, S. S., and Xing, L. (2001). Role of beam orientation optimization in intensity-modulated radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 50:551–560.

Shepard, D., Ferris, M., Olivera, G., and Mackie, T. (1999). Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, 41:721–744.

Shepard, D. M., Olivera, G., Reckwerdt, P. J., and Mackie, T. R. (2000). Iterative approaches to dose optimisation. *Phys. Med. Biol.*, 45:69–90.

Spirou, S. V. and Chui, C.-S. (1998). A gradient inverse planning algorith with dose–volume constraints. *Med. Phys.*, 25:321–333.

Spirou, S. V., Fournier-Bidoz, N., Jang, J., and Chui, C.-S. (2001). Smoothing intensity–modulated beam profiles to improve the efficiency of delivery. *Med. Phys.*, 25:321–333.

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

Starkschall, G., Pollack, A., and Stevens, C. W. (2001). Treatment Planning Using a Dose-Volume Feasibility Search Algorithm. *Int. J. Radiat. Oncol. Biol. Phys.*, 49:1419–1427.

Xiao, Y., Censor, Y., Michalski, D., and Galvin, J. M. (2003a). The least-intensity feasible solution for aperture-based inverse planning in radiation therapy. *Annals of Operations Research*, 119:183–203.

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

Xiao, Y., Michalski, D., Censor, Y., and Galvin, J. M. (2003b). Smoothness evaluation of intensity patterns resulting from simultaneous subgradient projection method for dose–volume objective satisfaction. Poster. Annual Meeting of the American Association of Physicists in Medicine, San Diego, CA. PO–T–125, Med. Phys., 30(6), 1324.

Xing, L. and Chen, T. (1996). Iterative methods for inverse treatment planning. *Phys. Med. Biol.*, 41:2107–2123.