

Instituto de Engenharia de Sistemas e Computadores de Coimbra
Institute of Systems Engineering and Computers
INESC - Coimbra

Humberto Rocha
Joana Matos Dias

On the Optimization of Radiation Therapy Planning

No. 15

2009

ISSN: 1645-2631

Instituto de Engenharia de Sistemas e Computadores de Coimbra
INESC - Coimbra
Rua Antero de Quental, 199; 3000-033 Coimbra; Portugal
www.inescc.pt



On the Optimization of Radiation Therapy Planning

Humberto Rocha * Joana Matos Dias *,†

September 21, 2009

Abstract

The number of cancer patients continues to grow worldwide. There are several different treatments commonly used, depending on the type and stage of the cancer and include surgery, radiation therapy, chemotherapy, immunotherapy, etc. Here, one will focus on radiation therapy, in particular on intensity modulated radiation therapy (IMRT), where optimization can have an important role in the improvement of the quality of the treatments delivered. The continuous development of new treatment machines contributes to the improvement of the accuracy and better control over the radiation delivery. Optimization research follows the evolution of the machines and technology and has made significant contributions to the improvement of radiation therapy planning. The objective of this state-of-the-art is threefold. First, to introduce the most significant radiation therapy concepts to newcomer researchers in this field. Second, to discuss the latest developments on the optimization of radiation therapy planning. And lastly, to highlight possible interesting directions to continue the work on this vast multidisciplinary field.

Key words. Optimization, Radiotherapy, Inverse Planning.

**INESC-Coimbra, Rua Antero de Quental, 199, 3000-033 Coimbra, Portugal.*

†*Faculdade de Economia da Universidade de Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal.*

1 Introduction

The number of cancer patients continues to grow worldwide. An American Cancer Institute study estimates that more than 33 thousand new cases of cancer surge everyday. Daily, around the globe, 20 thousand persons die of cancer. The same study predicts that, in 2050, there are 27 million of persons with cancer. The estimated number of new cases of cancer in the US for 2009 is close to 1,5 million [2]. This year more than half a million Americans are expected to die of cancer, more than 1500 a day. In Portugal, according to INE - Portuguese National Statistics Institute, around 23 thousand persons die every year victims of cancer. An increase of 20% of cancer cases is expected until 2020 in Portugal. The 5-year relative survival rate for all cancers diagnosed in the US between 1996-2004 is 66%, better than 50% in 1975-1977 [2]. This improvement reflects the effort in prevention and quick diagnostic, as well as an improvement on the quality of the set of treatments offered. There are several different treatments commonly used, depending on the type and stage of the cancer and include surgery, radiation therapy, chemotherapy, immunotherapy, etc. Frequently a combination of those treatments is used to obtain the best results.

This state-of-the-art will focus on optimization applied to radiation therapy treatment planning, in particular to intensity modulated radiation therapy (IMRT), where optimization can have an important role in the improvement of the quality of the treatments delivered. A citation search reveals that the terms radiation and optimization or optimal are linked since 1959 [10]. The use of the first linear programming model in 1968 to assist the design of radiotherapy models [5] started an interaction between operations research (OR) and medical physics leading to a florescent multidisciplinary area of work with an increasing importance. Many review papers have been produced on this multidisciplinary area (see eg. [9, 11, 23, 35]) which is another proof of the interest it has been suscitated.

The goal of radiation therapy is to deliver a dose of radiation to the cancerous region to sterilize the tumor minimizing the damages on the surrounding healthy organs and tissues. Radiation therapy is based on the fact that cancerous cells are focused on fast reproduction and are unable to repair themselves when damaged by radiation, unlike healthy cells. Therefore, the goal of the treatment is to deliver enough radiation to kill the cancerous cells but not so much that jeopardizes the ability of healthy cells to survive. The continuous development of new treatment machines contributes to the improvement of the accuracy



Figure 1: Linear accelerator (linac) rotating through different angles [57].

and better control over the radiation delivery. Optimization research follows the evolution of the machines and technology and has made significant contributions to the improvement of radiation therapy planning [5, 27, 28, 39, 44, 45, 58].

The two most common radiation therapy procedures are teletherapy (or external beam therapy) and brachytherapy. In the first, radiation is delivered from outside the body and directed to the patient's tumor location using specific machines. Each machine originates different types of radiation and they include Cobalt-60 machines (such as Gamma Knife radiosurgery), linear accelerators (such as intensity modulated radiation therapy - IMRT), neutron beam machines, proton beam machines, etc. Radioactive substances are inserted within the tumor region in brachytherapy. Even knowing that the cutting-edge research is nowadays performed for proton beam machines due to the characteristics of the proton radiation, since these machines are not yet widely spread in treatment institutes, the focus will be given to external beam therapy with photon beam machines (see Fig. 1) and particularly to IMRT.

We can consider two different types of radiation therapy treatments: conformal radiation therapy and conventional radiation therapy. In conventional radiation therapy the radiation dose is transmitted to the target through high energy radiation beams, being these beams large enough to irradiate all areas that need to be treated. In conformal radiation therapy

the objective is to be able to achieve a high conformity between the area to be treated and the doses absorbed by the tissues. This is accomplished by the use of small beams, each beam targeting a small area. An important type of conformal radiation therapy is IMRT where the radiation beam is modulated by a multileaf collimator, and a beam can be seen as constituted by several sub-beams (pencil beams, beamlets, bixels), each of them with a given fluence (intensity).

The optimization of a treatment planning can be interpreted as the optimal selection of a given treatment among a possible set of admissible solutions. Given the complexity of the problem, sometimes the treatment planning is done through a trial and error procedure. A given treatment planning is defined, and the absorbed doses are calculated. If these doses are acceptable considering the medical prescription, then the procedure terminates. Otherwise, it continues by manually changing the treatment planning. This is a time consuming process, and has no guarantees of producing high quality treatment plans. This trial and error procedure is usually known as forward planning. Inverse planning consists in calculating the optimal planning treatment given the prescribed doses, by using optimization models and algorithms. Inverse treatment planning allows the modeling of highly complex treatment planning problems and OR has a fundamental role in the success of this procedure.

The objective of this state-of-the-art is threefold. First, to introduce the most significant radiation therapy concepts to newcomer researchers in this field. Second, to discuss the latest developments on the optimization of radiation therapy planning. Third and last, to highlight possible interesting directions for OR researchers to continue to work in this vast multidisciplinary field. The paper is organized as follows. In the next section one describe the radiation therapy problem concepts. The following three sections address the radiation therapy subproblems: geometry problem, intensity problem, and realization problem. In the last section we have the concluding remarks.

2 Radiation Therapy Concepts

The radiation therapy treatment sequence is many times compared to a chain constituted by different links [71], as one can see in Figure 2.

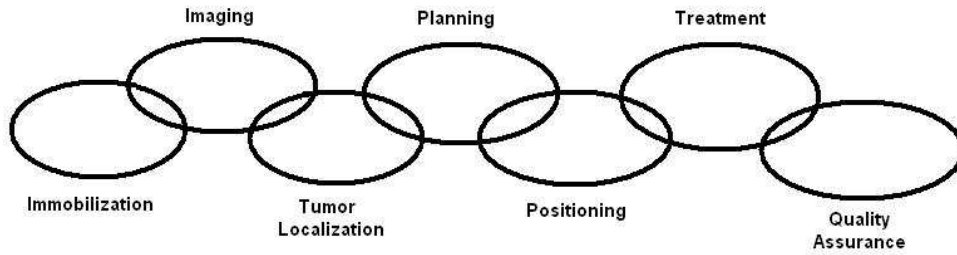


Figure 2: Radiation therapy treatment sequence.

Unlike the worldwide famous television quiz show, *The Weakest Link* (http://en.wikipedia.org/wiki/The_Weakest_Link), the objective is to maintain the chain robustness in Fig. 2, avoiding the existence of weakest links. The first three links in Fig. 2 concerns to the identification and delimitation of the three-dimensional (3D) shape of tumor, organs and tissue in the patient's body near the tumor. In order to do so, advanced 3D imaging techniques such as computer tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) are used. Based on the 3D images, a physician will delimit the following structures [35]:

- **Gross Target Volume (GTV):** represents the volume (macroscopic) of the known tumor.
- **Clinical Target Volume (CTV):** represents the volume of the known tumor (GTV) plus the possible microscopic spread.
- **Planning Target Volume (PTV):** represents the volume of the known tumor and microscopic spread (CTV) plus a marginal volume around the CTV. Usually is the structure used for designing treatment plans, and adding a marginal volume to the CTV is a safety measure to prevent possible inaccuracies or variations (organ or patient motion).
- **Organs at Risk (OARs):** represents the organs in the neighborhood of the tumor that could be damaged by radiation.

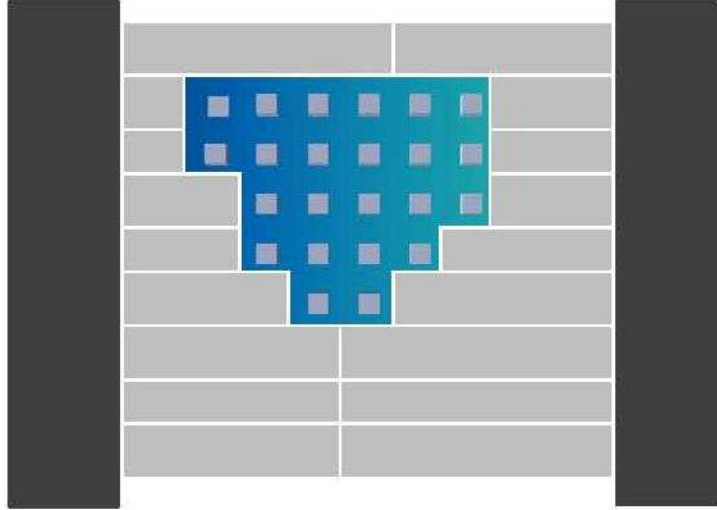


Figure 3: Illustration of a multileaf collimator (with 9 pairs of leaves).

- **Normal Tissue (NT):** healthy tissue where radiation should not accumulate.
- **Critical-normal-tissue-ring (CNTR):** band of normal tissue within Δmm of the PTV (is becoming more popular lately [39] as a device for obtaining conformal plans).

For optimization purposes, each of this volume structures is discretized in voxels (volume elements). The length and width of each voxel depends on the resolution and spacing between images (typically an image (2D) represents 3mm thickness).

Radiation therapy is delivered with the patient immobilized on a couch that can rotate. Typically, radiation is generated by a linear accelerator (linac) mounted on a gantry that can rotate along a central axis (see Fig. 1). The rotation of the couch combined with the rotation of the gantry allows a radiation from almost any angle around the tumor. The intersection point of the central axis of the linac and the rotation axis of the linac gantry is called isocenter. Basically, the isocenter is a geometric reference point, typically placed inside the tumor, to be strategically intersected by the radiation beam.

Despite the fact that almost every angle is possible for radiation delivery, except in rare cases [49], coplanar angles are always considered. This is a way to simplify an already complex problem, and the angles considered lay in the plane of the rotation of the gantry around the patient. Further studies can be pursued both for non-coplanar angles as well as for the case where more than a single isocenter is considered.

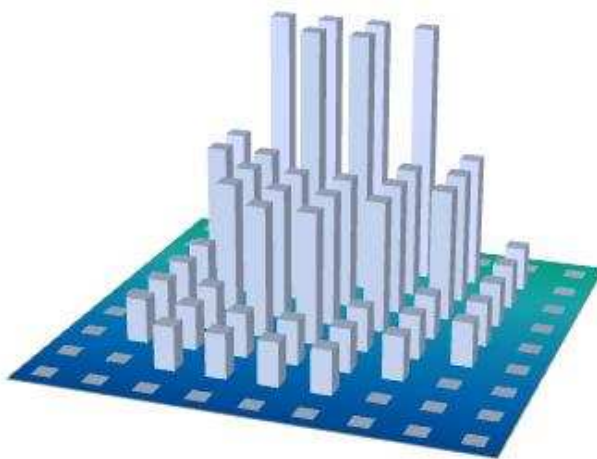


Figure 4: Illustration of a beamlet intensity map (9×9).

The face of the beam is relatively large (e.g. 10×10 cm) but multileaf collimators (MLC) (see Fig. 3) enable the transformation of the beam into a grid of smaller (as small as 3 mm) beamlets of independent intensities (see Fig. 4). Despite the illustration of Fig. 4, beamlets do not exist physically. Their existence is generated by the movement of the leaves of the MLC in Fig. 3 that block part of the beam during portions of the delivery time. The MLC has movable leaves on both sides that can be positioned at any beamlet grid boundary. MLC can operate in two distinct ways: dynamic collimation or multiple static collimation. In the first case, the leaves move continuously during irradiation. In the second case, the “step and shoot mode”, the leaves are set to open a desired aperture during each segment of the delivery and radiation is on for a specific fluence time or intensity. This procedure generates a discrete set (the set of chosen beam angles) of intensity maps like in Fig. 4. Here, one will consider multiple static collimation.

The generation of intensity maps as in Fig. 4 or Fig. 5 (b) allow a high degree of conformity of the delivered distribution dose with the shape of the PTV. Typically, the distributions of dose are represented graphically by the so called isodose distributions. The isodose lines are defined as a percentage of the prescribed dose, i.e. all tissue enclosed by a particular isodose line has received (at least) that percentage of radiation dose. The volume of tissue receiving (at least) that percentage of dose is called isodose volume. Isodose line graphic representation is one of the main tools to assess the quality of the treatment. The

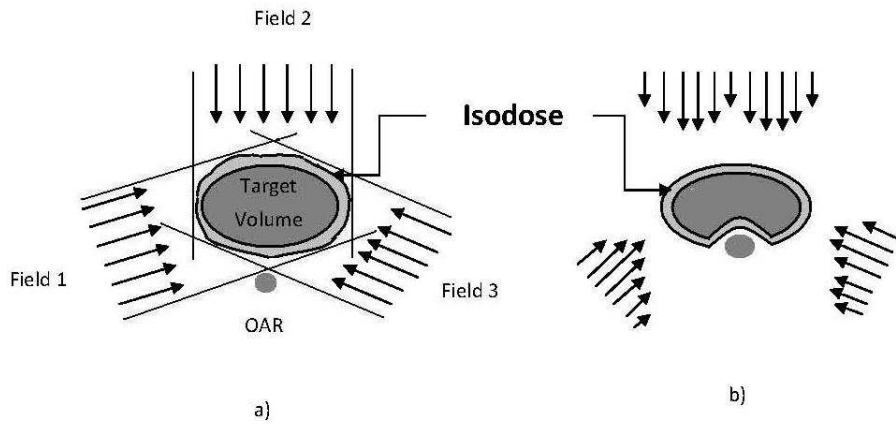


Figure 5: a) Conventional radiation therapy. The organ at risk (OAR) is sufficiently away from the target volume and a convex dose distribution is harmful; b) Conformal radiation therapy with modulated intensity. The isodose fits the target shape in order to contour the OAR with a concave dose distribution.

main advantage of conformal radiation therapy over conventional radiation therapy is the ability to take advantage of modern delivery systems (such as MLC) so that the delivered distribution dose can fit the shape of the PTV even for complex situations (see Fig. 5).

For most types of cancer, radiation therapy is administered 5 days each week for five to eight weeks [45]. Using small radiation doses daily instead of few larger doses helps protect healthy tissues in the tumor region.

In order to design a treatment plan, and since one will focus on dose-based models, the actual delivered dose from the prescribed dose, i.e., the dose that reaches each point of the body need to be calculated. The energy deposited per unit mass of tissue is called absorbed dose and is expressed in the unit called Gray (Gy). One Gy equals one Joule of energy deposited in one kilogram of matter. For each point in the body, the total absorbed dose can be computed as the weighted sum of the dose absorbed due to each beamlet intensity delivered from each angle. The places where the absorbed doses are calculated are called dose-points. Each dose-point is represented through a voxel. Therefore, the planning problem (4th link of Fig. 2) is to determine the angles and the intensity maps so that the resulting dose distribution fulfills the physician's prescription for tumor dose and

restrictions for healthy organs and tissues.

After an acceptable set of intensity maps is produced, one must find a suitable way for delivery (5th and 6th links of Fig. 2). Typically, beamlet intensities are discretized over a range of values (0 to 10, e.g.) and one of the many existing techniques ([6, 13, 65, 59]) is used to construct the apertures and intensities that approximately match the intensity maps previously determined. Due to leaf collision issues, leaf perturbation of adjacent beamlet intensities, etc., the intensity maps actually delivered may be substantially different from the optimized ones. Those problems need to be tackled and are still a prosperous field of research.

The quality of a radiation treatment (last link of Fig. 2) can be compared considering a variety of metrics and can change from patient to patient. Typically cumulative dose histograms (CDVH) are used to describe the amount of radiation received by each relevant structure (PTV, OARs, etc.). Isodose graphic representation is another straightforward tool for perception of the quality of the treatment. There are some other parameters that can complement this tools, including

- **coverage** – ratio of the PTV enclosed by the isodose surface prescribed to the total PTV volume (≤ 1).
- **conformity** – ratio between the volume inside the isodose surface prescribed and the volume of the PTV inside that isodose surface (≥ 1).
- **homogeneity** – ratio between the maximum and minimum dose received by PTV.

Another important aspect when verifying quality is the existence of cold spots – under radiated points or hot spots – over radiated points. Cold spots in the tumor area can jeopardize the whole treatment because the cancer can spread through there and hot spots in OAR can be lethal to the organ functioning.

However, in large measure, the quality of a radiation treatment (for inverse planning) depends on how well the following three subproblems are tackled.

3 Geometry Problem

Radiation therapy is delivered with the patient immobilized on a couch that can rotate. Typically, radiation is generated by a linear accelerator (linac) mounted on a gantry that can rotate along a central axis (see Fig. 1). The rotation of the couch combined with the rotation of the gantry allows a radiation from almost any angle around the tumor. In clinical practice, most of the times, beam directions are still manually selected by the treatment planner that relies mostly on his experience. Theoretically, the objective of the geometry problem, or beam angle optimization problem, is to find the minimum number of beams and corresponding directions that satisfy the treatment goals. However, except for rare exceptions (e.g. [39]), the number of beams is assumed to be defined a priori by the treatment planner. Therefore, the geometry problem consists in the determination of the linac's gantry positions for which radiation is delivered. This problem continues to be addressed by several researchers and further research should be carried on the optimization of the number of beams considered, at least for specific types of cancer.

Despite the fact that almost every angle is possible for radiation delivery, except for rare cases (e.g. [49]), coplanar angles are considered. This is a way to simplify an already complex problem, and the angles considered lay in the plane of the rotation of the gantry around the patient. Further studies should be pursued both for non-coplanar angles as well as for the case where more than a single isocenter is considered.

The beam selection problem is important due to two main reasons. First, the choice of adequate directions is decisive for the quality of the treatment, both for maximizing tumor doses and for OARs sparing. Second, changing beam directions during treatment is time consuming, and short treatments are desirable because the probability of the patient altering his position on the couch increases with the duration of the treatment. Selecting beam directions is still done manually in most health care centers. Typically it requires many trial and error iterations between selecting beam angles and computing fluence patterns until a suitable treatment is achieved. This process is tedious, time consuming (usually taking several hours), has no guarantees of producing good treatments and relies solely on the experience of the treatment planner. The goal of beam selection optimization is to release the treatment planner for other tasks, and at the same time improving the quality of the directions used.

Let us consider p to be the fixed number of (coplanar) beams, i.e., p beams are chosen on a circle around the CT-slice of the body that contains the isocenter (typically the center of mass of the tumor). Resolution of the geometry problem requires a quantitative measure to compare the quality of different sets of beam angles, i.e., for each set of p fixed beam directions, $\theta_1, \dots, \theta_p$, one needs to evaluate an objective function $f(\theta_1, \dots, \theta_p)$. A basic formulation for the beam optimization problem is obtained by selecting an objective function such that the best set of beams is obtained for the function's minimum:

$$\min f(\theta_1, \dots, \theta_p)$$

$$s.t. \quad \theta_1, \dots, \theta_p \in \Theta, \quad \text{where } \Theta \text{ is the set of all possible angles.}$$

Optimum beam orientations depend of a combination of factors including anatomy, radiation tolerance, and the target's prescribed dose. Some of the proposed objective functions for beam orientations are nonlinear and non-convex. Such functions may have numerous local optimums, which increase the difficulty of obtaining a good global solution. Thus, the choice of the solution method becomes a critical aspect for obtaining a good solution.

The objective $f(\theta_1, \dots, \theta_p)$ that measures the quality of the set of beam directions $\theta_1, \dots, \theta_p$ is chosen, in practice, in many different ways, expressing different criteria. Typically, f is defined to be the optimal solution of the beamlets intensities given by $\theta_1, \dots, \theta_p$. This choice of f is straightforward and is based on the fact that for a set of beam directions, a patient will be treated using an optimal fluence map. Let us start by describing a basic linear programming (LP) model to determine the optimal fluence map for a set of beams.

To determine optimal fluence maps, one needs to evaluate the dose distribution in the patient, i.e, it is necessary to calculate how radiation is deposited into the patient. There are many dose models in literature, with the gold standard being a Monte Carlo technique that simulates each particle's path through the anatomy. Available softwares that incorporate dose calculation models include RAD (Radiotherapy optimAl Design software) [1]. This software is written in MATLAB [48] and links to the CPLEX [15] solvers. Another MATLAB software available is CERR - Computational Environment For Radiotherapy Research (<http://radium.wustl.edu/CERR>). CERR is a software platform for developing and shar-

ing research results in radiation therapy treatment planning. The Optimization Research Applications in Radiation Therapy Collaborative Working Group (ORART CWG) [16] developed a suite of MATLAB routines (ORART Toolbox) integrated with CERR, that gives access to the dose distribution matrices.

Let us assume that there are $m \times n$ beamlets identified by the index pair (i, j) . Each voxel in a structure S , $S = PTV \cup OARs \cup NT$, is identified by a three dimensional coordinate (x, y, z) . The weight (intensity) of the beamlet (i, j) delivered over an angle $\theta \in \Theta$ is defined by $w(\theta, i, j)$. Using the superposition principle, the total dose, $D(x, y, z)$, that a voxel (x, y, z) receives is

$$D(x, y, z) = \sum_{(\theta, i, j)} w(\theta, i, j) \cdot d_{(\theta, i, j)}(x, y, z), \quad (1)$$

where $d_{(\theta, i, j)}(x, y, z)$ is the dose delivered to voxel (x, y, z) by beamlet (i, j) from angle θ .

Most of the optimization models in the literature(see e.g. [36, 46, 47, 53, 56, 66]) belong to a class of constrained optimization models such that an objective function is optimized while meeting dose requirements. Given a prescription with a target goal (TG_{PTV}), lower (LB_{PTV}) and upper (UB_{PTV}) bounds for the PTV dose (D_{PTV}), upper bound (UB_{OAR}) for the OAR(s) dose(s) (D_{OAR}), upper bound (UB_{NT}) for the NT dose (D_{NT}) and given an upper bound (M) for the beamlet weight, a simple formulation of the LP model is [46]:

$$\begin{aligned} \min_w \quad & f(D) \\ \text{s.t.} \quad & D(x, y, z) = \sum_{(\theta, i, j)} w(\theta, i, j) \cdot d_{(\theta, i, j)}(x, y, z), \forall (x, y, z) \in S \\ & LB_{PTV} \leq D_{PTV} \leq UB_{PTV}, \\ & D_{OAR} \leq UB_{OAR}, \\ & D_{NT} \leq UB_{NT}, \\ & 0 \leq w(\theta, i, j) \leq M, \forall \theta \in \Theta, i = 1, \dots, m, \\ & \qquad \qquad \qquad j = 1, \dots, n. \end{aligned} \quad (2)$$

A variety of criteria may be considered to be included in $f(D)$, leading to many different objective functions. A simple example of $f(D)$ is presented in Ref. [44] as the weighted sum of maximum deviation from tumor goal dose and maximum overdose to OARs and

normal tissue:

$$f(D) = \alpha_{ptv} \|D_{PTV} - TG_{PTV}\|_{\infty} + \alpha_{OAR} \|(D_{OAR} - UB_{OAR})_+\|_{\infty} + \alpha_{NT} \|(D_{NT} - UB_{NT})_+\|_{\infty}, \quad (3)$$

where $(\cdot)_+ = \max\{0, \cdot\}$ and $\alpha_{(\cdot)}$ are weight factors that can be tuned by the treatment planner. Therefore, even using inverse planning, this continues to be, at some stages, a trial and error process. Another possible objective function considers the minimization of the total number of angles [7].

The linear programming approach has the advantage of being efficiently solved. However, feasible solutions may not be found and producing extreme points as solution may not be desirable since prescription limits are often attained (due to simplex algorithms). Nonetheless, for beam angle optimization, computational time is of the utmost importance, therefore simple LP models are adequate. After deciding on the set of beam angles to use more elaborated models should be used to determine the optimal fluence map. That will be the subject of the next section.

One could think in all possible combinations of p beam angles as an exhaustive global search method. However, this requires an enormous amount of time to calculate and compare all dose distributions for all possible angle combinations. Therefore, an exhaustive search of a large-scale combinatorial problem is considered to be too slow and inappropriate for a clinical setting. For example if we choose $p = 5$ angles out of 72 candidate beam angles (considering angles 5 degrees apart starting from 0°), there are $C_5^{72} = 13991544$ combinations. Solving the fluence optimization problem (linear model (2)) will take more than a minute, therefore, at least 10000 days are required to solve this geometry problem. By decreasing the number of candidate beam angles to 36 (by considering angles 10° apart), the number of different combinations will decrease to $C_5^{36} = 376992$, but almost an year is still required to solve this geometry problem.

Random search techniques based on heuristic approaches such as simulated annealing ([18, 19, 22]), genetic algorithms ([19, 22]) or particle swarm ([43]) have been proposed as alternatives, as well as other heuristics incorporating a priori knowledge of the problem. Another common alternative is scoring methods where scores are assigned to beam angles based on geometric and dosimetric information (see e.g. [19, 46]). Despite the fact that

these methods reduce the computational time, they have the drawback of ignoring the inter-relationship between beam angles by calculating dosimetric parameters from a single incident beam angle plan. Set covering and vector quantization are two other single-step techniques used. A comparison of all those methodologies is presented in Ref. [21] leading to the conclusion that these techniques are very similar and intertwined even knowing that their clinical perspectives may radically differ.

The concept of beam's eye view has been a popular approach to address the geometry problem as well. The concept is similar to a bird's eye view, where the object being viewed is a patient as seen from a beam. The bigger the area of the PTV is seen by the beam, the better candidate the beam is to be used in the treatment plan. Other approaches include the projection of the OARs into the PTV. Pugachev and Xing (see Ref. [54]) present a computer assisted selection of coplanar angles based on scores assigned to each beam of every gantry position. The scores assigned to each beam are based on a variation of the beam's eye view concept. Many others attempts to address the geometry problem can be found in literature. Ehrgott et al. [22] propose a mathematical framework that unifies the approaches found in literature. Acosta et al. [1] focused on how different approximations of the anatomical dose affects the beam selection. Lee et al. [39] suggests a mixed integer programming (MIP) approach for simultaneously determining an optimal intensity map and optimal beam angles for IMRT delivery. This is an interesting large-scale approach but hard to solve even with the increasing computational capabilities of the modern days. It is a mixed integer programming model, with a continuous and a discrete variables associated with each voxel, leading to exhaustion of computer memory!! This approach allows a proper dose-volume control but, in our view, is preferable a 'soft' dose-volume criteria that is tractable. D'Souza et al. [19] proposed a beam ranking procedure (MOD - median OAR dose) that is suited for organs sparing. They also conclude that using 5 beams or more (up to 9) has no significant differences.

Clinical experience has shown that directly opposed beams yield undesirable hot spots in IMRT planning [19]. Therefore, beam angles θ and $\theta \pm 180^\circ$ should not be part of the same beam angle selection. Another common geometric constraint, for the same reason as before, is to consider a minimum angle distance (δ), i.e., angles θ_1 and θ_2 need to verify $|\theta_1 - \theta_2| \geq \delta$ in order to be included in the same beam angle selection. As corollary, one

4 Intensity Problem

After deciding what beam angles should be used, a patient will be treated using an optimal plan obtained by solving the intensity (or fluence map) problem - the problem of determining the optimal beamlet weights for the fixed beam angles. Many mathematical optimization models and algorithms have been proposed for the intensity problem, including linear models (e.g. [59, 60]), mixed integer linear models (e.g. [40, 55]), nonlinear models (e.g. [67]), and multiobjective models (e.g. [61]). One will go over those proposed methods to understand their advantages and disadvantages and what can still be improved.

Radiation dose distribution deposited into the patient, measured in Gray, needs to be assessed accurately in order to solve the intensity problem, i.e., to determine optimal fluence maps. As stated in Eq. 1, by the superposition principle, each voxel (x, y, z) in each structure S receives a total dose, $D(x, y, z)$:

$$D(x, y, z) = \sum_{\theta \in \Theta} \sum_{i=1}^m \sum_{j=1}^n w(\theta, i, j) \cdot d_{(\theta, i, j)}(x, y, z).$$

Typically, a dose matrix D is constructed from the collection of beamlet weights, by indexing the rows of D by (x, y, z) and the columns by (θ, i, j) , i.e., the number of rows of matrix D equals the number of voxels and the number of columns equals the number of beamlets from all angles considered. Usually the total number of voxels considered reaches the hundred thousands and sometimes millions, thus the row dimension of the dose matrix is of that magnitude. The size of D originates large-scale problems being the main reason for the difficulty of solving the intensity problem (as well as the geometry problem).

The first attempts to tackle the intensity problem used LP models. Some of the reasons for the use of LP models include the fact that dose deposition is linear, LP models are easy to implement and are broadly used. Given a prescription with lower and upper bounds for the PTV dose, upper bound for the OAR(s) dose(s), and upper bound for the NT dose, most of the formulations of the LP models belong to a class of constrained optimization models such that an objective function is optimized while meeting these dose requirements. A simple formulation of a LP model was given in the previous section, LP model (2), but here angles are not decision variables:

$$f(D) = \alpha_{ptv} \|D_{PTV} - TG_{PTV}\|_1 + \alpha_{OAR} \|(D_{OAR} - UB_{OAR})_+\|_1 + \alpha_{NT} \|(D_{NT} - UB_{NT})_+\|_1, \quad (5)$$

$$f(D) = \alpha_{ptv} \|D_{PTV} - TG_{PTV}\|_2^2 + \alpha_{OAR} \|(D_{OAR} - UB_{OAR})_+\|_2^2 + \alpha_{NT} \|(D_{NT} - UB_{NT})_+\|_2^2, \quad (6)$$

There are three common ways to define the objective:

- use the L_1 norm (5) – penalizes the absolute value of deviation from prescribed dose on each voxel, weighted according to the region in which voxel lies.
- use the L_2 norm (6) – penalizes the sum of squares of the deviations, weighted by the factors defined.
- use the L_∞ norm (3) – penalizes “hot spots” in sensitive regions (OARs and NT) and cold “spots” in the PTV.

The weight factors, $\alpha_{(\cdot)}$, need to be tuned by the treatment planner. In order to facilitate the choice of this weight parameters and reduce the dependence of the plan on the size of each structure (PTV, OARs, and NT) one can consider an average dose deviation in each structure for the objective function [44]:

$$f(D) = \alpha_{ptv} \frac{\|D_{PTV} - TG_{PTV}\|_p}{card(PTV)} + \alpha_{OAR} \frac{\|(D_{OAR} - UB_{OAR})_+\|_p}{card(OAR)} + \alpha_{NT} \frac{\|(D_{NT} - UB_{NT})_+\|_p}{card(NT)}, \quad p = 1, 2, \infty, \quad (7)$$

where $card(\cdot)$ denotes the total number of voxels in the considered structure. Note that the use of the L_2 norm originates nonlinear quadratic models.

Different goals, corresponding to different structures, can be considered by using different norms in the same objective function. Some of the most frequently used composite formulations include [44]

$$f(D) = \alpha_{ptv} \|D_{PTV} - TG_{PTV}\|_\infty + \alpha_{OAR} \frac{\|(D_{OAR} - UB_{OAR})_+\|_1}{card(OAR)} + \alpha_{NT} \frac{\|(D_{NT} - UB_{NT})_+\|_1}{card(NT)}, \quad (8)$$

when the goal is to have low maximum dose violation on the PTV and simultaneously to control the average dose deviation on the OARs and NT,

$$f(D) = \alpha_{ptv}^+ \frac{\|(D_{PTV} - UB_{PTV})_+\|_1}{\text{card}(PTV)} + \alpha_{ptv}^- \frac{\|(LB_{PTV} - D_{PTV})_+\|_1}{\text{card}(PTV)} + \quad (9)$$

$$\alpha_{OAR} \frac{\|(D_{OAR} - UB_{OAR})_+\|_1}{\text{card}(OAR)} + \alpha_{NT} \frac{\|(D_{NT} - UB_{NT})_+\|_1}{\text{card}(NT)},$$

or

$$f(D) = \alpha_{ptv}^+ \|(D_{PTV} - UB_{PTV})_+\|_\infty + \alpha_{ptv}^- \|(LB_{PTV} - D_{PTV})_+\|_\infty + \quad (10)$$

$$\alpha_{OAR} \frac{\|(D_{OAR} - UB_{OAR})_+\|_1}{\text{card}(OAR)} + \alpha_{NT} \frac{\|(D_{NT} - UB_{NT})_+\|_1}{\text{card}(NT)},$$

when the goal is to penalize differently cold and hot spots on the PTV. Other alternative objective functions can be formulated and have been tested.

The most important drawback of using LP models lies on the fact that simplex methods produce admissible solutions that are extreme points (where restrictions are satisfied as equalities), therefore the prescription limits are often attained. This means that, most of the times, the optimal solution assigns the maximum allowable dose to OARs and/or the tumor receives the lowest allowable dose. The LP approach has the advantage of being fast. However, if the constraints are too tight, a feasible solution is hard (or impossible) to find and the source of infeasibility is unknown. Holder [35] proposed a new LP formulation to overcome the infeasibility problem incorporating elastic constraints:

$$\begin{aligned} \min \quad & \lambda^T \alpha + u_{OAR}^T \beta + u_{NT}^T \gamma \\ \text{s.t.} \quad & LB_{PTV} - L\alpha \leq D_{PTV} \leq UB_{PTV}, \\ & D_{OAR} \leq UB_{OAR} + U_{OAR}\beta, \\ & D_{NT} \leq UB_{NT} + U_{NT}\gamma, \\ & 0 \leq L\alpha \leq LB_{PTV}, \\ & -UB_{OAR} \leq U_{OAR}\beta, \\ & 0 \leq U_{NT}\gamma, \\ & 0 \leq w \leq M. \end{aligned} \quad (11)$$

The constraints $LB_{PTV} - L\alpha \leq D_{PTV} \leq UB_{PTV}$, $D_{OAR} \leq UB_{OAR} + U_{OAR}\beta$, and $D_{NT} \leq UB_{NT} + U_{NT}\gamma$ are called elastic constraints because the bounds can vary with the vectors

α , β , and γ . The amount of elasticity is defined by the matrices L , U_{OAR} , and U_{NT} and the corresponding penalty terms in the objective function are l , u_{OAR} and u_{NT} , respectively. Scalar λ is the weight deciding the importance of tumor uniformity. These elastic constraints guarantee that both the primal and dual of (11) are feasible meaning that in those situations where medical prescription leads to an infeasible problem, the model can still originate a solution that is the closest possible to the infeasible prescription.

Sometimes the treatment objectives are clinically defined as a percentage of each type of tissue receiving a given radiation dose. This motivates the appearance of dose-volume constraints. It is typically acceptable that, for instance, the PTV receives 95% of the prescription dose. Moreover, different organs react differently to radiation. Organs like the lung or the liver are “modular” organs, i.e., they are capable of functioning even with substantial parts of their tissues destroyed. Organs like the spinal cord or the bowel are “chain” organs, i.e., they are unable of functioning even if only a small portion of tissue is destroyed. Dose-volume constraints are biologically meaningful and allow to model the physician’s goals in terms of percent tissue receiving a prescribed dose.

The use of mixed integer linear models (e.g. [40, 55]), appear naturally with the introduction of dose-volume constraints. By introducing binary variables (0/1 variables) and assigning a variable to each voxel it is very simple to introduce dose-volume constraints and to quantify the voxels that receive a dose lower or higher than the threshold in each structure.

The dose constraints in the LP models above will now be replaced by dose-volume constraints. As an example, the upper bound constraint on the OAR dose in LP (2)

$$D_i w \leq UB_{OAR}, \forall i \in OAR \quad (\text{each line of D is assigned to a voxel})$$

would be replaced by the dose-volume constraints,

$$D_i w \leq (1 + y_i F) UB_{OAR}, \forall i \in OAR$$

$$\sum_{i \in OAR} y_i \leq P \times \text{card}(OAR),$$

where y_i are the binary variables, F is maximum overdose allowed and P is the maximum percentage of points that can receive a dose above the maximum dose UB_{OAR} . The binary

variables y_i associated to a voxel i of a OAR are defined by

$$y_i = \begin{cases} 1, & \text{if radiation dose exceed the OAR upper bound dose} \\ 0, & \text{otherwise.} \end{cases}$$

The dose-volume constraints guarantee that no more than $P\%$ volume of the OAR can exceed $F\%$ of the radiation dose bound UB_{OAR} . Compared with the stricter bound of LP that imposes a bound UB_{OAR} for all OAR volume, the dose-volume constraints confer MIP models more flexibility avoiding feasibility issues. Note that this is a simple formulation for a single OAR that need to be split by the number of OARs with correspondent upper bounds for the cases where more than one OAR is at stake.

Similarly, LP constraints for PTV and normal tissue can be transformed in dose-volume constraints. Some authors (e.g. [39, 46]) argue that normal tissue in the neighborhood of the PTV should be allowed to receive more radiation than normal tissue further from the PTV. Therefore, they divide the voxels of normal tissue (NT) in two sets: critical-normal-tissue-ring (CNTR) – band of normal tissue within δmm of the PTV, and remaining normal tissue $\overline{NT} = NT/CNTR$.

Many MIP models have been proposed (e.g. [39, 55, 46]). A simple example of a MIP can be derived from LP (2) where dose-volume constraints replace the strict dose bounds:

$$\begin{aligned} \min \quad & f(D) \\ \text{s.t.} \quad & D(x, y, z) = \sum_{(\theta, i, j)} w(\theta, i, j) \cdot d_{(\theta, i, j)}(x, y, z), \\ & (1 - y_i^{PTV^-} F_{PTV}^-) LB_{PTV} \leq D_i w, \forall i \in PTV, \\ & D_i w \leq (1 + y_i^{PTV^+} F_{PTV}^+) UB_{PTV}, \forall i \in PTV, \\ & \sum_{i \in PTV} y_i^{PTV^-} \leq P_{PTV}^- \times \text{card}(PTV), \\ & \sum_{i \in PTV} y_i^{PTV^+} \leq P_{PTV}^+ \times \text{card}(PTV), \\ & D_i w \leq (1 + y_i^{OAR} F_{OAR}) UB_{OAR}, \forall i \in OAR, \\ & \sum_{i \in OAR} y_i^{OAR} \leq P_{OAR} \times \text{card}(OAR), \\ & D_i w \leq (1 + y_i^{NT} F_{NT}) UB_{NT}, \forall i \in NT, \\ & \sum_{i \in NT} y_i^{NT} \leq P_{NT} \times \text{card}(NT), \\ & 0 \leq w(\theta, i, j) \leq M, \forall \theta \in \Theta, i = 1, \dots, m, \\ & \quad \quad \quad j = 1, \dots, n, \end{aligned}$$

where $y_i^{(\cdot)}$ are binary variables, $F_{(\cdot)}^-$ are underdose maximum limits, $F_{(\cdot)}$ and $F_{(\cdot)}^+$ are overdose maximum limits and $P_{(\cdot)}$ are volume percentages.

All the objective functions considered before can be used in MIP models. Volume-based objective functions can also be considered. Conformity, homogeneity or total intensity are other possible objectives. Moreover, MIP problems can be used to solve combined problems (as seen in model (4) in previous section). Lee et al. [39] formulate a MIP model to simultaneously solve the beam geometry and the intensity map optimization problems. Preciado-Walters et al. [55] formulate a MIP model to simultaneously solve the intensity map optimization and the realization (also known as segmentation or aperture optimization) problems. In fact, MIP approach has the advantage of allowing the formulation of more complex and flexible models. However, an important drawback is that MIP models are much harder to solve, implying the use of, for instance, branch and bound algorithms. Since the number of integer variables is at least equal to the number of voxels (in the order of the hundred thousands or millions), finding an optimal solution is very difficult at least in a clinical acceptable time.

If one chooses to use the objective function (6), the resulting model is a quadratic program. The most prevalent formulation amongst nonlinear models is the quadratic programming formulation (e.g. [67]). A simple example of a quadratic programming model, considering a slight variation of the objective function (7) with $p = 2$, is

$$\begin{aligned} \min \quad & \alpha_{ptv} \frac{\|D_{PTV} - TG_{PTV}\|_2^2}{\text{card}(PTV)} + \alpha_{OAR} \frac{\|(D_{OAR} - UB_{OAR})_+\|_2^2}{\text{card}(OAR)} + \alpha_{NT} \frac{\|(D_{NT} - UB_{NT})_+\|_2^2}{\text{card}(NT)} \\ \text{s.t.} \quad & D(x, y, z) = \sum_{(\theta, i, j)} w(\theta, i, j) \cdot d_{(\theta, i, j)}(x, y, z), \forall (x, y, z) \in S \\ & 0 \leq w(\theta, i, j) \leq M, \quad \forall \theta \in \Theta, i = 1, \dots, m, \\ & \quad \quad \quad j = 1, \dots, n. \end{aligned}$$

Most of the linear models considered previously can be modeled as a quadratic model just by considering the L_2 norm instead of the L_1 or the L_∞ norms. However, most of them don't have any biological meaning. A different modeling approach is to include biological perspectives. A common approach to incorporate radiobiological meaning on objective functions is to use tumor control probability (TCP) or normal-tissue complication probability (NTCP). Typically, the treatment goal is to maximize the tumor control probability

and simultaneously control the growth of normal-tissue complication probability. The difficulty in radiobiological models is to accurately define mathematical formulas for TCP and NTCP. The most used formulas for TCP and NTCP are Poisson-based formulas [23]:

$$TCP = \prod_{i=1}^{card(PTV)} \exp\left[-\frac{\mathcal{O}}{card(PTV)} \exp(-rd_i)\right], \quad (12)$$

$$NTCP = 1 - \prod_{k=1}^K (1 - P^k),$$

$$P^k = \left(1 - \prod_{i \in OAR_k} [1 - (P(d_i))^s]^{\frac{v_i}{\sum_{i \in OAR_k} v_i}}\right)^{\frac{1}{s}},$$

where r is the radiosensitivity of the tumor clonogens (cells capable of producing a colony of similar cell types), \mathcal{O} is the total number of clonogens in the tumor, d_i is the i^{th} voxel dose, v_i is the i^{th} voxel volume, K is the number of OARs, and s describes the relative seriality of tissue organization. Assuming homogeneous dose,

$$P(d_i) = 2^{-\exp(eg(1-d_i/D_{50}))},$$

where e is the Neper number, g is the normalized dose-response gradient, and D_{50} is the dose of 50% response.

Different radiobiological models can be formulated based on TCP and NTCP. Currently, one of the most used models is the maximization of the uncomplicated tumor control probability, P_+ , given by

$$P_+ = TCP \times (1 - NTCP),$$

assuming TCP and NTCP are independent. The advantage of this objective function is that the probabilities of eliminating the tumor avoiding fatal second effects are gathered in a single scalar quantity.

Another radiobiological approach commonly used is the intuitive concept that helps the analysis of dose distributions, the concept of equivalent uniform dose (EUD). EUD is the dose which, if administered uniformly to a structure, would lead to the same biological/clinical effect as a non-uniform dose distribution. EUD was originally based on TCP model (12) and has also been defined as ‘‘generalized mean’’ dose and can be considered as

being between pure physical dose objective and biological planning objectives. EUD can be defined as [23]:

$$EUD = \left(\frac{1}{m} \sum_{i=1}^m d_i^a \right)^{\frac{1}{a}},$$

where m is the number of voxels in the structure at stake and a is a radiation tissue response parameter that depends on the structure considered. Further experiments should be carried out using EUD, since there is little experience with this concept.

The radiotherapy optimization problems discussed so far are multiobjective by nature: there are (at least) two conflicting goals that have to be considered. On one hand it is desirable to achieve high radiation doses in the area to be treated, on the other hand healthy tissues have to be spared [24].

When dealing explicitly with more than one objective, the idea of calculating the optimal solution is replaced by the need of calculating efficient (non-dominated or Pareto-optimal) solutions. The calculation of non-dominated solutions can be done a priori, a posteriori or using an interactive procedure. If solutions are calculated a priori this means that a set of non-dominated solutions are calculated (sometimes using generator methods, that calculate all or a great number of solutions), and then presented to the decision maker. The decision maker chooses one out of the set of solutions presented. In a posteriori methods, the preferences of the decision maker have to be incorporated into the problem (the task of eliciting the decision maker's preferences is a complex one), so that the solution that optimizes the decision maker's utility function is reached. In interactive procedures, there is a calculation phase where non-dominated solutions are calculated and a dialogue phase where the non-dominated solutions are presented to the decision maker, giving the opportunity of expressing preferences, which will guide the next calculation phase and so forth. This process stops when the decision maker finds the preferred solution.

There are several references in the literature that deal explicitly with multiobjective problems. An extensive review of different multiobjective approaches can be found in Ref. [23]. A comparison of different multiobjective models via a unifying mathematical framework is presented in Ref. [61]. Ehrgott and Burjony [24], present a multiobjective problem where the objective function considered represent the minimization of the deviation between the radiation received and prescribed in the several structures defined. In a

first phase, the authors calculate a priori a set of non-dominated solutions, followed by a second phase of interaction with the decision maker (considering the set of solutions already known). Ehrgott and Winz [25, 72], develop a software (CARINA) that calculates efficient solutions considering the minimization of the deviations between the radiation received and the bounds imposed. Cotrutz, Lahanas et al. [14], present a multiobjective model for conformal therapy, calculating non-dominated solutions by the use of a weighted mono-objective function. Hamacher and Kufer [34], consider a multiobjective model, with dose restrictions (upper and lower bounds) and homogeneity restrictions (where a reference voxel is defined, and then the differences between the absorbed dose in this voxel and the others in the PTV are bounded). Kufer, Scherrer et al. [38], present a multicriteria approach, making use of clustering techniques and redundancy tests to deal with the dimension of the problem, developing an a priori generating method. Schreibmann, Lahanas et al. [64], present a genetic algorithm capable of generating solutions when the objective functions considered are: the minimization of the variance of the received dose in relation to the prescribed dose for the PTV; the minimization of the sum of the square of the received doses for normal tissue and critical structures; the minimization of the number of beams used. The genetic algorithm is responsible for finding an optimal number of beams, and the optimization of intensities is performed by other dedicated algorithm. Thieke, Kufer et al. [70], present a multicriteria approach in two phases: first a generator algorithm is used to calculate possible planning treatments. Then the decision maker can interactively explore the non-dominated solutions found. The authors consider as objectives the generalized equivalent uniform dose for OAR, that accounts for the specificities of each organ; for PTV they use a metric that considered the values above or below defined thresholds for the median dose received and standard deviation. Further work should be developed in order to improve the existing approaches and avoiding the trial and error processes.

Regardless the formulation used, size is always the biggest challenge to overcome. One of the strategies to reduce the computing time of the overall process is sampling [29]. By aggregating voxels in certain structures (mainly in normal tissue and OARs) a decrease on the problem size leads to a fast resolution. Usually normal tissue contains many more voxels than any other structure and since it is relative unimportant compared with other vital structures, is therefore the main candidate for aggressive sampling. Further work

should be developed here, since this is one of the best strategies to speed up the overall inverse planning optimization process.

Treatments are usually judged by their cumulative dose-volume histogram (DVH) and by analyzing isodose curves, i.e., the level curves of the radiation by slice. An ideal DVH would have the tumor at 100% for the entire volume and then drop immediately to zero, indicating that the target is treated exactly as prescribed. The curves for the organs at risk would instead drop immediately to zero, meaning that they receive no radiation. The appropriate isodose curve should fit the tumor area for each slice. The quality of the treatment depends on the model and new models can be proposed for specific cases, where a tailored model can improve the general existing ones. However, the quality of the treatment also depends on the algorithm used to solve the proposed model. Many algorithms are presented in literature (e.g. [8, 52, 63, 68, 73, 74]) but new algorithms should be explored, namely for nonlinear models, where the gold standard is simulated annealing. Many other derivative free algorithms should be attempted and compared to acknowledge if some naturally design better solutions.

Typically, most of the commercial software requires that the planner furnishes the beam angles to use. This first human loop is time consuming and was discussed in previous section. Unfortunately, for a given set of beam angles, commercial software finds the optimal fluence map, but requires a second human loop. The objective functions considered lie in one of the (3), (5) – (10) objective functions. Planners need to tune the weights to achieve DVH control on the PTV and OARs. Another alternative is to update the minimum and/or maximum fractions of the prescribed dose that the planner wants to deliver to the PTV and/or OARs. This is again a trial and error process, time consuming, and automated mechanisms should be developed to assist the planner. Multiobjective models should have an important role here.

5 Realization Problem

In order to deliver the optimized intensities resulting from the previous intensity problem optimization, one needs to find an efficient way for MLC devices to produce the exact same optimized intensity profiles. Due to MLC device restrictions, this transposition is

not straightforward. Moreover, reproducing the optimized intensity maps efficiently, i.e. minimizing the radiation exposure time, is a challenging optimization problem.

Many mathematical optimization models and algorithms have been proposed for the realization problem, similarly to the intensity problem. After an acceptable set of intensity maps is produced, one must find a suitable way for delivery. Typically, beamlet intensities are discretized over a range of values (0 to 10, e.g.) and one of the many existing techniques ([3, 6, 13, 50, 59, 65, 69, 75]) is used to construct the apertures and intensities that approximately match the intensity maps previously determined.

There are two distinct methods using MLCs: dynamic collimation or multiple static collimation. In the first case, the leaves move continuously during irradiation while the beam is always on. In the second case, the “step and shoot mode”, the leaves are set to open a desired aperture during each segment of the delivery and radiation is on for a specific fluence time. The mathematical methods used to determine optimal treatment plans are similar, but, here, one will only consider multiple static collimation.

An intensity or fluence map can be expressed by a matrix of $m \times n$ beamlet weights (intensity assigned to each beamlet), i.e., there are m leaf pairs and for each leaf there are $n + 1$ possible positions. For instance, the fluence map of Fig. 4 can be represented by the following matrix of beamlet weights:

$$W = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 6 & 6 & 6 & 3 & 1 & 0 & 0 \\ 0 & 3 & 3 & 6 & 6 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 3 & 3 & 3 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 3 & 3 & 3 & 3 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 3 & 3 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (13)$$

Since the weights of the different beamlets are different, and MLC devices produce the same radiation intensity for all the exposed beamlets (when beam is on), beamlet variation can be achieved by changing the beam aperture and superimposing the correspondent number of segments. The problem of choosing the leaf positions and corresponding apertures to consider for delivery is equivalent to find the decomposition of W into K binary matrices

B^k , called shape matrices, such that

$$W = \sum_{k=1}^K \mu_k B^k, \quad (14)$$

where μ_k is the intensity value (exposure time) for each aperture and B^k are 0–1 matrices where a 0-entry means the radiation is blocked and a 1-entry means that the radiation go through.

The resolution of Eq. 14 would be straightforward if the following constraint were not required: for each row of the binary matrices B^k , if there is more than one nonzero element (more than an exposed beamlet), then the sequence (of 1's) must be consecutive, i.e., zeros are not allowed to break the nonzero sequence (intermittence between exposed beamlets in the same row is impossible).

As examples of possible (feasible) rows we may have

$$0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0,$$

or

$$1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1,$$

where all radiation is blocked and radiation go through all beamlets, respectively. An unfeasible row might be

$$0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0,$$

because left and right leaves block 1st and 9ht beamlet, respectively, but cannot block central beamlets without blocking the ones immediately at left or right.

More than a decomposition can be found for a given fluence matrix. For matrix in (13), a possible decomposition is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 6 & 6 & 6 & 3 & 1 & 0 & 0 \\ 0 & 3 & 3 & 6 & 6 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 3 & 3 & 3 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 3 & 3 & 3 & 3 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 3 & 3 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} +$$

$$2 \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + 3 \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

which is equivalent to the superimposition of the apertures of Fig. 6.

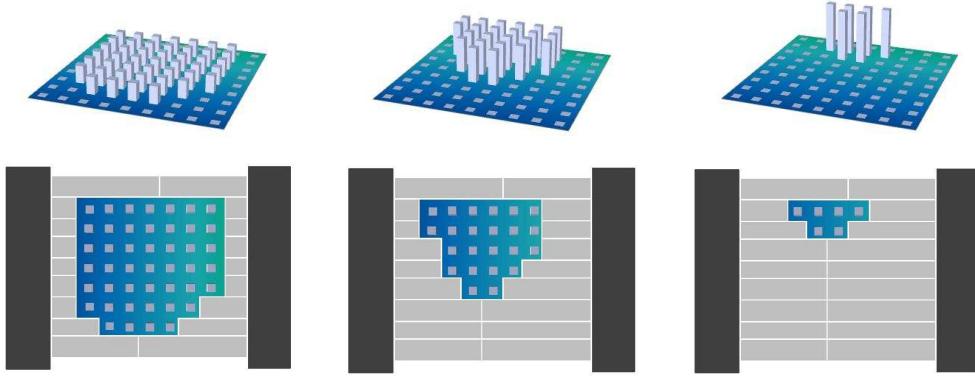


Figure 6: Sequence of apertures and intensities for a decomposition of the fluence map of Fig. 4.

The difficulty of the problem is increased when some objectives are aimed, including:

- minimizing the sum of the matrix weights, $\sum_{k=1}^K \mu_k$, reducing the total beam time;
- minimizing the number of apertures, K , reducing the number of times the beam is turned off.

The decomposition problem with these goals and the referred constraint forms difficult combinatorial optimization problems. The first problem is usually called the decomposition time (DT) problem and the second is called the decomposition cardinality (DC) problem. Obviously, one can also minimize some weighted sum of decomposition time and cardinality, forming objective functions of the form

$$\sum_{k=1}^K \mu_k + \lambda K,$$

where λ is some positive scalar.

The optimal value of the DT problem can be easily obtained while the DC problem is computationally very hard. Typically, the DT optimal solution is found first, and then, heuristically, a decomposition that realizes the optimal DT with a small DC is found.

There are many algorithms to determine the optimal DT, but all fall on the following characterization of the minimal DT [26]. Associate to W its difference matrix D of dimension $m \times (n + 1)$:

$$d_{i,j} = w_{i,j} - w_{i,j-1}, i = 1, \dots, m, j = 1, \dots, n + 1,$$

considering

$$w_{i,0} = w_{i,n+1} = 0, \forall i = 1, \dots, m.$$

The complexity of W , $c(W)$, is defined to be

$$c(W) = \max_{i=1, \dots, m} \{0, c_i(W)\},$$

where $c_i(W)$ is the i^{th} row complexity of W defined by

$$c_i(W) = \sum_{j=1}^n \max\{0, d_{i,j}\}.$$

The following result gives an exact mathematical formula for the minimal DT.

Theorem [26]: The minimal DT for a matrix W equals $c(W)$.

From its proof an algorithm for the construction of an optimal DT can be derived immediately [45].

Algorithm: (DT-optimal leaf sequence)

$k = 0;$

while $W \neq \mathbf{0}$ **do**

$k = k + 1;$

for $i = 1, \dots, m$ **do**

if $w_{i,j} = 0, \forall j = 1, \dots, n$ **then**

$l_i = n;$

$r_i = n + 1;$

else

$l_i = \min\{j \mid 0 \leq j \leq n, w_{i,j} < w_{i,j+1}\};$

$r_i = \min\{j \mid l_i < j \leq n + 1, w_{i,j-1} > w_{i,j}\};$

for $j = 1, \dots, n$ **do**

if $l_i < j < r_i$ **then**

$b_{i,j}^k = 1;$

else

$b_{i,j}^k = 0;$

$W = W - B^k;$

return $B^1, \dots, B^k.$

The choice of l_i and r_i (left and right leaf movements) is not unique. One possible choice, called sweep technique, is to move the leaves always from left to right (see [75], e.g.). Using this algorithm with the sweep technique, the decomposition of W in Eq. 13 is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 6 & 6 & 6 & 3 & 1 & 0 \\ 0 & 3 & 3 & 6 & 6 & 3 & 3 & 1 & 0 \\ 0 & 1 & 3 & 3 & 3 & 3 & 3 & 1 & 0 \\ 0 & 1 & 3 & 3 & 3 & 3 & 1 & 1 & 0 \\ 0 & 1 & 1 & 3 & 3 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} =$$

$$\begin{aligned}
& \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \\
& \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \\
& \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Note that B^4 and B^5 in this decomposition are equal, corresponding to the same aperture. Therefore, the number of segments K for this decomposition of W is 5, more than the number of segments of the first decomposition, $K = 3$. The optimal DT, $\sum_{k=1}^K \mu_k$, is 6, which is also obtained by the first decomposition. No doubt, both decompositions are DT-optimal, but the first decomposition beats the latter in terms of cardinality (DC). Thus, the first decomposition is preferable to the latter.

Typically, after solving the DT, one must solve the DC in order to find the DT-optimal decomposition that behaves better in terms of DC. This problem is not as simple as the DT problem. In fact, the DC problem is strongly NP-hard, even for matrices with a single row [45] (even for a single pair of leaves). Therefore, the most natural way to tackle this problem is to use heuristics. Since one wants to solve the DC problem for the set of decompositions with optimal DT, the DC becomes

$$\begin{aligned}
& \min && K \\
& \text{s.t.} && W = \sum_{k=1}^K \mu_k B^k, \\
& && \sum_{k=1}^K \mu_k \text{ is minimal,} \\
& && \mu_k \in \mathbb{N}.
\end{aligned} \tag{15}$$

A common greedy strategy used to address (15) is to use the following general algorithm [26]:

Algorithm: (general DC algorithm)

while $W \neq \mathbf{0}$ **do**

 find a coefficient $\mu > 0$ and a segment B such that:

$W' = W - \mu B$ is nonnegative;

$c(W') = c(W) - \mu$;

 output (μ, B) ;

$W = W - \mu B$;

Each of the output pairs (μ, B) is called admissible segmentation pair. The success of this algorithm lies on the strategies used to find admissible pairs (μ, B) . Many heuristics can be found in literature (e.g. [26, 75]) but none proved to be the best. Many work can still be done on the development of strategies to address this problem.

Due to MLC physical limitations, a delivery plan for a given optimal fluence map may not be executable. Therefore, some approaches choose to optimize directly apertures and fluences instead of solving the realization problem after the intensity problem. By solving both problems at once, one has the guarantee that delivery is always executable. Methods that integrate intensity optimization and delivery have been recently considered [55, 59]. These methods perform column generation to represent the segments at each iteration.

Due to leaf collision issues, leaf perturbation of adjacent beamlet intensities, tongue-and-groove constraints, etc., the intensity maps actually delivered may be substantially different from the optimized ones. Those problems have been tackled ([37], e.g.) and are still a prosperous field of research.

6 Concluding Remarks

Radiation therapy planning, namely inverse planning, has made an amazing development from the first use of optimization in radiation therapy in 1968 to today's widespread clinical application around the world. Optimization research followed (and contributed to) the evolution of new treatment machines and technology and has made significant contributions to the improvement of radiation therapy planning. However, new (and old) challenges need to be tackled, and this multidisciplinary field continues to require optimization research contributions for further relevant clinical improvements.

One of the main challenges yet to be properly addressed is the choice of the beam angles, as well as the number of angles to consider. In clinical practice, most of the times, beam directions are still manually selected by the treatment planner that relies mostly on his experience. Typically it requires many trial and error iterations between selecting beam angles and computing fluence patterns until a suitable treatment is achieved. This process is tedious, time consuming, has no guarantees of producing good treatments and rely solely on the experience of the treatment planner. Moreover, despite the fact that almost every angle is possible for radiation delivery, except in rare cases, coplanar angles are always considered. This is a way to simplify an already complex problem, and the angles considered lay in the plane of the rotation of the gantry around the patient. Further studies should be pursued both for non-coplanar angles as well as for the case where more than a single isocenter is considered. Further research should also focus on the optimization of the number of beams considered, at least for specific types of cancer, trying to reduce the number of beam angles typically used. Reliable automated process should substitute the planner on this mission and release him to other important tasks of the overall process, namely quality assurance.

Many mathematical optimization models and algorithms have been proposed for the inverse radiation treatment planning, including linear models, integer models, nonlinear models, and multiobjective models. The linear programming approach has the advantage of being fast, but feasible solutions may not be found and producing extreme points as solutions may not be desirable, since prescription limits are often attained. Nonetheless, for beam angle optimization, the most important is computational time, therefore simple LP models are adequate. MIP approach has the advantage of allowing the formulation of

more complex and flexible models. However, an important drawback is that MIP models are solved by branch and bound algorithms with LP. Since the number of integer variables is at least equal to the number of voxels (in the order of the hundred thousands or millions), finding an optimal solution is very difficult at least in a clinical acceptable time. Nonlinear and multiobjective approaches are a fertile field for further research. The quality of the treatment depends on the model and new linear and MIP models can be proposed for specific cases, where a tailored model can improve the general existing ones. However, the quality of the treatment also depends on the algorithm used to solve the proposed model. New algorithms should be explored, namely for nonlinear models, where the gold standard is simulated annealing. Many other derivative free algorithms should be attempted and compared to acknowledge if some naturally design better solutions.

Even using sophisticated inverse planning techniques, radiation therapy planning continues to be, at some stages, a trial and error process. As said above, in clinical practice, most of the times, beam directions are still manually selected by the treatment planner in a trial and error tedious, not accurate, and time consuming process. Unfortunately, for a given set of beam angles, commercial software finds the optimal fluence map, but requires a second human trial and error loop. The objective functions considered depend on weight factors. Planners need to tune the weights or to update the minimum and/or maximum fractions of the prescribed dose that they want to deliver to the PTV and/or OARs to achieve DVH control on the PTV and OARs. The planner should be spared to another time consuming trial and error process, and automated mechanisms should be developed to assist him/her.

Regardless the formulation used, size is always the biggest challenge to overcome. One of the strategies to reduce the computing time of the overall process is sampling. By aggregating voxels in certain structures (mainly in normal tissue and OARs) a decrease in the problem size leads to a faster resolution. Further work should be developed here, since this is one of the best strategies to speed up the overall inverse planning optimization process.

With the existing issues with MLC's, including leaf collision issues, leaf perturbation of adjacent beamlet intensities, tongue-and-groove constraints, etc., the intensity maps actually delivered may be substantially different from the optimized ones. Those problems as

well as the development of new strategies and heuristics to address the realization problem are still a prosperous field of research.

Some unexplored topics for future work can be the target of further research, including the fractionation of radiation treatment planning [12, 17, 30]. Further research should also focus on the development of tailored models for certain types of cancer. The specificities of each type of cancer should be incorporated into general models, by the use of specific constraints, diminishing the computational time needed to calculate solutions by using tailored algorithms.

Inverse planning models and algorithms can also be of use in other related problems (that are out of the scope of this report), like brachytherapy [4, 20, 33, 41, 42, 51] or gamma knife [28, 31, 32, 35]. Much work is still to be done and significant clinical improvements might be accomplished with further optimization research.

References

- [1] R. Acosta, M. Ehrgott, A. Holder, D. Nevin, J. Reese, and B. Salter, Comparing beam selection strategies in radiotherapy treatment design: the influence of dose point resolution, in: C. Alves, P. Pardalos, L. Vicente (Eds.), *Optimization in Medicine, Springer Optimization and Its Applications*, International Center for Mathematics, Springer, 2007.
- [2] American Cancer Society. Cancer facts and figures, 2009. <http://www.cancer.org/downloads/STT/500809web.pdf>.
- [3] M. Alber and F. Nusslin, Optimization of intensity modulated radiotherapy under constraints for static and dynamic MLC delivery, *Phys. Med. Biol.*, Vol. 46, 2001, 3229–3239.
- [4] R. Alterovitz, E. Lessard, J. Pouliot, I.-C. J. Hsu, J. F. O’Brien, and K. Goldberg, Optimizing of HDR brachytherapy dose distributions using linear programming with penalty costs, *Med. Phys.*, Vol. 33, 2006, 4012–4019.

- [5] G. K. Bahr, J. G. Kereiakes, H. Horwitz, R. Finney, J. Galvin, and K. Goode, The method of linear programming applied to radiation treatment planning, *Radiology*, Vol. 91, 1968, 686–693.
- [6] G. Bednarz, D. Michalski, C. Houser, M.S. Huq, Y. Xiao, P. R. Anne, and J. M. Galvin, The use of mixed-integer programming for inverse treatment planning with pre-defined segments, *Phys. Med. Biol.*, Vol. 47, 2002, 2235–2245.
- [7] S. C. Billups and J. M. Kennedy, Minimum-support solutions for radiotherapy planning. *Ann. of Oper. Res.*, Vol. 119, 2003, 229–245.
- [8] V. Bevilacqua, G. Mastronardi, and G. Piscopo, Evolutionary approach to inverse planning in coplanar radiotherapy, *Image and Vision Computing*, Vol. 25, 2007, 196–203.
- [9] T. Borffeld, IMRT: a review and preview, *Phys. Med. Biol.*, Vol. 51, 2006, 363–379.
- [10] W. D. Carlson and D. Morkovin, Time-dose relations for optimal radiation treatment of malignancy with minimal damage to normal tissues, *Miss. Valley Med. J.*, Vol. 81, 1959, 287–288.
- [11] Y. Censor, Mathematical optimization for the inverse problem of intensity-modulated radiation therapy, Y. Censor, in: J.R. Palta and T.R. Mackie (Eds.), *Intensity-Modulated Radiation Therapy: The State of The Art*, American Association of Physicists in Medicine (AAPM), Medical Physics Monograph No. 29, Medical Physics Publishing, Madison, Wisconsin, USA, 2003, 25–49.
- [12] M. Chu, Y. Zinchenko, S. G. Henderson, and M. B. Sharpe, Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty, *Phys. Med. Biol.*, Vol. 50, 2005, 5463–5477.
- [13] C. Cotrutz and L. Xing, Segmented-based optimization using a genetic algorithm, *Phys. Med. Biol.*, Vol. 48, 2003, 2987–2998.
- [14] C. Cotrutz, M. Lahanas, et al., A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy, *Phys. Med. Biol.*, Vol. 46, 2001, 2161–2175.

- [15] CPLEX, ILOG CPLEX, <http://www.ilog.com/products/cplex>.
- [16] J. Deasy, E. K. Lee, T. Bortfeld, M. Langer, K. Zakarian, J. Alaly, Y. Zhang, H. Liu, R. Mohan, R. Ahuja, A. Pollack, J. Purdy, and R. Rardin, A collaboratory for radiation therapy planning optimization research, *Ann. Oper. Res.*, Vol. 148, 2006, 55–63.
- [17] G. Deng and M. C. Ferris, Neuro-dynamic programming for fractionated radiotherapy planning, in: C. Alves, P. Pardalos, L. Vicente (Eds.), *Optimization in Medicine, Springer Optimization and Its Applications*, International Center for Mathematics, Springer, 2007.
- [18] D. Djajaputra, Q. Wu, Y. Wu, and R. Mohan, Algorithm and performance of a clinical IMRT beam-angle optimization system, *Phys. Med. Biol.*, Vol. 48, 2003, 3191–3212.
- [19] W. D. D’Souza, R. R. Meyer, and L. Shi, Selection of beam orientations in intensity-modulated radiation therapy using single beam indices and integer programming, *Phys. Med. Biol.*, Vol. 49, 2004, 3465–3481.
- [20] W. D. D’Souza, R. R. Meyer, B. R. Thomadsen, and M. C. Ferris, An iterative sequential mixed-integer approach to automated prostate brachytherapy treatment plan optimization, *Phys. Med. Biol.*, Vol. 46, 2001, 297–322.
- [21] M. Ehrgott and R. Johnston, Optimisation of beam directions in intensity modulated radiation therapy planning, *OR Spectrum*, Vol. 25, 2003, 251–264.
- [22] M. Ehrgott, A. Holder, and J. Reese, Beam selection in radiotherapy design, *Linear Algebra and its Applications*, Vol. 428, 2008, 1272–1312.
- [23] M. Ehrgott, C. Guler, H. W. Hammacher, and L. Shao, Mathematical optimization in intensity modulated radiation therapy, *4OR*, v6, 2008.
- [24] M. Ehrgott and M. Burjony, Radiation therapy planning by multicriteria optimisation, *Proceedings of the 36th Annual Conference of the Operational Research Society of New Zealand*, 2001.
- [25] M. Ehrgott and I. Winz, Interactive decision support in radiation therapy treatment planning, *OR Spectrum*, Vol. 30, 2008, 311–329.

- [26] K. Engel, A new algorithm for optimal multileaf collimator field segmentation, *Discrete Appl. Math.*, Vol. 152, 2005, 35–51.
- [27] M. C. Ferris, J. -H. Lim, and D. M. Shepard, Optimization approaches for treatment planning on a Gamma Knife, *SIAM J. Optim.*, Vol. 13, 2003, 921–937.
- [28] M. C. Ferris, J. -H. Lim, and D. M. Shepard, Radiosurgery treatment planning via nonlinear programming, *Ann. of Oper. Res.*, Vol. 119, 2003, 247–260.
- [29] M. C. Ferris, R. Einarsson, Z. Jiang, and D. M. Shepard, Sampling issues for optimization in radiotherapy, *Ann. of Oper. Res.*, Vol. 148, 2006, 95–115.
- [30] M. C. Ferris and M. M. Voekler, Fractionation in radiation treatment planning, *Math. Program.*, Vol. 101, 2004, 387–413.
- [31] M. C. Ferris, J. Lim, and D. M. Shepard, An optimization approach for radiosurgery treatment planning, *SIAM J. Optim.*, Vol. 13, 2001, 921–937.
- [32] M. C. Ferris, R. R. Meyer, and W. D’Souza, Radiation treatment planning: mixed integer programming formulations and approaches, in: G. Appa, L. Pitsoulis, and H. P. Williams (Eds.), *Handbook on Modelling for Discrete Optimization*, Springer, 2006.
- [33] R. J. Gallagher and E. K. Lee, Mixed integer programming optimization models for brachytherapy treatment planning, *Proc AMIA Annu Fall Symp*, 1997, 278-282.
- [34] H. W. Hamacher and K. H. Kufer, Inverse radiation therapy planning – a multiple objective optimization approach, *Discrete Appl. Math.*, Vol. 118, 2002, 145–161.
- [35] A. Holder and B. Salter, A tutorial on radiation oncology and optimization, in: H. Greenber (Eds.), *Emerging Methodologies and Applications in Operations Research*, Kluwer Academic Press, Boston, USA, 2004.
- [36] A. Holder, Designing radiotherapy plans with elastic constraints and interior point methods, *Health Care Management Science*, Vol. 6, 2003, 5–16.
- [37] T. Kalinowski, A duality based algorithm for multileaf collimator field segmentation with interleaf collision constraint, *Discrete Appl. Math.*, Vol. 152, 2005, 52–88.

- [38] K. H. Kufer, A. Scherrer, et al., Intensity-Modulated radiotherapy - a large scale multi-criteria programming problem, *OR Spectrum*, Vol. 25, 2003, 223–249.
- [39] E. K. Lee, T. Fox, and I. Crocker, Simultaneous beam geometry and intensity map optimization in intensity-modulated radiation therapy, *Int. J. Radiat. Oncol. Biol. Phys.*, Vol. 64, 2006, 301–320.
- [40] E. K. Lee, T. Fox, and I. Crocker, Integer programming applied to intensity-modulated radiation therapy treatment planning, *Ann. Oper. Res.*, Vol. 119, 2003, 165–181.
- [41] E. K. Lee, R. J. Gallagher, D. Silvern, C. S. Wuu, and M. Zaider, Treatment planning for brachytherapy: an integer programming model, two computational approaches and experiments with permanent prostate implant planning, *Phys. Med. Biol.*, Vol. 44, 199, 145–165.
- [42] E. K. Lee and M. Zaider, Mixed integer programming approaches to treatment planning for brachytherapy-application to permanent prostate implants, *Ann. Oper. Res.*, Vol. 119, 2003, 147–163.
- [43] Y. Li, J. Yao, D. Yao, and W. Chen, A particle swarm optimization algorithm for beam angle selection in intensity-modulated radiotherapy planning, *Phys. Med. Biol.*, Vol. 50, 2005, 3491–3514.
- [44] G. J. Lim, M.C. Ferris, S. J. Wright, D. M. Shepard, and M. A. Earl, An optimization framework for conformal radiation treatment planning, *INFORMS Journal on Computing*, Vol. 19, 2007, 366–380.
- [45] G. J. Lim and E. K. Lee, *Optimization in Medicine and Biology*. Taylor and Francis, Auerbach Publications, USA, 2007.
- [46] G. J. Lim, J. Choi, and R. Mohan, Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning, *OR Spectrum*, Vol. 30, 2008, 289–309.
- [47] W. A. Lodwick, S. McCourt, F. Newman and S. Humphries, Optimization Methods for Radiation Therapy Plans, in: C. Brgers and F. Natterer (Eds.), *Computational, Radiology and Imaging: Therapy and Diagnostics*, Springer-Verlag, 1998.

- [48] MATLAB, The MathWorks Inc., <http://www.mathworks.com>.
- [49] G. Meedt, M. Alber, and F. Nusslin, Non-coplanar beam direction optimization for intensity-modulated raditherapy, *Phys. Med. Biol.*, Vol. 48, 2003, 2999–3019.
- [50] C. Men, H. E. Romeijn, Z. C. Taskin, and J. F. Dempsey, An exact approach to direct aperture optimization in IMRT treatment planning, *Phys. Med. Biol.*, Vol. 52, 2007, 7333–7352.
- [51] R. R. Meyer, W. D. D’Souza, M. C. Ferris, and B. R. Thomadsen, MIP models and BB strategies in brachytherapy treatment optimization, *Journal of Global Optimization*, Vol. 25, 2003, 23–42.
- [52] S. M. Morrill, R. G. Lane, G. Jacobson, and I. Rosen, Treatment planning optimization using constrained simulated annealing, *Phys. Med. Biol.*, Vol. 36, 1991, 1341–1361.
- [53] A. Olafsson and S. J. Wright, Linear programming formulations and algorithms for radiotherapy treatment planning, *Optimization Methods and Software*, Vol. 21, 2006, 201–231.
- [54] A. Pugachev and L. Xing, Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy, *Phys. Med. Biol.*, Vol. 46, 2001, 2467–2476.
- [55] F. Preciado-Walters, M. P. Langer, R. L. Rardin, and V. Thai, Column generation for IMRT cancer therapy optimization with implementable segments, *Ann. Oper. Res.*, Vol. 148, 2006, 65–79.
- [56] F. Preciado-Walters, R. Rardin, M. Langer, and V. Thai, A coupled column generation, mixed-integer approach to optimal planning of intensity modulated radiation therapy for cancer, *Mathematical Programming*, Vol. 101, 2004, 319–338.
- [57] RadiologyInfo. <http://www.radiologyinfo.org/en/info.cfm?pg=linac>.
- [58] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, and A. Kumar, A new linear programming approach to radiation therapy planning problems, *Operations Research*, Vol. 54, 2006, 201–216.

- [59] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, and A. Kumar, A column generation approach to radiation therapy treatment planning using aperture modulation, *SIAM J. Optim.*, Vol. 15, 2005, 838–862.
- [60] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, A. Kumar, and J. Li, A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planing, *Phys. Med. Biol.*, Vol. 48, 2003, 3521–3542.
- [61] H. E. Romeijn, J. F. Dempsey, and J. Li, A unifying framework for multi-criteria fluence map optimization models, *Phys. Med. Biol.*, Vol. 49, 2004, 1991–2013.
- [62] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, Design and Analysis of Computer Experiments, *Statistical Science*, Vol. 4, 1989, 409–423.
- [63] A. Scherrer and K. H. Kufer, Accelerated IMRT plan optimization using the adaptive clustering method, *Linear Algebra and Its Applications*, Vol. 428, 2008, 1250–1271.
- [64] E. Schreibmann, M. Lahanas, et al., Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy, *Phys. Med. Biol.*, Vol. 49, 2004, 747–770.
- [65] D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu, Direct aperture optimization: a turnkey solution for the step-and-shoot IMRT, *Med. Phys.*, Vol. 29, 2002, 1007–1018.
- [66] D. M. Shepard, M. C. Ferris, G. H. Olivera, and T. R. Mackie, Optimizing the Delivery of Radiation Therapy to Cancer Patients, *SIAM Review*, Vol. 41, 1999, 721–744.
- [67] S. Spirou and C. -S. Chui, A gradient inverse planning algorithm with dose-volume constraints, *Med. Phys.*, Vol. 25, 1998, 321–333.
- [68] P. Suss and K. H. Kufer, Balancing control and simplicity: A variable aggregation method in intensity modulated radiation therapy planning, *Linear Algebra and Its Applications*, Vol. 428, 2008, 1388–1405.
- [69] J. Tervo, P. Kolmonen, T. Lyyra-Laitinen, J. D. Pintr, and T. Lahtinen, An optimization-based approach to the multiple static delivery technique in radiation therapy, *Ann. of Oper. Res.*, Vol. 119, 2003, 205–227.

- [70] C. Thieke, K. H. Kufer, et al., A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation, *Radiotherapy and Oncology*, Vol. 85, 2007, 292–298.
- [71] S. Webb, *The physics of three-dimensional radiation therapy - Conformal radiotherapy, radiosurgery and treatment planning*, Institute of Physics Publishing, Bristol and Philadelphia, 1993.
- [72] I. Winz, *A decision support system for radiation therapy treatment planning*, Department of Engineering Science, University of Auckland, MsC., 2004.
- [73] X. Wu, Y. Zhu, J. Dai, and Z. Wang, Selection and determination of beam weights based on genetic algorithms for conformal radiotherapy treatment planning, *Phys. Med. Biol.*, Vol. 45, 2000, 2547–2558.
- [74] X. Wu, Y. Zhu, and L. Luo, Linear programming based on neural networks for radiotherapy treatment planning, *Phys. Med. Biol.*, Vol. 45, 2000, 719–728.
- [75] P. Xia, L. Verhey, Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple segments, *Med. Phys.*, Vol. 25, 1999, 1424–1434.