

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.Sciencedirect.com)

## Computers &amp; Operations Research

journal homepage: [www.elsevier.com/locate/caor](http://www.elsevier.com/locate/caor)

# A hybrid approach to beam angle optimization in intensity-modulated radiation therapy

D. Bertsimas<sup>a</sup>, V. Cacchiani<sup>b,\*</sup>, D. Craft<sup>c</sup>, O. Nohadani<sup>d</sup>

<sup>a</sup> Operations Research Center, Massachusetts Institute of Technology, Cambridge, USA

<sup>b</sup> DEIS, University of Bologna, Bologna, Italy

<sup>c</sup> Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, USA

<sup>d</sup> School of Industrial Engineering, Purdue University, Grissom Hall, West Lafayette, USA

## ARTICLE INFO

### Keywords:

Beam angle optimization  
Intensity-modulated radiation therapy  
Heuristic algorithm  
Simulated annealing  
Gradient descent  
Linear programming  
Computational experiments

## ABSTRACT

Intensity-Modulated Radiation Therapy is the technique of delivering radiation to cancer patients by using non-uniform radiation fields from selected angles, with the aim of reducing the intensity of the beams that go through critical structures while reaching the dose prescription in the target volume. Two decisions are of fundamental importance: to select the beam angles and to compute the intensity of the beams used to deliver the radiation to the patient. Often, these two decisions are made separately: first, the treatment planners, on the basis of experience and intuition, decide the orientation of the beams and then the intensities of the beams are optimized by using an automated software tool. Automatic beam angle selection (also known as Beam Angle Optimization) is an important problem and is today often based on human experience. In this context, we face the problem of optimizing both the decisions, developing an algorithm which automatically selects the beam angles and computes the beam intensities. We propose a hybrid heuristic method, which combines a simulated annealing procedure with the knowledge of the gradient. Gradient information is used to quickly find a local minimum, while simulated annealing allows to search for global minima. As an integral part of this procedure, the beam intensities are optimized by solving a Linear Programming model. The proposed method presents a main difference from previous works: it does not require to have on input a set of candidate beam angles. Indeed, it dynamically explores angles and the only discretization that is necessary is due to the maximum accuracy that can be achieved by the linear accelerator machine. Experimental results are performed on phantom and real-life case studies, showing the advantages that come from our approach.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

Intensity-Modulated Radiation Therapy (IMRT) consists of delivering radiation to cancer patients by modulating the intensities of the rays (*beams*), which are typically delivered from five to seven different directions (*angles*). The tumor shape is analyzed by the doctor, who outlines the so-called *target volume* and decides the *prescribed dose* that must be delivered to the tumor cells. Each beam is divided into *beamlets*, all having the same direction but which can be assigned different intensities, achieved by sliding the leaves of a multi-leaf collimator in the beam path while the beam is on or by using the step and shoot approach (in which the radiation is off whenever the leaves move). The

intensities of the beamlets are optimized with the aim of achieving the prescribed dose requested by the doctor for the target volume while sparing the *organs at risk* (OARs).

This technique is increasingly becoming common in the hospitals and it requires an automated tool which captures many different features, in order to produce good treatment plans. Three can be considered as the main phases for building a planning process (see [11] for a survey on this topic):

- the selection of the number of beams and the directions from which to deliver the radiation
- the selection of the intensities for the beamlets
- the selection of a delivery sequence

The aim of the first phase is to find the best selection of radiation angles. Once the directions have been obtained, the intensities are determined. While the process of optimizing the intensities is generally automated, the selection of the beam angles is often

\* Corresponding author.

E-mail addresses: [dbertsim@MIT.EDU](mailto:dbertsim@MIT.EDU) (D. Bertsimas), [valentina.cacchiani@unibo.it](mailto:valentina.cacchiani@unibo.it) (V. Cacchiani), [dcraft@partners.org](mailto:dcraft@partners.org) (D. Craft), [nohadani@purdue.edu](mailto:nohadani@purdue.edu) (O. Nohadani).

based on the planners' experience and on trial-and-error procedures. The last phase focuses on the problem of realizing the intensity by using a multi-leaf collimator.

In this paper, we address the problem of optimizing the intensities of the beamlets together with the choice of the angles from which the beams are delivered. This problem is called Beam Angle Optimization (BAO) and is a non-convex problem, with many local minima. Several approaches have been developed for dealing with the BAO problem, many of which are based on heuristic methods, due to the difficulty of the problem.

In particular, simulated annealing is often applied to tackle the problem: see e.g. [2,5,10,25,28]. In these works, sets of directions are considered: they can be predefined on input or determined in a first phase of the algorithm that takes into account the quality of each beam orientation. A set of beam angles are then selected by using simulated annealing or by combining simulated annealing and local neighborhood search as in [2].

Genetic algorithms are also used (see e.g. [13,15]), as well as gradient search (see e.g. [6,22]) and sometimes the two techniques are combined together (see e.g. [19,26]): the genetic algorithm is used to select suitable beam angles and then gradient search is used to determine the intensity profiles of the beams. A metaheuristic approach is developed by Zhang et al. [32], in which computational efficiency is achieved by utilizing high-throughput computing.

Mixed Integer Programming approaches have also been developed (see e.g. [7,18,21,23,30,31]): in this case, usually, a set of candidate beam angles is given on input (or determined in a first phase of the algorithm), among which the best angles can be chosen. Branch and cut or branch and bound algorithms are proposed to solve the problem. Beam angle elimination is also applied (see e.g. [14,20]). Ehrgott et al. [12] present a mathematical development that provides a unified framework for the problem and several techniques are compared to demonstrate how they behave.

In other works, a score function, introduced to measure the "goodness" of each beamlet at a given angle, is used to select some angles among a set of candidate beam angles (see e.g. [8,24,27,29]). The score function can give preference to beamlets that can deliver a higher dose to the target without exceeding the tolerance of the sensitive structures. The overall score of a beam is calculated as the sum of the scores of all the beamlets belonging to it. The beam orientations with the highest value are then selected.

The method that we develop to tackle the BAO problem differs from previous works for the following reasons. First of all, it does not require to have on input a set of candidate beam angles: indeed it dynamically explores angles and the only discretization that is necessary is due to the maximum accuracy that can be achieved by the linear accelerator machine. In other words, we allow to select any angle that can be obtained by the machine. This is why we present the model with continuous variables corresponding to the angle selection (the proposed method will take into account the maximum accuracy of the machine, as explained in Section 3). In order to model continuous beam angle space, we perform dose computations at a fine angular spacing (e.g.  $\Delta = 2^\circ$ ): this means that the dose matrix is computed at  $360/\Delta$  angles. The beamlet dose computations are done using CERR (Computational Environment for Radiotherapy Research, St. Louis, MO) [9] with QIB (quadratic infinite beam) method, which uses an algorithm based on [1]. Then we use linear interpolation for obtaining the beamlet dose-influence values at any continuous angle.

Since the problem appears to be highly non-convex and with many local minima (as shown e.g. by Bortfeld and Schlegel [5]), we propose a heuristic hybrid method (HM), which combines a

Simulated Annealing procedure (SA) with a Gradient Descent method (GD). HM consists of an iterative approach: it alternates few steps of SA for quickly finding a local minimum, with few steps of GD for jumping out of the local minima and starting to search in a different part of the solution space. In this way, we overcome the difficulty of searching over a given power set of angles by using calculus, i.e. gradient computation (as in [6]) and by using simulated annealing to avoid being trapped in a local optimum. We wish to mention that the proposed method does not address beamlet implementation issues, i.e. how the obtained solution could be implemented by means of the multi-leaf collimator. However, note that there is nothing fundamentally incorrect about using a beamlet based approach, see for example the work by Jelen et al. [17].

The problem is initially tested for a simple phantom case. Then, we study a real-life case. In both cases, we compare the solutions found by HM with the solutions obtained with equispaced angles and with the solutions obtained by applying a pure simulated annealing approach, which is often used for BAO.

The paper is organized as follows. In Section 2 we formally define the problem, and present a Non-Linear Programming formulation which takes into account clinical requirements (see Hong et al. [16]). In Section 3 we describe the heuristic hybrid approach and in Section 4 we present computational results on a phantom case and on a real-life case study. Finally, we draw conclusions in Section 5.

## 2. Problem description and model formulation

The first step for building a treatment plan consists of outlining the shape of the tumor (Clinical Target Volume) and of the organs at risk. This is done by the doctor, who also decides the prescribed dose of radiation which is needed in order to kill the tumor cells. The image of the body of the patient is then discretized, by building a grid where each point is called *voxel*.

Usually, the treatment planners choose the directions from which the beams are delivered. This choice is done manually, based on previous experience and intuition. Often, the beam directions are chosen such that they go through the isocentre of the tumor and are almost equispaced. However, investigating the simultaneous choice of beam directions and intensities is very important from a practical point of view, as one might find solutions which hardly could be found manually. In addition, an automated decision can reduce the trial-and-error process, and also find solutions of comparable quality with a smaller number of used directions, thus reducing the overall treatment time. This is very important since a long treatment can more likely lead to inaccurate positions of the patient and consequently to a faulty delivery of radiation.

Each beam is divided into beamlets, having the same direction but which can be assigned different intensities. BAO aims at determining a set of beam angles and the corresponding beamlet intensities, so that the prescribed dose to the tumor is reached, while the organs at risk are spared. We focus on the co-planar treatment, i.e. beam angles are chosen on a circle around a slice of the body (containing the center of the tumor) of the patient.

### 2.1. Clinical model

In this section, we present a Non-Linear Programming formulation of the BAO Problem, which can well capture real-life requirements, according to what is presented in [16]. In their work, Hong et al. study a multicriteria optimization approach to deal with different planning goals. They focus on the study of pancreatic patients and produce a database of treatment plans

(Pareto surface); the physician can choose among those plans, finding a trade-off between different objectives (i.e., deciding which OAR should receive the lowest dose). In the following we introduce some notation and describe the clinical model. We refer the readers to [16] for a more detailed description.

Let  $S$  be the number of Volumes of Interest (VOIs), i.e. OARs, normal tissue (i.e., all the body voxels which do not belong to any particular structure) and target volumes (i.e. tumor), that need to be considered in the construction of a treatment plan. Let  $V_h$  be the set of all voxels (considered in the discretized grid) associated with  $VOI_h$  ( $h=1, \dots, S$ ), and  $N_h$  be the number of voxels in  $VOI_h$  ( $h=1, \dots, S$ ). Let  $LB_i$  be the prescribed dose for each voxel  $i \in V_h$  and  $UB_i$  be the maximum dose that can be delivered to any voxel  $i \in V_h$  ( $h=1, \dots, S$ ). Let  $B$  be the set of beamlets, partitioned into subsets  $B_1 \cup \dots \cup B_n$ , where  $n$  represents the number of beams, or equivalently the number of angles from which the radiation will be delivered. Finally, let  $D$  be the matrix describing the dose influence, where each element  $D_{ij}(\theta_k)$ ,  $i \in V_h$ ,  $j \in B_k$ ,  $k=1, \dots, n$ ,  $h=1, \dots, S$ , represents the dose delivered to voxel  $i$  by a unit intensity of beamlet  $j$  from direction  $\theta_k$ . The matrix  $D$  is obtained by running an IMRT beamlet calculation every  $\Delta$  degrees, i.e. we compute the dose matrix at  $360/\Delta$  angles. In order to obtain the values of the doses for a generic angle, we use linear interpolation.

We introduce continuous variables  $x_j$  ( $j \in B$ ), representing the beamlets intensities, and continuous variables  $\theta_k$  ( $k=1, \dots, n$ ), representing the angles. We introduce for sake of clarity variables  $d_i$  (linked with equality constraints to variables  $x$ ), representing the dose delivered to the voxel  $i$  ( $i \in V_h$ ,  $h=1, \dots, S$ ).

Usually, a weight is assigned to each voxel in the objective function and typically the same weight is given to voxels in the same structure. A high weight is usually given to selected OARs that must be spared by radiations. However, often the doctor does not provide a specific weight for each voxel, but rather he describes how to limit underdosing to the tumor voxels and overdosing to the OARs. This can be expressed by means of max/mean functions, min/mean functions and ramp functions that specify how to penalize the underdosing or overdosing. To this aim, the VOIs are grouped in different sets, according to the different objectives and constraints in which they are involved. Auxiliary continuous variables are introduced for expressing these objectives and constraints. The objectives and the constraints are of three different types ( $O_p$  and  $C_p$  are sets indexing respectively objectives and constraints of type  $p = \{1, 2, 3\}$ ):

1. *Type 1 ( $O_1$  and  $C_1$ )*: this type is used for OARs. In objectives  $O_1$ , a weighted convex combination (given by parameter  $0 \leq \alpha_h \leq 1$ ) of the maximum dose  $y_h$  to OAR  $h \in O_1$  and of the mean dose  $1/N_h(\sum_{i \in V_h} d_i)$  to the same OAR is penalized. Different non negative weights  $w_h$  can be chosen for the OARs considered in this type of objective, in order to give them more or less importance in the optimization. Based on the value of  $\alpha_h$ , we penalize, in the objective function, either the maximum dose or the mean dose or a convex combination of them. In constraints  $C_1$ , the same convex combination of the maximum dose  $y_h$  to the OAR and the mean dose  $1/N_h(\sum_{i \in V_h} d_i)$  to the OAR is bounded by parameter  $g_h$ .
2. *Type 2 ( $O_2$  and  $C_2$ )*: this type is used for target volumes (i.e. tumor). In this case, the minimum dose  $y_h$  delivered to target  $h$  is taken into account. In objectives  $O_2$ , a weighted convex combination (given by parameter  $0 \leq \alpha_h \leq 1$ ) of the minimum dose  $y_h$  to the target and the mean dose  $1/N_h(\sum_{i \in V_h} d_i)$  to the target is considered. Note that we will use negative weights  $w_h$  for the targets, since we maximize such objectives. Also in this case, different weights can be assigned to different targets, according to their relevance. Constraints  $C_2$  are used to impose

a lower bound  $g_h$  on the convex combination of the minimum dose and mean dose to be delivered to target  $h$ .

3. *Type 3 ( $O_3$  and  $C_3$ )*: this type is used both for targets and OARs. A target dose  $t_h$  is considered for VOI  $h$ . Then, a “V” function or ramp function is defined, where each side of the “V” can have its own slope (i.e. a generalization of an absolute deviation from a target dose function). The lower (i.e. left) slope is given by the parameter  $s_h^l$ , and the upper (i.e. right) slope by  $s_h^u$ . We use this function to limit underdosing to a target and overdosing to an OAR. The auxiliary variables  $z_i^h$  are used to track the upper or lower dose penalties. In objectives  $O_3$ , we penalize the weighted mean  $1/N_h \sum_{i \in V_h} z_i^h$  (over the number of voxels) underdosing to the targets or overdosing to the OARs. Again, weights  $w_h$  depend on the importance that we give to the VOIs in the corresponding objective. In constraints  $C_3$ , we bound the same mean by  $g_h$ .

The problem formulation reads as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{h \in O_1 \cup O_2} w_h \left( \alpha_h y_h + (1 - \alpha_h) 1/N_h \left( \sum_{i \in V_h} d_i \right) \right) \\ & + \sum_{h \in O_3} w_h \left( 1/N_h \sum_{i \in V_h} z_i^h \right) \end{aligned} \tag{1}$$

$$\sum_{k=1}^n \sum_{j \in B_k} D_{ij}(\theta_k) x_j = d_i, \quad i \in V_h, \quad h = 1, \dots, S \tag{2}$$

$$d_i \geq LB_i, \quad i \in V_h, \quad h = 1, \dots, S \tag{3}$$

$$d_i \leq UB_i, \quad i \in V_h, \quad h = 1, \dots, S \tag{4}$$

$$x_j \geq 0, \quad j \in B_k, \quad k = 1, \dots, n \tag{5}$$

$$0 \leq \theta_k \leq 360, \quad k = 1, \dots, n \tag{6}$$

$$y_h \geq d_i, \quad i \in V_h, \quad h \in O_1 \cup C_1 \tag{7}$$

$$y_h \leq d_i, \quad i \in V_h, \quad h \in O_2 \cup C_2 \tag{8}$$

$$\alpha_h y_h + (1 - \alpha_h) 1/N_h \sum_{i \in V_h} d_i \leq g_h, \quad h \in C_1 \tag{9}$$

$$\alpha_h y_h + (1 - \alpha_h) 1/N_h \sum_{i \in V_h} d_i \geq g_h, \quad h \in C_2 \tag{10}$$

$$z_i^h \geq s_h^u (d_i - t_h), \quad h \in O_3 \cup C_3, \quad i \in V_h \tag{11}$$

$$z_i^h \geq s_h^l (t_h - d_i), \quad h \in O_3 \cup C_3, \quad i \in V_h \tag{12}$$

$$1/N_h \sum_{i \in V_h} z_i^h \leq g_h, \quad h \in C_3 \tag{13}$$

The model presents an objective function that helps to express the requirements of the doctors, as described above. More precisely, the objective is to minimize the weighted convex combination of maximum dose and mean dose delivered to the OARS ( $O_1$ ), while maximizing the weighted convex combination of the minimum dose and the mean dose delivered to the targets ( $O_2$ ) and while penalizing overdosing to the OARs and underdosing to the target ( $O_3$ ). Thus, depending on the VOIs considered in the different objectives  $O_1$ ,  $O_2$  and  $O_3$ , the goal can also be to maximize the dose to the tumor (weights  $w_h$  are negative in  $O_2$ ). Constraints (2) express the dose delivered to each voxel: these constraints are used to derive knowledge on the gradient in the HM approach (see [6]). Constraints (3) impose to achieve the prescribed dose for the voxels: in particular, lower bound  $LB_i$  corresponds to the prescribed dose for tumor voxel  $i$ , and is often

set to zero for voxels belonging to the other VOIs. Constraints (4) impose a maximum amount of dose for each voxel. Constraints (5) express the non-negativity of the beamlet intensity variables. Constraints (6) impose the bounds on the angles. Finally, constraints (7)–(13) are used to express the objective function. In particular, constraints (7) are used to define the maximum dose  $y_h$  delivered to VOI (OAR)  $h$  belonging objective  $O_1$  and/or constraint  $C_1$ . Constraints (8) are used to define the minimum dose  $y_h$  delivered to VOI (target)  $h$  belonging objective  $O_2$  and/or constraint  $C_2$ . Constraints (9) impose an upper bound  $g_h$  on the convex combination of maximum and mean dose delivered to VOI (OAR)  $h \in C_1$ . Constraints (10) impose a lower bound  $g_h$  on the convex combination of minimum and mean dose delivered to VOI (target)  $h \in C_2$ . Constraints (11) and (12) are used to define the  $z_i^h$  variables that track the overdosing to the OARs and underdosing to targets, respectively. Finally, constraints (13) impose a bound on the overdosing to the OARs and underdosing to targets belonging to constraints  $C_3$ . Note that the dependence of the doses  $D_{ij}$  on the angles makes the problem non-linear.

The selection of the parameters used in the model is derived by the interaction with doctors. They provide the lower bound for the target dose and the upper bounds for the doses to the OARs. In addition, since they want to have an insight on what happens when sparing one OAR with respect to the others, usually a large weight  $w_h$  is given in the objective function to an OAR belonging to  $O_1$  or  $O_3$  and smaller weights to the other ones. For OARs and for the target in  $O_3$  and  $C_3$  the values  $t_h$  for limiting overdosing and underdosing respectively are determined by the doctors, who also give an insight on how the shape of ramp function should look like. Note that doctors provide these values based on experience.

### 3. Hybrid method

The hybrid method that we have developed is based on the combination of a Simulated Annealing procedure with a Gradient Descent method. The GD allows to quickly find a local minimum. On the other hand, since the objective function presents many local minima, SA is used to avoid being trapped in a local minimum and explore other parts of the solution space. HM consists of an iterative procedure, which alternates some iterations of GD and some iterations of SA. The gradient information is obtained as in [6]. In [6], Craft studies local beam angle optimization, i.e. how a beam angle set can be refined by using gradient information. The gradient is derived by using linear programming duality theory to get the change in the objective function when matrix  $D$  is perturbed, and then by using the known  $D(\theta)$  and the slope  $dD(\theta)/d\theta$  of  $D(\theta)$ , with the chain rule to get how  $f$  changes with a change of  $\theta$  (we refer the readers to [6] for further details).

HM starts from a set of  $n$  (e.g. 5) equispaced angles  $(\theta_1, \dots, \theta_n)$ . This choice is because it is reasonable to start with angles that are not too close to each other. In addition, it is a common approach to choose angles that are almost equispaced. Thus we use the starting solution for the comparison of HM. As already mentioned, the dose influence matrix  $D$  is obtained by running an IMRT beamlet calculation every  $\Delta$  degrees i.e. we compute the dose matrix at  $360/\Delta$  angles. Given the angle choice, the values of the corresponding entries in the  $D$  matrix are computed by linear interpolation, and the obtained LP-problem is solved to optimality by means of CPLEX solver. More precisely, in model (1)–(13), Eq. (2) are replaced by the following equations:

$$\sum_{k=1}^n \sum_{j \in B_k} D_{ij}(\bar{\theta}_k) x_j = d_i, \quad i \in V_h, \quad h = 1, \dots, S \quad (14)$$

and constraints (6) are removed.

Note that  $D_{ij}(\bar{\theta}_k)$  ( $i \in V, j \in B_k, k = 1, \dots, n$ ) are known values. We solve the LP-model (1), (3)–(5), (7)–(13), (14) to optimality, once that the angles are fixed ( $\theta = \bar{\theta}$ ). Let  $x^*$  be the primal optimal solution of the LP-model and  $z^*$  be the corresponding optimal objective value. The algorithm performs  $K_{GD}$  iterations of GD. At each iteration, the LP-problem corresponding to the current set of angles is solved and the dual solution is used to obtain an approximation of the gradient  $\nabla f(\theta) = (\partial f / \partial \theta_1, \dots, \partial f / \partial \theta_n)$ . As mentioned, the approximation is computed by means of sensitivity analysis (see [6]). A better approximation for the gradient can be obtained by applying a finite-difference method as in (Bertsimas et al. [4]). We use the method presented in [6] due to its simplicity.

A step  $\bar{\gamma}$  is taken from the current set of angles  $\bar{\theta}$  along the direction opposite to the gradient, obtaining the new set of angles, ( $\bar{\gamma}$  indicates the length of the step). Sometimes this step happens to be “too big”, i.e. the objective function value does not decrease. This is because of the approximation that we introduce with linear interpolation. In this case a smaller step (e.g.  $\bar{\gamma}/10$ ) is considered, until the new objective function value decreases or the step becomes smaller than a threshold  $\bar{\gamma}_0$  (i.e., we have reached a local minimum). Note that threshold  $\bar{\gamma}_0$  is taken based on the maximum accuracy provided by the linear accelerator machine. In case the change of the angles cannot be implemented by the machine, HM behaves as if it has reached a local minimum. This process turns out to be more time consuming if  $\Delta$  is larger: indeed, when  $\Delta$  is larger, the gradient information is less trustable (we have the correct information on the dose influence matrix, computed with IMRT beamlet calculation, at fewer angles) and  $\bar{\gamma}$  needs to be reduced many times. Thus, if  $\Delta$  is strictly larger than  $2^\circ$ , we apply the following alternative method.

If the 2-norm of the gradient is less than a given threshold (e.g. 1.25 in our tests), then the gradient information is neglected (the set of angles does not change). Otherwise, a change of  $\delta$  degrees is applied from the current set of angles  $\bar{\theta}$  along the direction opposite to the gradient. It may happen that the absolute value of some components of the gradient is very big and leads to a very big change in the angles (this can happen if  $\Delta$  is large and the gradient information is not very precise). Since we want to perform a local search using the gradient information, we decide the following approach:

- if the absolute value of a component of the gradient is at least 10, then we change the corresponding angle by  $\delta = 2^\circ$ ,
- otherwise, if the absolute value of a component of the gradient is at least 1, then we change the corresponding angle by  $\delta = 1^\circ$ ,
- otherwise if the absolute value of a component of the gradient is at least 0.1, then we change the corresponding angle by  $\delta = \delta_{min}$  degrees (where  $\delta_{min}$  corresponds to the smallest change in degrees that can be achieved by the linear accelerator machine).

This means that we consider each component of the gradient and update the corresponding angle, according to the absolute value of this component. The best choice for these threshold values is likely to be case specific, but it seems reasonable to suspect that a set of values that works well for one patient of a certain disease type will work for other patients of the same type. The fact that the same set works well for vastly different objective function weightings, see Fig. 10, is encouraging.

We change the angles only if the 2-norm is larger than the given threshold: indeed, if the 2-norm is small, it is likely that the gradient step does not lead to a relevant improvement of the solution. In addition, the computing time for the LP solution is dominant in HM, with respect to the other steps. Thus, we avoid spending time for computing the LP solution if the expected

improvement might not be significant. For the same reason, we follow the described rule for updating the angles, according to the absolute values of the components of the gradient. This helps to perform a local search around the solution, without moving from each current angle to a new angle that differs more than  $2^\circ$  from it. Note that both approaches avoid changes of the angles if they are not of practical relevance or achievability, due to the accuracy of the linear accelerator machine.

After  $K_{GD}$  iterations of GD, the algorithm performs  $K_{SA}$  iterations of SA, in order to escape from the local minimum. At each iteration  $l$ , ( $l = 1, \dots, K_{SA}$ ), a new set of angles is generated starting from the previous one  $\theta^{new} = \theta^{old} + r * \alpha$ , where:

- $\theta^{new} = (\theta_1^{new}, \dots, \theta_n^{new})$  is the new set of angles,
- $\theta^{old} = (\theta_1^{old}, \dots, \theta_n^{old})$  is the previous set of angles (i.e., when  $l = 1$ , it is the set of angles obtained in the last gradient descent iteration),
- $r = (r_1, \dots, r_n)$  is a set of random numbers with Gaussian distribution (with mean 0 and standard deviation 1),
- $\alpha$  is a parameter for the step size.

```

Hybrid Method
begin
  initialize the set of angles with equispaced angles:  $\bar{\theta} = (\bar{\theta}_1, \dots, \bar{\theta}_n)$ ;
  create the LP-problem corresponding to the set of angles  $\bar{\theta}$ ;
  solve it and obtain  $x^*$  and  $z^*$ ;
  repeat
    repeat
      gradient descent
    until  $K_{GD}$  iterations have been executed;
    repeat
      simulated annealing
    until  $K_{SA}$  iterations have been executed;
  until time limit is reached;
end.

```

Fig. 1. General structure of the hybrid method.

The random numbers are chosen to be Gaussian distributed, because this allows to locally explore the neighborhood of  $\theta^{old}$ , i.e. we focus “nearby” the old set of angles but occasionally allow to jump further. The step size  $\alpha$  is chosen such that the new set of angles can be obtained by the linear accelerator machine. The LP model is solved for  $\theta^{old}$  and  $\theta^{new}$  giving optimal objective function values  $z^{old}$  and  $z^{new}$ , respectively. According to standard SA, the move to the new set of angles  $\theta^{new}$  is accepted whenever the change improves the objective value or according to the probability  $\exp(-(z^{new} - z^{old})/t_l)$ , where  $t_l$  is the temperature of the system at the iteration  $l$ . The adopted cooling scheme is the *Adaptive Simulated Annealing*:  $T_0 * \exp(-c * l^{1/d})$ , where  $T_0$  is the initial temperature of the system,  $c$  is a parameter, and  $d$  is the dimension of the problem, as introduced in [3]. The temperature is updated every  $K_T$  iterations, according to this cooling scheme.

The general structure of the method is outlined in Fig. 1 and the first approach for GD and SA are described in Figs. 2 and 3, respectively.

#### 4. Computational results

We illustrate the experimental results obtained by the proposed method HM discussed in Section 3 on a phantom case and on a 3D real-life pancreatic case study provided by the Massachusetts General Hospital (MGH) of Boston, MA. HM was implemented in C and CPLEX was used as a general purpose solver for solving the LP-model (1), (3)–(5), (7)–(13), (14), when the set of angles is fixed. For the phantom case we performed a comparison of the solutions obtained by HM with a pure gradient descent, a pure simulated annealing and with the solution obtained with equispaced angles. Since the gradient descent turns out to be the worst method (as it will be shown in the next section), for the real-life case study the comparison is performed with a pure simulated annealing and with the solution obtained with equispaced angles. We wish to mention that we tried several different parameter settings, and we describe in the next section the choice that experimentally turns out to be the best one.

```

Gradient Descent Method
begin
   $k := 1$ ;  $\theta^k := \bar{\theta}$ ;  $\gamma := \bar{\gamma}$ ;  $z^{*0} = \infty$ ;  $\gamma_0 := \bar{\gamma}_0$ ;
  repeat
    compute the gradient of  $f$  at the current set of angles  $\bar{\theta}$ ;
    take a step  $\gamma$  in the direction opposite to the gradient and obtain the new set of angles  $\theta^k$ ;
    create and solve the LP-problem corresponding to  $\theta^k$ ;
    obtain  $x^{*k}$  and  $z^{*k}$ ;
    while ( $\gamma > \gamma_0$ )
      if ( $z^{*k} < z^{*(k-1)}$ ) break (we have found an improved solution, corresponding to  $\theta^k$ , which becomes the new set of angles);
      reduce the step:  $\gamma := \gamma / 10$ 
      take a step  $\gamma$  in the direction opposite to the gradient and obtain the new set of angles  $\theta^k$ ;
      create and solve the LP-problem corresponding to  $\theta^k$ ;
      obtain  $x^{*k}$  and  $z^{*k}$ ;
    end
     $\bar{\theta} := \theta^k$ ;  $k := k + 1$ ;  $\gamma := \bar{\gamma}$ ;
  until ( $k = K_{GD}$ );
end.

```

Fig. 2. General structure of the gradient descent method.

```

Simulated Annealing Method
begin
   $l := 1$ ;  $k := 1$ ;  $\theta := \bar{\theta}$ ; (set of angles corresponding to the local minimum found by
  GD, the corresponding objective value is  $\bar{z}^*$ )
  repeat
    generate a new set of angles  $\theta^l$  in the neighborhood of  $\theta$  ( $\theta^l := \theta + r^* \alpha$ );
    create and solve the LP-problem corresponding to  $\theta^l$ ;
    obtain  $x^{*l}$  and  $z^{*l}$ ;
    if ( $z^{*l} < \bar{z}^*$ ) then set  $\theta := \theta^l$  (and  $\bar{z}^* := z^{*l}$ );
    else set  $\theta := \theta^l$  (and  $\bar{z}^* := z^{*l}$ ) with probability  $\exp(- (z^{*l} - \bar{z}^*) / t_l)$ ;
    if ( $k = K_T$ ) then  $t_l = T_0^* \exp(-c^* l^{1/d})$  and  $k := 1$ ;
     $l := l + 1$ ;  $k := k + 1$ ;
  until ( $l = K_{SA}$ )
end.

```

Fig. 3. General structure of the simulated annealing method.

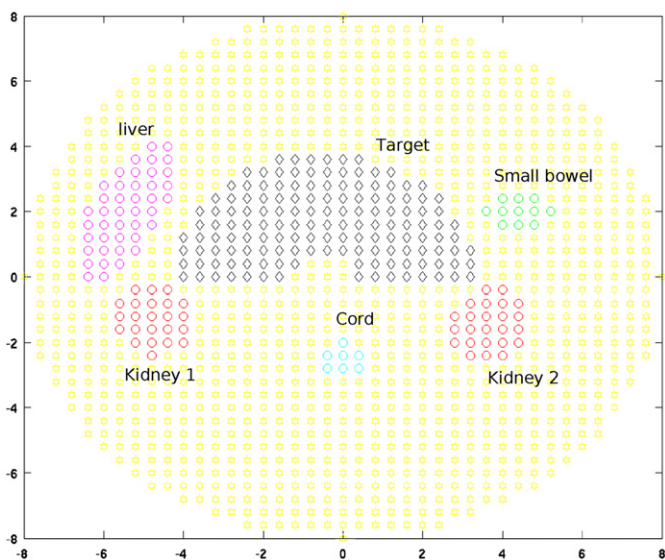


Fig. 4. Geometry of the 2D phantom case.

#### 4.1. Phantom case

The geometry of the phantom case that we consider is shown in Fig. 4. It is a 2D pancreatic tumor case study, and the organs at risk are outlined: kidneys, liver, spinal cord and bowel.

We consider  $n=5$  angles and  $|B_k| = 16$ ,  $k = 1, \dots, n$ , (16 beamlets per angle). The VOIs considered are the following: the tumor set that contains 145 voxels, the set of OARs that contains 42 voxels and the set of the normal tissue that contains 1005 elements (globally 1192 voxels). The matrix  $D$  is computed at 180 angles (i.e.  $\Delta = 2^\circ$ ). The computational tests are executed on a Intel T2300, 1.6 GHz processor, 1 Gb Ram, and the LP-solver used is CPLEX 9.0.

We have performed 40 runs of HM, with a time limit of 1800 s per run. Since this instance is small, we have been able to perform extensive tests on it. This testing has been done in order to get a better feeling of the behavior of HM with respect to gradient descent, to simulated annealing and to the solution obtained with equispaced angles. At each run we started with five equispaced angles. The number of GD iterations was chosen as  $K_{GD} = 10$  and the number of SA iterations as  $K_{SA} = 2$ . The step  $\alpha = 4$  and  $\bar{\gamma} = 0.01$  and  $\bar{\gamma}_0 = 0.00001$ . The initial temperature was set to 1000, the final temperature was set to 0.00001 and  $K_T = 10$ . The minimum

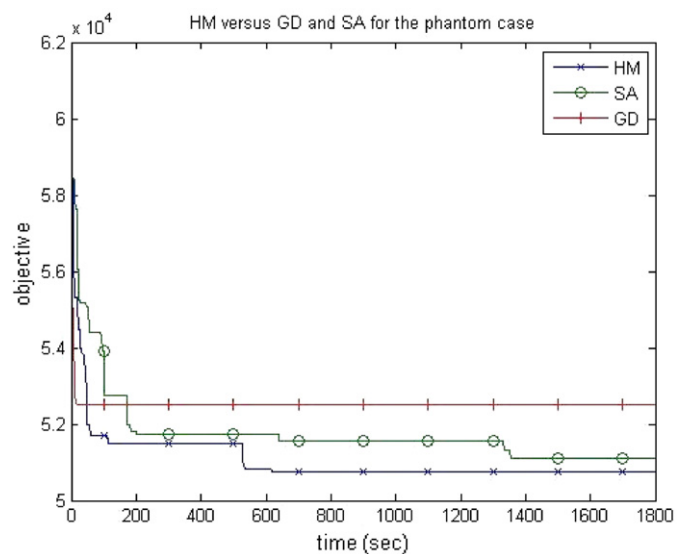


Fig. 5. Comparison of best solutions found by gradient descent, simulated annealing and hybrid method on the phantom case.

dose for each voxel in the tumor was set to 65 Gy (and 0 for voxels in the OARs and in the normal tissue) and the maximum dose for each voxel was set to 75 Gy. Since  $\Delta$  is equal to  $2^\circ$ , we applied the first approach for GD (see Fig. 2).

We present a comparison of the results obtained with HM with the results obtained running (with the same parameters and time limit) the following two methods:

- Gradient Descent procedure described in Fig. 2
- Simulated Annealing procedure illustrated in Fig. 3.

This comparison is to show that HM as a combination of GD and SA outperforms the other two methods. In addition, the starting set of angles is chosen to be equispaced. In Fig. 5 we show the best solution found (over the 40 runs) per time instant for each method. As one can see, HM finds better solutions than the other methods, with an improvement of about 18% in the objective function from the starting set of angles. Moreover, GD turns out to be the worst method.

In order to analyze the sensitivity of HM to the choice of the initial set of angles, we computed the average objective function value and the worst objective function value over the 40 runs

(at each run we started with five different equispaced angles), and compared them to the best solution found. It turns out that the average is 2.4% from the best value and the worst is 4.2% from the best value. As expected, HM is influenced by the choice of the initial set of angles. However, on average, the solution found is not far from the best solution found.

In the following, we show the advantage of allowing HM to choose any angle that can be obtained according to the accuracy of the linear accelerator machine, instead of having a limited choice of predefined candidate angles. To this aim, we develop an MILP model, according to what described in [31].

More precisely, we consider a set  $\Psi = \{\bar{\theta}_1, \dots, \bar{\theta}_m\}$  of candidate equispaced angles as given on input. Let  $m = |\Psi|$ , e.g.  $m=36$  or  $m=72$  equispaced angles. We consider model (1), (3)–(5), (7)–(13), (14), and modify it as follows. We introduce a binary variable  $\psi_k$  for each candidate angle  $\bar{\theta}_k$ ,  $k=1, \dots, m$ , assuming value 1 if the corresponding angle is selected in the solution:

$$\psi_k \in \{0, 1\}, \quad k = 1, \dots, m \quad (15)$$

In addition, we insert the following constraint limiting the number of angles that can be chosen in the solution to  $n$  (e.g.  $n=5$ ):

$$\sum_{k=1}^m \psi_k \leq n \quad (16)$$

Finally, we add linking constraints that impose not to use any beamlet belonging to a direction whose corresponding angle is not selected:

$$x_j \leq \bar{I} \psi_k, \quad j \in B_k, \quad k = 1, \dots, m \quad (17)$$

where  $\bar{I}$  corresponds to the maximum intensity achievable by a beamlet. We perform two comparison tests: in the first one, we consider a set of  $m=36$  equispaced angles as candidate angles while in the second one we consider a set of  $m=72$  equispaced angles. The number  $n$  of angles to be chosen is set to five in both cases. For both tests we set a time limit of 3600 s. For the first case, the optimal solution to the MILP model, computed by CPLEX, was obtained in 2236 s with value 53,464.2, i.e. 5% worse than the best solution found by HM. For the second case, the time limit is reached and the best solution found is 53,182.2, i.e. 4.5% worse than the best solution found by HM; the optimal solution for the second case is obtained in 123,823 s (about 34 h) with value 52,251.9, i.e. 2.8% worse than the best solution found by HM. We also compare the average (over 40 runs) results obtained by HM with the solutions obtained by considering 36 or 72 equispaced candidate angles with a time limit of 1800 s: it turns out that the best solution found in both cases (36 or 72) is 2.7% worse than the average obtained by HM and 0.9% worse than the worst solution obtained by HM.

#### 4.2. Real-life case study

We consider a real-life 3D pancreatic case study provided by the Massachusetts General Hospital (MGH) of Boston, MA. The VOIs considered for building a treatment plan are the following: left kidney, right kidney, liver, stomach, skin and pancreas (i.e. clinical target volume). In Fig. 6 we show the geometry of the considered case. As was done in [16], objectives of type 1 and 3 are considered, and in particular: left kidney, right kidney and liver belong to  $O_1$ ; skin and stomach belong to  $O_3$  (described in Section 2.1). This choice, as in [16], is mainly suggested by the experimentation results and the interaction with doctors, that accepted this formulation as the most clinically relevant formulation to use.

We consider different weights for the VOIs in the objective function: in particular, every VOI has weight 1 except from the

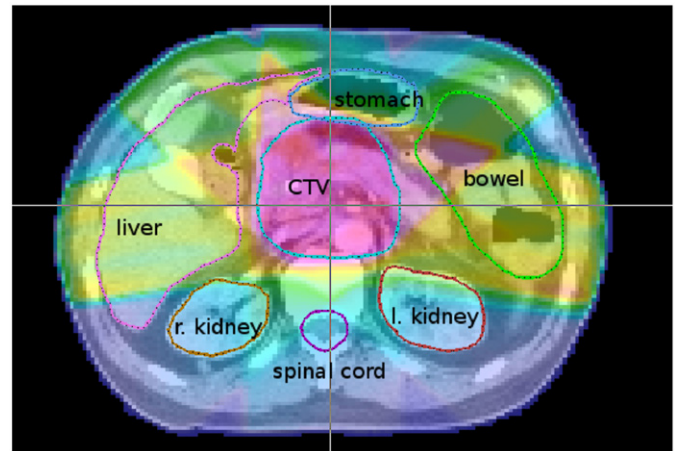


Fig. 6. Geometry of the 3D pancreatic case. For the remaining figures, spinal cord and bowel are not shown since they are not dose limiting structures for the pancreas case study.

one selected as the most important, which gets weight 10. More precisely, we consider 4 instances, each one aiming at a different OAR: left kidney, right kidney, liver and stomach, respectively. This choice is done to simulate the choice of the doctor to spare one OAR with respect to the others. In addition, we consider the instance in which all the OARs get the same weight (equal to 1): we will identify this instance as “All”. The following lower and upper bounds on the doses delivered to the VOIs are used:  $LB=0$  for all the VOIs (left kidney, right kidney, liver, stomach, bowel, spinal cord and skin) except from the target volume, which gets  $LB=50.4$  Gy;  $UB=56.4$  Gy for all the VOIs, except from the spinal cord which gets  $UB=45$  Gy. For skin and stomach, which belong to  $O_3$ , the left slope  $s_h^l$  is set to 0 and the right slope  $s_h^r$  to 1, with  $t_h=37$  Gy for the skin and  $t_h=30$  for the stomach. The number  $N_h$  of voxels in each  $VOI_h$  are the following: bowel=6707, CTV=7822, left kidney=2398, right kidney=1854, liver=4574, skin=36,945, spinal cord=1172 and stomach=1124. Note that a bigger VOI does not necessarily contain more voxels, since we downsample the voxels differently for different structures. In particular, we use the following sample rates: bowel=4, CTV=2, left kidney=2, right kidney=2, liver=4, skin=8, spinal cord=2 and stomach=4. A sample rate of  $s$  means that the dose is computed for one out of every  $s$  voxels in that structure. We consider five angles from which to send the radiation to the patient and 112 beamlets per angle. Beamlets are  $1 \text{ cm} \times 1 \text{ cm}$ .

The dose influence matrix  $D$  is computed every  $\Delta = 5^\circ$ , due to the very large amount of data in the real-life case. Thus, the second approach for GD is used (described in Section 3). The value of  $\delta_{min}$  is set to 0.5. In addition,  $K_{CD}$  is set to 3 and  $K_{SA}$  is set to 1. This choice is due to the following reasons: on one hand, we deal with a larger amount of data when considering the real-life case, thus the computing time for solving the LP problem is larger; on the other hand, the gradient information is less trustable, since we compute the dose influence matrix every  $5^\circ$  (instead of  $2^\circ$  as in the phantom case). The initial and final temperatures and  $K_T$  are set as for the phantom case.

The hybrid method was tested on a Linux machine with 2.66 GHz processor, 16 GB ram, Intel Xeon and we used CPLEX 11.0 as a general purpose solver for solving the LP problem, when the set of angles is fixed. We set a time limit of 80,000 s (22.2 h), which is a reasonable time for computing a treatment plan if one can guarantee that the final result will be acceptable without needing further modifications. We have observed that higher computing time brings little (or even irrelevant) percentage improvement of the solution.

In Fig. 7 we show as an example the execution of the hybrid method (within the time limit of 80,000 s) for the case strongly weighting the left kidney. In particular, this is the plot of the evolution of the objective function value versus the number of iterations of HM. We show in dotted line the iterations when the gradient information is used. As we can see, most of the time the gradient information is reliable even if it is an approximation (due to the interpolation, indeed we evaluate the dose influence matrix every 5°), and it helps to quickly get an improvement in the objective value. When the gradient step leads to a worsening of the cost function, it is due to the gradient approximation error, since a shrinking of the step size in these cases did not correct the situation.

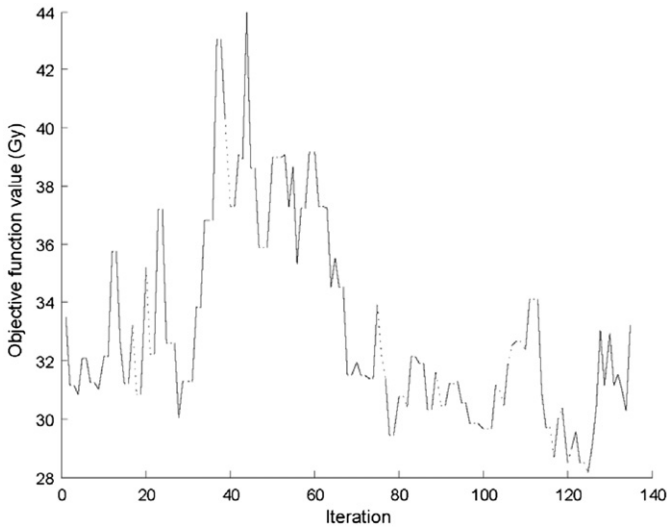


Fig. 7. Hybrid execution for left kidney objective.

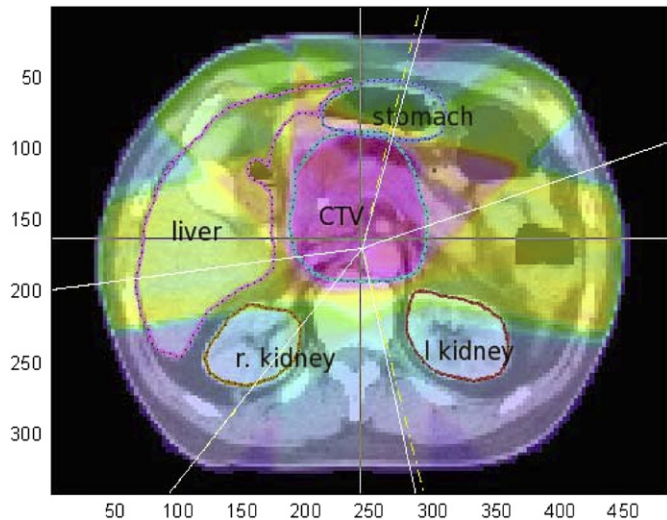


Fig. 8. Change of the angles (solid lines) caused by the gradient descent step for left kidney.

In Fig. 8 we consider the same case, aiming at sparing the left kidney. We show the set of angles (dotted lines) chosen at one iteration of HM and the change of the angles (solid lines) caused by the gradient descent (in this case, the best solution found by HM is obtained after a gradient step). For a better understanding, the angles and the doses delivered to the OARs are listed in Table 1, before and after the execution of the gradient descent.

As Table 1 shows, two angles out of five have been changed by 2° each (14.9 changed to 16.9 and 164.3 changed to 166.3). In this case, the best solution found by HM is obtained by this final step of GD.

We perform a comparison of HM with the solution obtained with a pure SA approach. We also compare the best solution found by HM with the solution obtained with equispaced angles (0,72,144,216,288). The comparison is presented in Table 2, where we show the objective values obtained (for each instance aiming at a different OAR and for the “All” instance) in the case of equispaced angles, HM and SA. We also show the percentage improvement obtained by HM and by SA with respect to the equispaced solution. As we can see, HM turns out to be better than the solution obtained with equispaced angles and than the solution of SA, taking advantage from the gradient information in all the cases.

In Table 3 we show the delivered mean doses (expressed in Gray (Gy)) to the OARs for each instance in the case of equispaced angles (the initial setting of HM) and in the best final solution found, according to the described objective types. In addition, we show the dose delivered to the target. As one can see, both the

Table 2 Comparison of HM with the solution obtained with equispaced angles and SA.

Instance	Equisp. obj.	HM obj.	%Impr.	SA obj.	% Impr.
Left kidney	33.48	28.15	15.9	30.04	10.3
Right kidney	32.49	25.76	20.7	27.5	15.3
Liver	120.75	106.48	11.8	111.95	7.3
Stomach	40.81	35.50	13.0	37.07	9.2
All	24.70	19.87	19.5	21.45	13.1

Table 3 OAR mean doses and target dose (in Gy) for the initial case of equispaced angles and the final case (best solution found by HM).

Instance	Left kidney	Right kidney	Liver	Stomach	Skin	Target
Left kidney obj in. dose	0.5	8.2	12.5	3.1	4.2	51.9
Left kidney obj fin. dose	0.3	4.3	13.5	3.5	3.9	52.8
Right kidney obj in. dose	7.6	0.4	13.9	2.5	3.9	53.2
Right kidney obj fin. dose	2.0	0.4	13.5	2.1	3.9	53.1
Liver obj in. dose	4.8	7.1	10.1	4.2	3.7	52.9
Liver obj fin. dose	2.4	4.2	9.1	4.8	4.0	53.2
Stomach obj in. dose	3.7	3.4	13.4	1.6	3.8	52.9
Stomach obj fin. dose	1.3	1.9	12.5	1.5	3.9	52.9
All obj in. dose	2.9	3.6	12.2	3.7	2.2	53.2
All obj Fin. dose	1.1	1.2	11.5	3.9	2.1	52.9

Table 1 HM objective value and doses delivered to the OARs before and after applying the gradient step, and corresponding change in the angles.

Iteration	HM Obj.	LK	RK	Liver	Skin	Stomach	Ang1	Ang2	Ang3	Ang4	Ang5
Before GD	28.47	0.36	3.65	14.09	3.89	3.20	14.9	72.8	164.3	221.3	263.3
After GD	28.15	0.28	4.33	13.55	3.94	3.48	16.9	72.8	166.3	221.3	263.3



intensities and the angles are selected so as to keep a low dose to the selected organ and also to reduce the total sum of the doses delivered to all the OARs in the final setting. We also notice that sometimes we get an increase in the dose to the normal tissue or to a specific organ when changing the set of angles: this is mainly due to reducing the total sum of the doses delivered to all the organs, while satisfying the constraints on the prescribed dose to the target.

The comparison between HM and the pure SA, when starting from the set of equispaced angles (0,72,144,216,288), as shown in Table 2, is also shown in the graphics in Fig. 9. In particular, we present a figure for each instance aiming at a different OAR and for the “All” instance and we show the best objective value (expressed in Gy) against computing time (expressed in seconds until the time limit of 80,000 s) for HM and SA.

In Figs. 10 and 11 we present in solid lines the set of angles chosen in the best solution found and in dotted lines the set of equispaced angles. These solutions correspond to the objective values in Table 2 obtained by HM. As one can see, the final set of angles is often not intuitive if compared to the equispaced angles, but leads to better treatment plans, in reasonable computing times.

Finally, in Table 4, we show the results obtained by HM within a shorter computing time: in particular, we consider a time limit of 30 min, 1 h and 5 h, respectively. We report as a comparison the solution derived with equispaced angles and the corresponding percentage improvement obtained by HM. As it can be seen from the results, the improvement obtained by HM is significant already after 1 h of execution and it increases after 5 h. This means that HM is also effective when a shorter time limit is imposed. If a larger computing time is available (which is usually the case when a treatment plan is computed), HM can lead to better results (see Table 2).

We can deduce from the presented results that HM takes advantage from the gradient information, leading to better results than using equispaced angles or pure simulated annealing both when aiming at a single OAR and when all the OARs get the same weight. The combination of gradient descent and simulated annealing is very effective: gradient information helps to quickly find a local minimum, while simulated annealing allows to escape from local minima. In addition, the comparison to the method described in [31], which selects the set of angles for the treatment plan out of a set of candidate angles, shows the advantage of HM

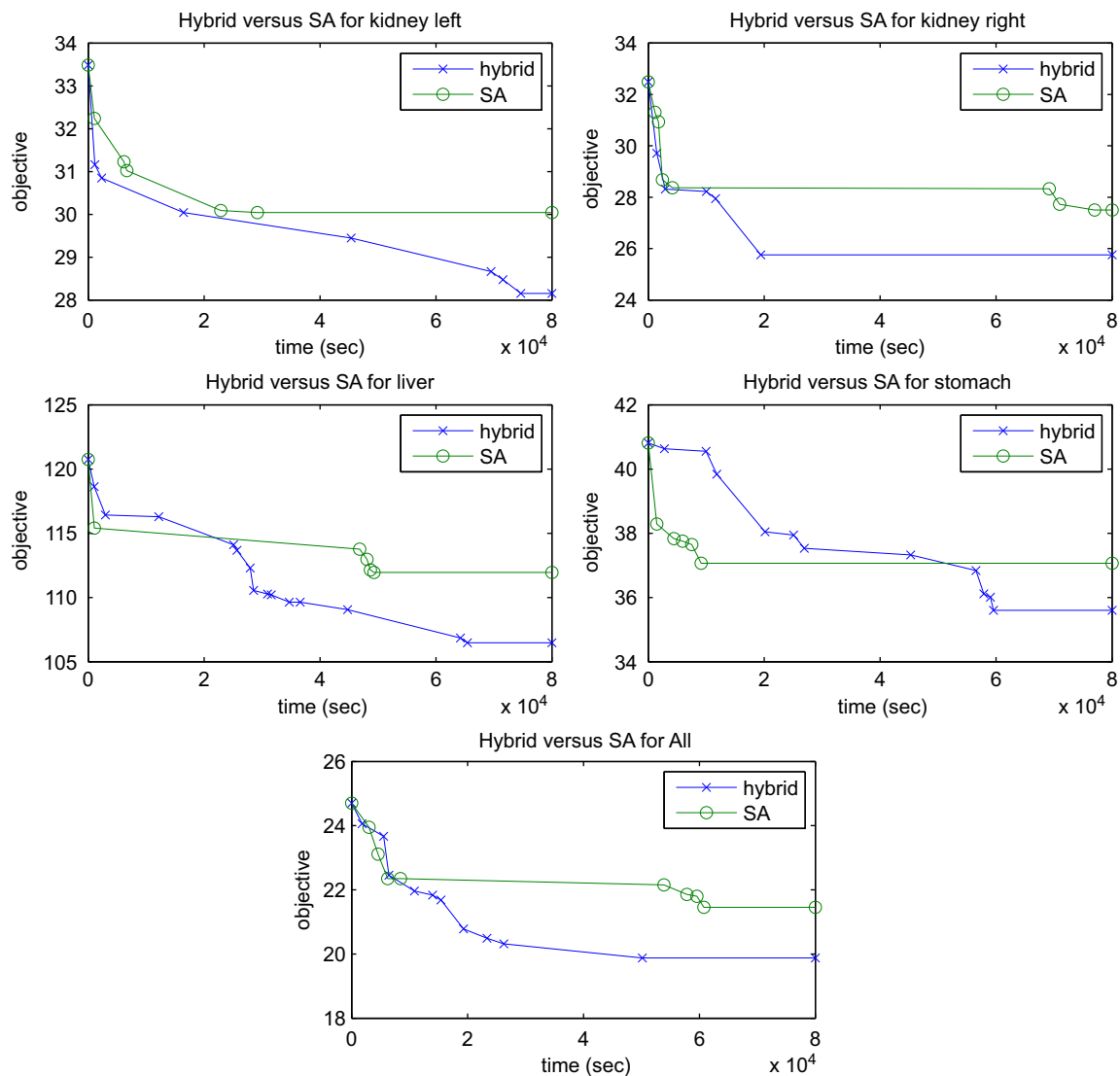
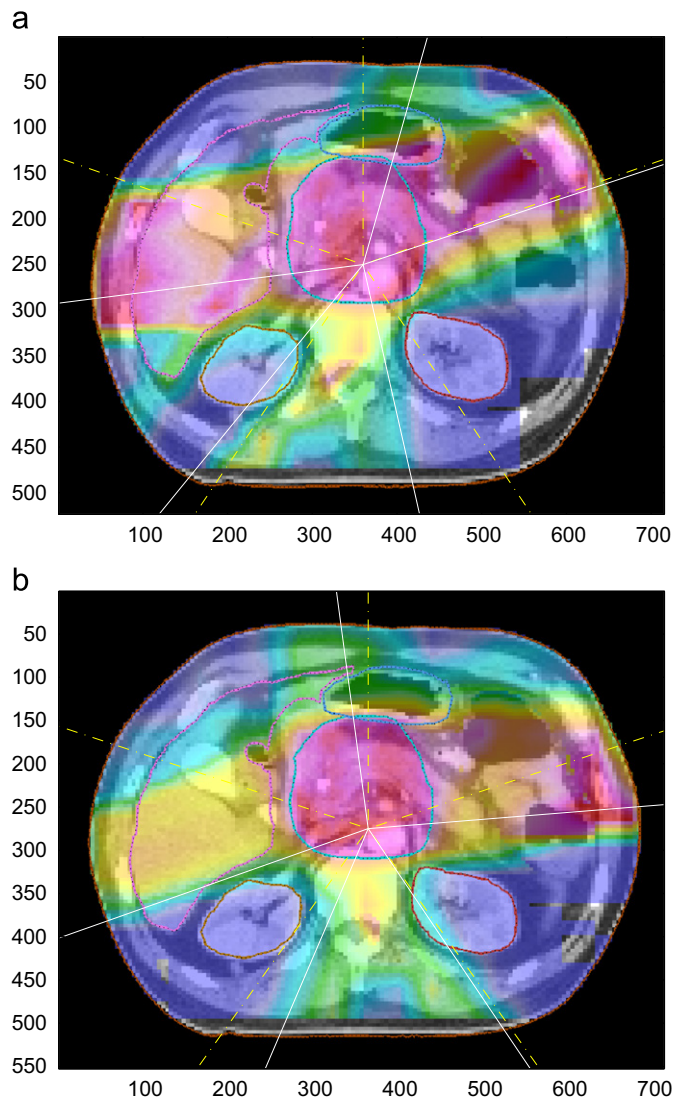


Fig. 9. Comparison between the hybrid method and the pure SA, starting from equispaced angles (one comparison for each organ as main objective, and the comparison for the combined objectives).

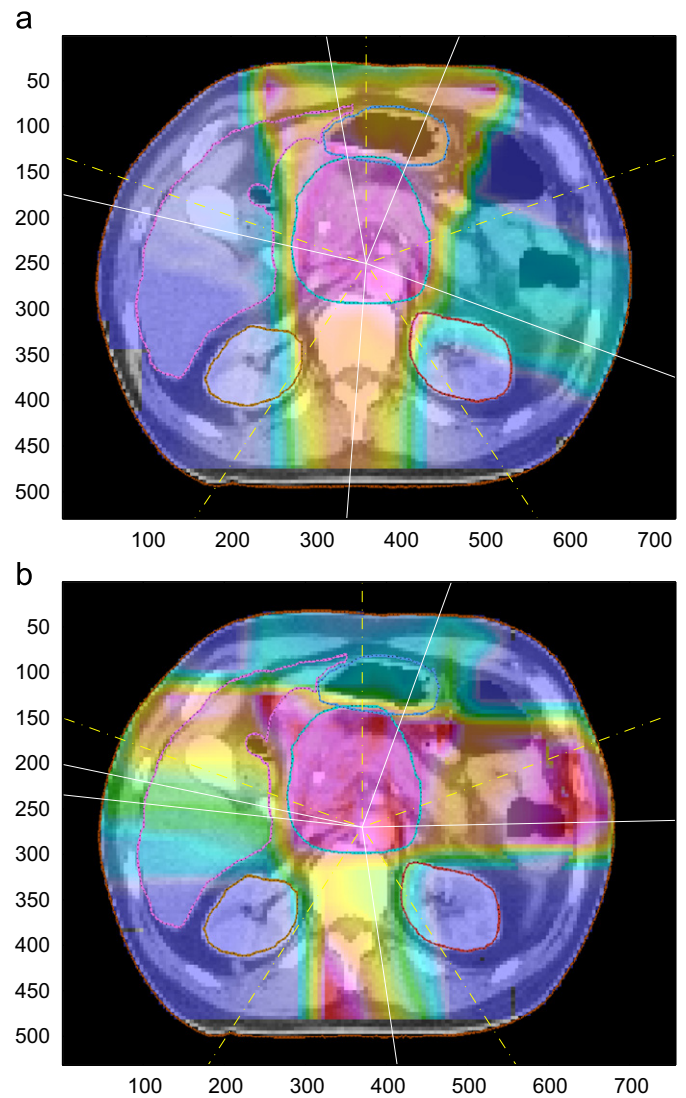


**Fig. 10.** Final set of angles for the left kidney objective (a) and for the right kidney objective (b).

of allowing to choose any angle that can be obtained according to the accuracy of the linear accelerator machines.

## 5. Conclusions

We have studied the Beam Angle Optimization problem, where the directions for delivering radiation to cancer patients are optimized together with beam intensities for building an optimal treatment plan. The aim is to spare the organs at risk, while reaching the prescribed dose to the tumor. The problem is often solved in two phases: firstly, the treatment planners decide the delivery directions, and secondly the intensities of the beams are optimized in an automated way. We have developed a hybrid heuristic algorithm for finding good solutions to the Beam Angle Optimization problem. It alternates some iterations of gradient descent with some iterations of simulated annealing: gradient information is used to quickly find a local minimum, while simulated annealing is aimed to escape from local minima. The optimization of the intensities, when a set of angles has been chosen, is done by solving a linear programming model with Cplex. The presented method differs from previous approaches since it does not require to have on input a set of candidate



**Fig. 11.** Final set of angles for the liver objective (a) and for the stomach objective (b).

**Table 4**

Results obtained by HM in shorter computing time.

Instance	Equisp. obj.	HM obj.					
		30'	% Impr.	1 h	% Impr.	5 h	% Impr.
Left kidney	33.48	31.16	6.9	30.84	7.9	30.04	10.3
Right kidney	32.49	29.71	8.5	28.31	12.9	27.95	13.9
Liver	120.75	118.63	1.7	116.42	3.6	116.29	3.7
Stomach	40.81	40.81	0.0	40.63	0.4	39.84	2.4
All	24.70	24.06	2.6	24.06	2.6	21.68	12.2

angles: angles could vary in the continuous space, provided the accuracy of the linear accelerator machine used. Moreover, the proposed method overcomes the difficulty of searching over a power set of angles by using calculus. In addition, it takes advantage both from the global search (simulated annealing phase) and local search (gradient descent phase). Combining these two techniques leads to a more trustable search (instead of just applying “guessing” heuristics), which often gives as best set of angles a non-intuitive choice (see also Stein et al. [28]). The presented method has been tested on a phantom case and on a real-life case. A comparison with pure simulated annealing, often

used to solve Beam Angle Optimization, has been performed, showing the effectiveness of the hybrid approach, which has obtained an improvement in the solution value between 4% and 7%. In addition, a comparison with the solution obtained using equispaced angles has been shown, reaching an improvement in the solution value always above 10%, in reasonable computing times. Future research can be devoted to speed up the computing time of the hybrid method in order to be able to execute it starting from different sets of angles, which can lead to further improvement. Moreover, additional real-life requirements can be taken into account: for example, it would be interesting to evaluate how the obtained solution can be implemented by using a multi-leaf collimator. We aim to incorporate the direct-aperture optimization with the angle optimization that may circumvent the uncertainties arising when transforming beam intensities to leaf sequences.

### Acknowledgements

We are very grateful to Thomas Bortfeld for the fruitful discussion and help during the project. We are very grateful to the anonymous referees for their helpful comments.

### References

- [1] Ahnesjö A, Saxner M, Trepp A. A pencil beam model for photon dose calculation. *Medical Physics* 1992;19(2):263–73.
- [2] Aleman DM, Kumar A, Ahuja RK, Romeijn HE, Dempsey JF. Neighborhood search approaches to beam orientation optimization in intensity modulated radiation therapy treatment planning. *Journal of Global Optimization* 2008;42(4):587–607.
- [3] Bertsimas D, Nohadani O, Teo KM. Robust optimization with simulated annealing. *Journal of Global Optimization* 2010;48:2,323.
- [4] Bertsimas D, Nohadani O, Teo KM. Nonconvex robust optimization for problems with constraints. *INFORMS Journal on Computing* 2010;22:44–58.
- [5] Bortfeld T, Schlegel W. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Physics in Medicine and Biology* 1993;38:291–304.
- [6] Craft D. Local beam angle optimization with linear programming and gradient search. *Physics in Medicine and Biology* 2007;52:127–35.
- [7] D'Souza WD, Meyer RR, Shi L. Selection of beam orientations in intensity-modulated radiation therapy using single-beam indices and integer programming. *Physics in Medicine and Biology* 2004;49:3465–81.
- [8] Das S, Cullip T, Tracton G, Chang S, Marics L, Anscher M, Rosenman J. Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. *International Journal of Radiation Oncology, Biology, Physics* 2003;55:215–24.
- [9] Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. *Medical Physics* 2003;30(5):979–85.
- [10] Djajaputra D, Wu Q, Wu Y, Mohan R. Algorithm and performance of a clinical IMRT beam-angle optimization system. *Physics in Medicine and Biology* 2003;48:3191–212.
- [11] Ehr Gott M, Güler Ç, Hamacher HW, Shao L. Mathematical optimization in intensity modulated radiation therapy. *4OR A Quarterly Journal of Operations Research* 2008;6(3):199–262.
- [12] Ehr Gott M, Holder A, Reese J. Beam selection in radiotherapy design. *Linear Algebra and its Applications* 2008;428:1272–312.
- [13] Haas OCL, Burnham KJ, Mills JA. Optimization of beam orientation in radiotherapy using planar geometry. *Physics in Medicine and Biology* 1998;43:2179–93.
- [14] Holder A. Partitioning multiple objective optimal solutions with applications in radiotherapy design. *Optimization and Engineering* 2006;7(4):501–26.
- [15] Hou Q, Wang J, Chen Y, Galvin J. Beam orientation optimization for IMRT by a hybrid method of the genetic algorithm and the simulated dynamics. *Medical Physics* 2003;30:2360–7.
- [16] Hong TS, Craft DL, Carlsson F, Bortfeld T. Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head. *International Journal of Radiation Oncology, Biology, Physics* 2008;72:1208–14.
- [17] Jelen U, Sohn M, Alber M. A finite size pencil beam for IMRT dose optimization. *Physics in Medicine and Biology* 2005;50:1747–66.
- [18] Lee E, Fox T, Crocker I. Integer programming applied to intensity-modulated radiation therapy treatment planning. *Annals of Operations Research* 2003;119:165–81.
- [19] Li Y, Yao J, Yao D. Automatic beam angle selection in IMRT planning using genetic algorithm. *Physics in Medicine and Biology* 2004;49:1915–32.
- [20] Lim G, Choi J, Mohan R. Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning. *OR Spectrum* 2008;30:289–309.
- [21] Lim G, Ferris M, Wright S, Shepard D, Earl M. An optimization framework for conformal radiation treatment planning. *INFORMS Journal on Computing* 2007;19:366–80.
- [22] Liu HH, Jauregui M, Zhang X, Wang XC, Dong L, Mohan R. Beam angle optimization and reduction for intensity-modulated radiation therapy of non-small-cell lung cancers. *International Journal of Radiation Oncology, Biology, Physics* 2006;65(2):561–72.
- [23] Preciado-Walters F, Rardin R, Langer M, Thai V. A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer. *Mathematical Programming B* 2004;101:319–38.
- [24] Pugachev A, Xing L. Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy. *Physics in Medicine and Biology* 2001;46:2467–76.
- [25] Pugachev A, Xing L. Incorporating prior knowledge into beam orientation optimization in IMRT. *International Journal of Radiation Oncology, Biology, Physics* 2002;54:1565–74.
- [26] Schreibmann E, Xing L. Feasibility study of beam orientation class-solutions for prostate IMRT. *Medical Physics* 2004;31:2863–70.
- [27] Schreibmann E, Xing L. Dose-volume based ranking of incident beam direction and its utility in facilitating IMRT beam placement. *International Journal of Radiation Oncology, Biology, Physics* 2005;63:584–93.
- [28] Stein J, Mohan R, Wang XH, Bortfeld T, Wu Q, Preiser K, Ling CC, Schlegel W. Number and orientation of beams in intensity-modulated radiation treatments. *Medical Physics* 1997;24:149–60.
- [29] Vaitheeswaran R, Narayanan VKS, Bhangle JR, Nirhali A, Kumar N, Basu S, Maiya V. An algorithm for fast beam angle selection in intensity modulated radiotherapy. *Medical Physics* 2010;37:6443–52.
- [30] Wang C, Dai J, Hu Y. Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Physics in Medicine and Biology* 2003;48:4065–76.
- [31] Yang R, Dai J, Yang Y, Hu Y. Beam orientation optimization for intensity-modulated radiation therapy using mixed integer programming. *Physics in Medicine and Biology* 2006;51:3653–66.
- [32] Zhang HH, Shi L, Meyer R, Nazareth D, D'Souza W. Solving beam-angle selection and dose optimization simultaneously via high-throughput computing. *INFORMS Journal on Computing* 2009;21(3):427–44.