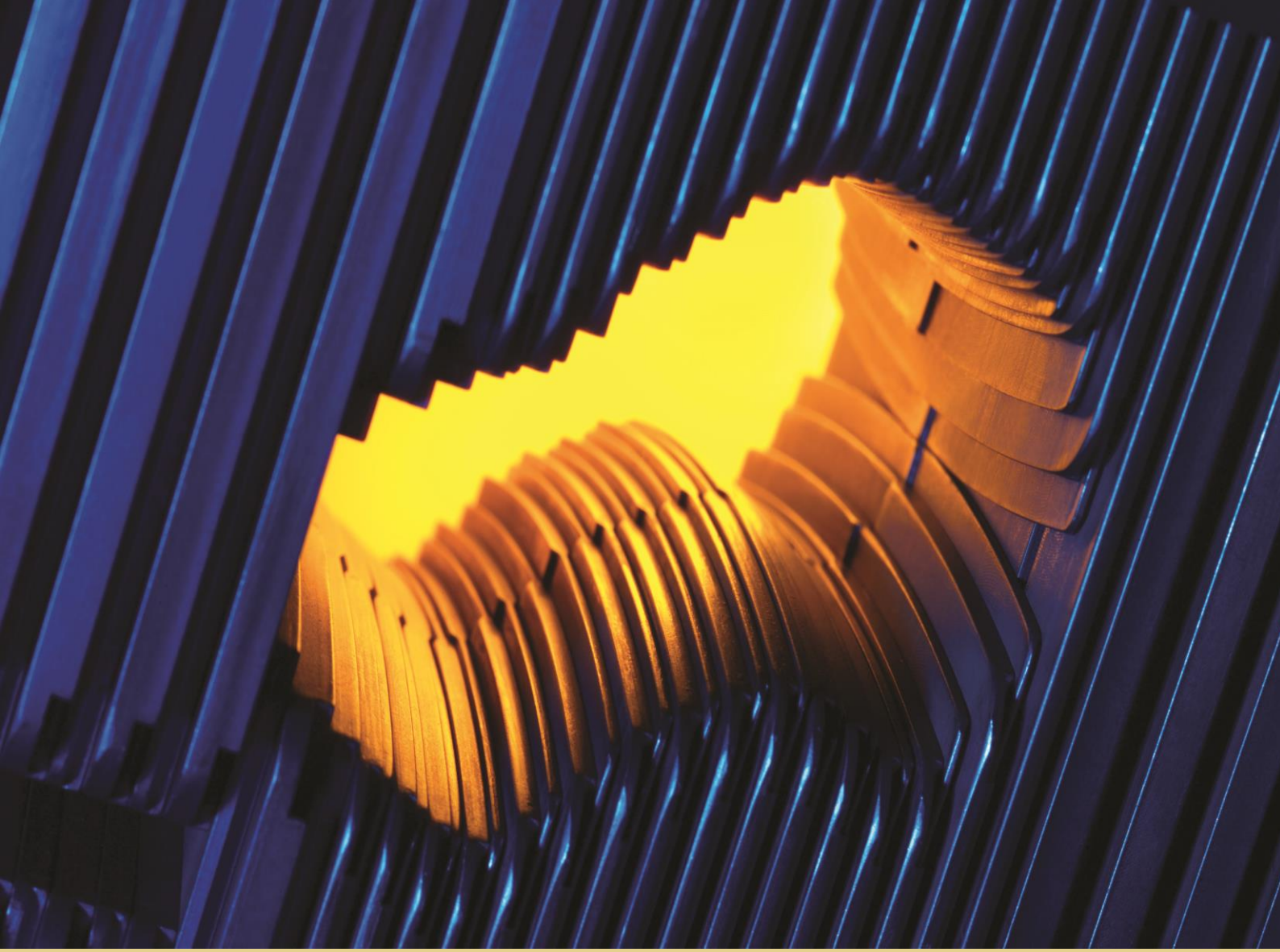


Rui Guilherme da Silva Valente

Computational methods for beam angle optimization in intensity modulated radiotherapy

Master Dissertation in Mathematics, Area of Specialization in Applied Analysis and Computation, supervised by Professor Marta Margarida Braz Pascoal and presented to the Department of Mathematics of the Faculty of Sciences and Technology of the University of Coimbra.

July 2018

· U C ·

Universidade de Coimbra

Computational methods for beam angle optimization in intensity modulated radiotherapy

Rui Guilherme da Silva Valente

UNIVERSIDADE DE COIMBRA

Master in Mathematics

Mestrado em Matemática

MSc Dissertation | Dissertação de Mestrado

July 2018

Acknowledgements

For the conclusion of this master's thesis, the support of several people was essential, to whom I now turn to express my gratitude.

Firstly, I would like to express my sincere acknowledgements to my advisor Professor Marta Pascoal for the continuous support of my master's studies and related research, for her patience and knowledge.

Furthermore, I would like to thank Dr. Paulo César Simões for introducing me to the Oncology Unit of *Centro Hospitalar e Universitário de Coimbra*, as well as for all the clarifications.

I must express my very profound gratitude to my family, in particular my parents, for providing me with unfailing support and continuous encouragement throughout my academic studies.

I am profoundly grateful to Rita for all her patience, unconditional support and sweetness.

I also take this opportunity to thank my classmates Tiago, Carolina, Andreia and Daniel, who have always given me invaluable help whenever necessary.

Lastly, my special thanks are extended to FAN-Farra Académica de Coimbra, my second home, which makes the end of this journey even more difficult.

Abstract

To plan the treatment of tumor cells in the human body can be a very complex task for radiation oncology. Radiation therapy applies ionizing radiation to the patient's body with the aim of damaging the DNA of cancerous tissue leading to cellular death. However, the incidence of radiation could be fatal to the vital organs surrounding the cancerous tissues, leading to their deterioration. In this thesis, a formulation proposed by Bertsimas *et al* in [3] is considered, which aims at finding a set of angles and the beamlet intensities while minimizing the effect of incident radiation on vital bodies and maximizing the radiation absorbed by the malignant cells representing the biological consequences of radiation. In order to circumvent the nonlinearity of the formulation, heuristic methods are described and applied to one phantom case, composed by tumor cells and one surrounding organ, both discretized in voxels. The set of beamlets per beam are defined by bixels and the combination of one, three, five and seven angles are applied. Experimental results compare different variants of the methods, in order to assess them in terms of the objective function values and the run times.[1]

[1]The copyright of the cover image belongs to: Maxom Precision Motors, "Intensity-Modulated Radiosurgery System", January 31, 2012. Accessed on May 23, 2018. https://maxonmotorusa.wordpress.com/category/fda/

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The present work deals with an optimization problem related to the treatment of tumor forms, known as intensity modulated radiotherapy. This treatment method is based on the incidence of X-ray radiation, which aims to destroy malignant cells, whilst preserving, as far as possible, the surrounding vital organs. This text reviews this type of treatment and its planning with the goal of making its application effective.

1.1 Intensity modulated radiation therapy

Radiotherapy or radiation therapy is a medical treatment of cancerous cells on the human tissue. This treatment consists on using ionizing radiation against the tumor, which is absorbed and damages the deoxyribonucleic acid (DNA) of the cancerous cells, destroying them. The intensity required to disable the reproduction of the tumor cells is smaller than for non-cancerous cells. Moreover, these non-cancerous cells have the merit of reproducing themselves even with damaged DNA, unlike the cancerous-cells. This turns into an advantage for the radiotherapy treatment, but it is not enough. Although tumor cells are more sensitive to radiation, most of the times, their eradication involves the destruction of non-cancerous cells. In radiation therapy it is almost impractical to deliver zero dose to the normal tissue (all the body cells which do not belong to any particular tumor structure) adjoining the target volumes (i.e., the tumor). As such, the goal is to deliver a reasonable cumulative dose to the tumor body while minimizing the dose delivered to normal tissue [8, 20]. This is one of the biggest difficulties of this process, since there are critical cells that cannot be destroyed in order to certify the sustention of the corresponding organs at risk (OARs).

In the last decades, several computational approaches have been developed in order to improve the accuracy and effectiveness of radiation therapy. Distinct scanning techniques have emerged, such as computed tomography (CT), magnetic resonance imaging (MRI), 3D treatment planning systems, among others, which meant a huge step in this area. These techniques were crucial not only to diagnose the tumors, but also to characterize their composition and to image them in a better way. These advances were followed by other improvements that could overcome early radiography equipments. The radiotherapy used to be managed manually and intuitively by a physician, which incurs, evidently, in a large margin of possible errors. The development of the software and hardware in medical technology (3D planning, velocity of calculations, delivery of radiation) allowed this

process to be more accurate in the determination of the most appropriate patient position, the number and distribution of beam angles, the type of energy and intensity of radiation and dose disposal through the tumor. Additionally, it also improved the outline of tumors' shape (Clinical Target Volume) and the OARs. This technological development eliminated the hand planning methodology, but not the trial and error approach, in which the parameters are fixed and the dose distribution is made through several attempts on the patient, until a reasonable outcome is found.

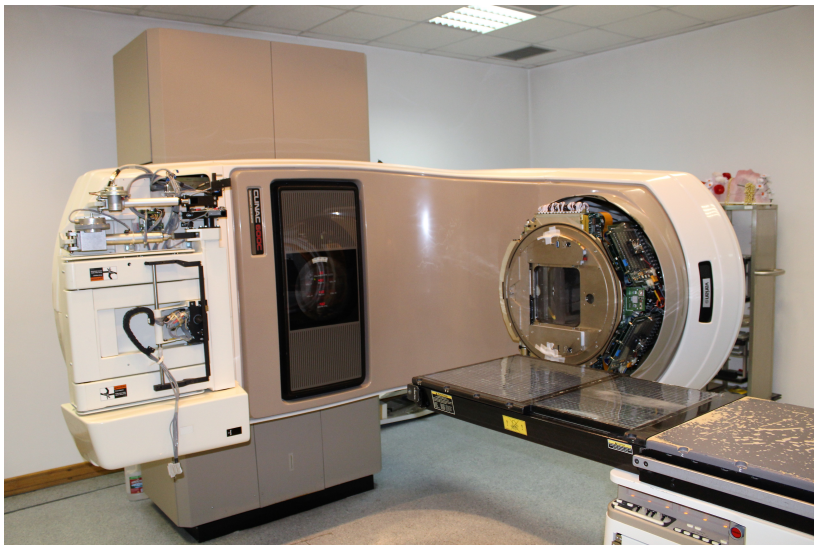


Fig. 1.1 Linear accelerator at the Oncology Unit of *Centro Hospitalar e Universitário de Coimbra*

An innovative radiotherapy technology emerged and it lasts until today: the multileaf collimator (MLC). The MLC is a sophisticated electrical/mechanical device that provides shaped, independent and modulated beams' radiation, using individual leaves, as depicted in Figures 1.2 and 1.1. These multiple leaves are metal barriers that form a kind of board, situated right after the radiation source and such that they can move along the radiation path to control and modulate the passage of a particle beam. This mechanism is discretized into *bixels*, or *beamlets*, which subdivide the fluency distribution of a certain beam in n rectangular parts, all having the same direction but of which some could be blocked, thanks to the leaves of the MLC that slide in the beam path while the radiotherapy is being performed. Such machine holds and rotates on a gantry 360º around the patient, who is lying on the stretcher, with total independence, which allows for a more appropriate approach towards the treatment [25]. When this is the only movement in the treatment session, the treatment is said to be *coplanar*. However, other forms of movement are also allowed, which leads to a *non-coplanar* treatment. The first one will be considered in the rest of this work.

This arrangement allowed the creation of an intensity modulated radiotherapy (IMRT): a state-of-the-art treatment method, with a more efficient and precise behavior against the cancerous cells, laying aside OARs and other normal tissues. This new technology improved several points from traditional radiotherapy. Since the conventional treatment has been limited to the use of a very small number of inadequate shaped beams, by now, the number of geometrical positions could have been increased in order to effectively cover the tumor volume.

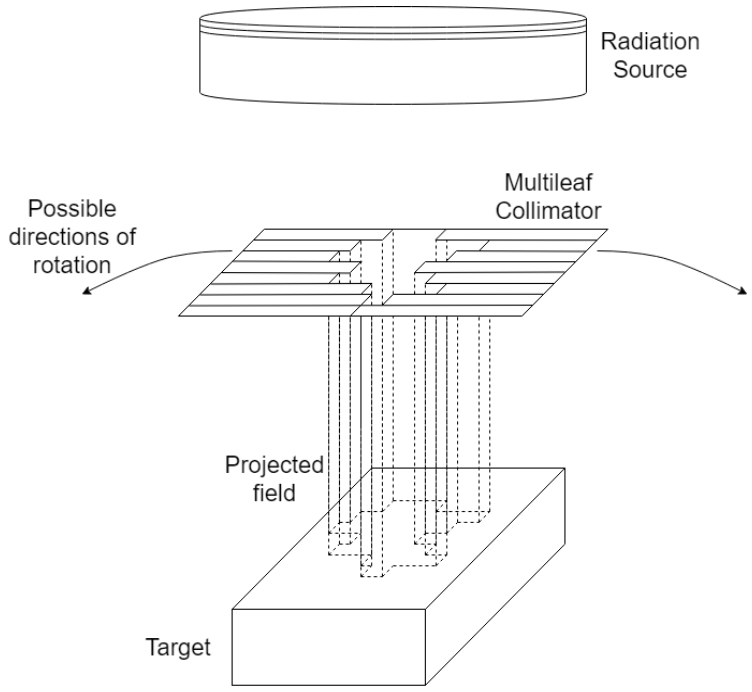


Fig. 1.2 Simplified representation of a multileaf collimator mechanism

For this therapy, a discretization of the patient's tumor and OARs in 3D hyper-rectangles, called *voxels*, will be considered. As such, every organ near the tumor structure will be represented as a set of voxels and is susceptible to the beamlets during the treatment. Each voxel is associated with a certain radiation dose (RD) value. This RD has an important role in this medical area and can be defined as the quantity of energy absorbed by the voxel, by unit of mass. The RD is calculated taking into account the type and stage of the tumor being treated and the condition and biological response of the patient.

We shall now consider n possible beams for a treatment and assume that each beam is segmented in p bixels. Moreover, m voxels are considered in this treatment, irrespective of the structure they compose. Furthermore, $a_{ij}(\theta_k)$ denotes the ratio of intensity dose delivered to the voxel i by the beamlet j, from the beam k, for any $i = 1,\ldots,m$, $j = 1,\ldots,p$, $k = 1,\ldots,n$. Given the properties of the patient's anatomy, the beamlet source (modality, energy) and orientation k, $a_{ij}(\theta_k)$ is calculated by mathematical models of the physical behavior of radiation as it travels through the body [8]. Thus, $a_{ij}(\theta_k)$ can be computed by projection models that mimic radiation's orientation and behavior through the patient's body. Supposing that the matrix $A \in [0,1]^{m\times np}$, called the dose deposition matrix, is known, then we need to establish a vector $x \in \mathbb{R}^{np}_{\geq 0}$, representing the beams' different intensities, where x_j stands for the intensity of the bixel j, for $j = 1,\ldots,p$, positioned in angle k, $k = 1,\ldots,n$, such that

$$d = Ax \tag{1.1}$$

where $d \in \mathbb{R}^m$ is the dose vector and its element d_i represents the total dose deposited to the voxel i (of a certain organism's structure) after all process, for any $i = 1,\ldots,m$.

The planning process for this therapy is divided into three phases:

1. Geometry or beam angle optimization problem: which aims at finding the optimal set of beam combinations, in terms of its number and directions, with the goal of providing the best dose distribution to each patient.
2. Intensity or fluency map optimization problem: which has the goal of calculating the intensity profile of the beams selected in the first step.
3. Realization problem: which marshals the delivery sequence of the beams in a sensitive way, in order to find a consistent order to deliver the radiotherapy treatment planned in the previous points.

These stages will be further explained below, with special focus on the beam angle optimization problem and on the intensity problem.

The beam angle optimization problem Selecting the best possible set of beams is not a very intuitive process — therefore, the beam angle optimization problem has an important role on the radiotherapy process. Over the last years, optimization algorithms have come to find a suitable approach for this problem. The goal of using them is mainly to avoid the manual selection of beam angles made by physicians, leaving the calculations for a computer to do. Nonetheless, the large size of the search space involved in the beam-angle search brings some hurdle to this problem. Assuming that, e.g. a pursuit only considers integer angles, 360 possible angles can be used in the IMRT, which means that for n sets of beams there are around C_n^{360} possibilities and, therefore, exhaustive searches may be insufficiently broad for most clinical cases.

Moreover, the treatment should be as fast and brief as possible since the patient's movements can put at risk its effectiveness and performance. Changing beam directions or using a great number of directions is not advisable because of the time it requires. The IMRT, also known as 'conformal therapy', has enlarged the number of possible potential beams, although it has been proven that, in some simple cases, the number of beams should not exceed 5, whilst a more complicated clinical case, such as situations where OARs are more vulnerable or neighboring with a complex tumor, can require up to 9 beams or more [12]. Due to this, physicians sometimes prefer to spread short sessions throughout the day, to make sure that the treatment needed is provided.

The intensity problem To complement the beam angle optimization, we optimize the intensity vector $x = [x_j]_{j=1,\ldots,np}$, for each point j on the MLC aperture, for a certain angle (or direction) proposed to the radiation process. The dose is calculated discretely in all points defined in the critical area, called *dose points* (DP), that represent the RD attributed to every voxel that is affected by the radiation.

The correct process of calculating the vector x, denominated as *fluency map optimization* (FMO), should be followed by the conversion to MLC leaf sequences, in order to be able to proceed the treatment correctly [25]. The regions that might be affected by the radiation treatment must also be studied. The various tissues toleration (target structures and any nearby critical structures) should be taken into account since each tissue has a different toleration to the absorption of radiation. As such, during the FMO, the radiation oncologist should calculate a prescription dose for the tumor, taking into account the bounds of the OARs and nearby normal tissue.

Essentially, the main goal of this problem is to fit the optimal values for the intensity vector. On the one hand, the intensity cannot be exaggerated since that could irreversibly affecting some OARs, but, on the other, this intensity has to be powerful enough to achieve the results needed for fighting the tumor. These requirements seem to be an obstacle to most of the intensity problem because of the substantial amount of time needed to process and optimize the intensity of a single set of predefined beams computationally. Depending on the complexity of the tumor and the specifications of the organ, the required time can vary between several minutes and a couple of hours. This problem is particularly significant for the most difficult cases where the tumor is wrapped around a healthy organ or the tumor has a non-convex shape.

Along the years, several models and algorithms were studied and reviewed in order to optimize this problem. These mathematical models can be categorized as: feasibility problems, linear programming models, nonlinear programming models, mixed integer programming models and multiple objective programming models [8].

The realization problem The realization problem is the last problem of the IMRT optimization model. Assuming the two previous subproblems have been solved, the last goal is to find an optimal solution to implement the given intensity matrix I, obtained by the reshaping of vector x. This matrix represents the intensity of the radiation through the MLC and is calculated as the overlap of the different leaves that irradiate the target for a fixed amount of time. The MLC can behave in three different ways [8]:

1. Dynamic mode: the opposing leaves move continuously and independently in order to generate the desired intensity proposed by the matrix I.

2. Static mode (step-and-shoot method): the desired intensity matrix is divided into a set of binary matrices. These matrices reproduce a set of subfields that are designed by the MLC and delivered sequentially for the required amount of time. The radiation is suspended while the leaves spawn the next subfield.

3. Arc therapy: in this mode the MLC moves dynamically to shape each subfield while the gantry is rotating and the radiation is on [24].

Given this, we can conclude that the realization problem has been in continuous development until today and no perfect method has been found, since the number of specifications is uncommonly vast. As such, this last problem can also be considered as a matrix decomposition optimization problem, where the best solution could be possibly unreachable given its complexity. Nevertheless, the successive studies have brought new approaches and developed the existing ones in order to improve and reach a higher success rate for the radiation therapy.

1.2 Organization

After this description of the mechanical process behind the intensity modulated radiation therapy and the usual organization of such treatments, this dissertation is composed of five other chapters. The second chapter presents the formulation in the form of a nonlinear optimization problem that includes

the constraints resulting from the effects of radiation on the various types of organs. In the third chapter, the behavior of radiation when going through the human body is analyzed – more specifically, its attenuation – in order to incorporate the effect on the formulation previously described. In the fourth chapter, several heuristic methods are discussed. The fifth chapter presents computational results of the heuristic methods, when tested for an artificial example. Moreover, these results are compared in terms of the solution obtained and the time of execution. Finally, the last chapter presents the concluding remarks of this dissertation.

Chapter 2

The beam angle optimization problem

Two decisions are of fundamental importance to project the radiation medical treatment: the selection of the beam angles and the computation of the beams' intensity 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 are optimized by using an automated software tool. However, in this chapter, a nonlinear formulation will be presented that addresses both angle selection and the corresponding intensity computation simultaneously. This formulation, last presented by Bertsimas *et al.* [3], will be crucial to the implementation of methods to solve the IMRT studied later.

2.1 Literature review

The beam angle optimization (BAO) problem has been extensively studied for a large number of years, and different approaches and models have been proposed. Mathematical models were developed, based on medical treatment requirements, in order to replace the trial and error approach used traditionally by technicians. This section will briefly overview the most relevant approaches to this topic.

The first linear optimization model for the BAO was proposed in 1968 by Bahr *et al.* [2]. In this first article, the authors propose to remove the "trial" procedure from spatial treatment planning as much as possible. In principle, the therapist needs only to provide contour exhibiting areas of anatomical interest with his requirements of dosages (regarding minimum and maximum doses) in the areas to be treated. In the first formulation, the treatment area is partitioned into a target region T, consisting of m_T voxels, and a normal region N, consisting of m_N voxels. The dose delivered to the tumor structure is constrained to lie between a lower bound vector, TLB, and an upper bound vector, TUB, while the dosage delivered to the normal region is bounded above by NUB. This formulation proposed to minimize a weighted sum of the doses delivered to the normal voxels, where the weights are components of a cost vector w. Considering these structures, the dose deposition matrix A could be divided into two submatrices: A_T and A_N, which are the dose matrices for the target and normal regions, respectively. Defining the decision variables as the vector of beam weights, denoted by x, the vector of doses to the target voxels can be calculated as $d_T = A_T x$, and the vector of doses to the

normal voxels, as $d_N = A_N x$. Then, the model can be expressed as follows.

$$\begin{array}{lrcl} \min & w_N^{\mathsf{T}} d_N & & \\ \text{subject to} & d_N & \leq & NUB \\ & TLB & \leq & d_T \leq \quad TUB \\ & x & \geq & 0 \end{array} \tag{2.1}$$

Since then, other researchers have used linear models to formulate the BAO problem. For instance, Langer and Joseph [11] describes a method for formulating and solving this optimization problem as a linear combinatorial program, yielding a rigorous and efficient determination of the beam weights for the case of a thoracic plan. In 1991 Morril *et al.* [15] presented a method of incorporating dose-volume considerations within the framework of conventional linear programming; and Rosen *et al.* [16] designed a linear objective function, in such a way that it defines the goal of optimization, including dose constraint points, target volume and normal structures with minimum and maximum dose values assigned to each point. A more detailed survey on linear models for the BAO problem can be found in [17].

Besides the linear optimization models, some nonlinear approaches are commonly applied as well. At present, a nonlinear programming model, the weighted least squares model, is one of the most well-known formulations among researchers. In 1989, Webb [20] applied a method that begins with the treatment dose prescription and calculates the beam profiles rather than the reverse 'conventional' planning technique. In 1990 Bortfeld *et al.* [4] analyzed the problem of reconstructing a 3D image from its 2D projections and compared this to the problem of optimizing the dose distribution for conformational radiotherapy with intensity modulated external beams. Five years later, Xing and Chen [22] discussed the relationship between iterative algorithms used in imaging science and several existing inverse treatment planning methods in radiotherapy. This model calculates the weighted sum of average squared deviations from the prescribed dose for each organ. Its general model is presented as:

$$\min \frac{w_T}{m_T} \parallel d_T - TG \parallel_2^2 + \frac{w_C}{m_C} \parallel d_C - CUB \parallel_2^2 + \frac{w_N}{m_N} \parallel d_N - NUB \parallel_2^2 . \tag{2.2}$$

In this formulation, besides the target and normal regions, a critical region C, composed by m_C voxels, is considered. Furthermore, weight factors w_T and w_C are attributed to target and critical organs respectively. The factor TG represents the desired dose to tumor voxels, while CUB represents the upper bound to critical voxels. As such, $d_C \leq CUB$ should be added to the constraints of this problem, assuming d_C as the vector of doses to the critical region C, calculated from the submatrix A_C of A, related to the voxels in region C. The remaining constraints coincide with the constraints of problem (2.1) – therefore, the whole model is formulated as:

$$\begin{array}{ll} \min & \frac{w_T}{m_T} \parallel d_T - TG \parallel_2^2 + \frac{w_C}{m_C} \parallel d_C - CUB \parallel_2^2 + \frac{w_N}{m_N} \parallel d_N - NUB \parallel_2^2 \\ \text{subject to} & d_C \leq CUB \\ & d_N \leq NUB \\ & TLB \leq d_T \leq TUB \\ & x \geq 0 \end{array}$$

Apart from the research on nonlinear models, others works introduced and formulated heuristic methods to optimize the orientation and the intensity of the beams. These methods automatically select the beam angles, turning the nonlinear optimization problem into a linear problem, solved computationally. For every solution considered, the heuristic method is performed, and the corresponding objective function value is used as a measure to find the global minimum. In 1993, Mageras and Mohan [13] applied simulated annealing to optimize radiation treatment plans, in which a set of beam weights are iteratively adjusted so as to minimize a cost function. More recently, Bertsimas *et al.* [3] introduced a rather complete formulation for the problem and proposed a hybrid heuristic method which combines a simulated annealing procedure with the knowledge of the gradient descendent. The idea behind the method is to take advantage of the best features of simulated annealing, which allows to search for global minima, and the gradient information, used to quickly find a local minimal. In the following section this formulation will be analized with more details. In 2015, Dias *et al.* [7] have proposed a convex penalty function formulation, where each voxel is penalized considering the square difference of the amount of dose received by the voxel and the amount of dose desired/allowed for the voxel. In this work, several heuristic and local search methods were tested for different instances, revealing some promising performances to create sophisticated treatment plans. Besides these, some other methods are used with less frequency for this problem [8].

2.2 Problem formulation

The first step to build a treatment plan consists of outlining the shape of the tumor structure and of the OARs involved in the process. This is usually done by a doctor, who has the control over the prescribed dose of radiation that is needed in order to kill the tumor cells, as well as other important values that characterize each case. These dose values are set intuitively based on his/her personal experience and also on stored data from other treatments. Besides this, the doctor also decides the number of beams required according to the complexity of the tumor. As mentioned earlier, it will be considered that beam angles are chosen according to a coplanar treatment, i.e., they are chosen on a circle around the patient's body, on the same vertical plane as the center of the tumor. In this section, the nonlinear programming formulation of the BAO problem proposed by Bertsimas *et al.* in [3] is reviewed.

For a given fixed number of beams, n, the following notation is introduced:

- S: denotes the number of volumes of interest (VOIs), i.e., OARs, normal tissue and target volumes, that need to be considered for the treatment plan,
- V_h: denotes the set of all voxels associated with a certain VOI h, $h = 1,\ldots,S$,
- N_h: denotes the number of voxels in VOI h, $h = 1,\ldots,S$,
- LB_i: denotes the prescribed dose for each voxel $i \in V_h$,
- UB_i: denotes the maximum dose for each voxel $i \in V_h$,
- B: denotes the set of beamlets, partitioned into subsets $\bigcup_{k=1}^{n} B_k$.

The decision variables of the program are

- $x_j \geq 0$, denoting the beamlets intensities, for $j \in B_k$, $k = 1, \ldots, n$,
- $0 \leq \theta_k \leq 360$, denoting the angle, or direction, of the beam, for $k = 1, \ldots, n$. The variable θ_k is assumed to range continuously from 0° to 360°, which means that the gantry of the linear accelerator can rotate freely around the patient, having total access to every possible angle of this imaginary circle.

Additionally, three groups of structures are considered, each one identified by an index $p = 1,2,3$, $p = 1$ for OARs, $p = 2$ for target volumes – that is, the tumor, and $p = 3$ for both OARs and target volumes. The set of voxels involved in the objective function (constraints) and associated with group p is denoted by O_p (C_p), for $p = 1,2,3$. The objective function aggregates three components, one for each group of structures, and each one intends to measure differences between the dose delivered and the dose prescribed by the doctors for each structure. Namely,

1. For $p = 1$ the objective function includes a weighted convex combination of the maximum dose delivered for voxels in O_1 and of the mean dose in the same voxels – that is, given $0 \leq \alpha_h \leq 1$,

$$\alpha_h y_h^{\max} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i, \quad h \in O_1,$$

   where d_i is an auxiliary variable that represents the dose delivered to voxel i, $i \in V_h$ and $h = 1, \ldots, S$.

2. The objective function for $p = 2$ is similar to that for $p = 1$, but now the combination involves the minimum dose delivered for voxels in O_2 and of the mean dose in the same voxels, that is, given $0 \leq \alpha_h \leq 1$,

$$\alpha_h y_h^{\min} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i, \quad h \in O_2.$$

3. For $p = 3$ the objective function includes the mean for auxiliary variables z_i^h, defined later on, with the purpose of penalizing underdosing the target volumes or overdosing the OARs,

$$\frac{1}{N_h} \sum_{i \in V_h} z_i^h, \quad h \in O_3.$$

As mentioned earlier, the objective function, in order to be minimized, aggregates these three components as a weighted sum with weights $w_h^+ \geq 0$, for $h \in O_1$, $w_h^- \leq 0$, for $h \in O_2$, and $w_h \geq 0$, for $h \in O_3$, which depend on the importance of the respective structures. The used weights may depend on the voxels, but usually a common value is used for voxels in the same structure. It should be remarked that the weights w_h^- are non-positive because, for voxels in target volume structures, the goal is to maximize the delivered dose. In conclusion, the objective function can be written as

$$f(\theta, x) = \sum_{h \in O_1} w_h^+ \left(\alpha_h y_h^{\max} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i \right) + \sum_{h \in O_2} w_h^- \left(\alpha_h y_h^{\min} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i \right) + \sum_{h \in O_3} \frac{w_h}{N_h} \sum_{i \in V_h} z_i^h.$$

As before, let $A = [a_{ij}(\theta_k)]$ be the dose influence matrix, where each $a_{ij}(\theta_k) \in [0,1]$ identifies the dose deposited in voxel i by unit intensity of beamlet j from direction θ_k, for $i \in V_h$, $j \in B_k$, $k = 1,\ldots,n$. The problem constraints include defining the dose delivered to each voxel, d_i,

$$d_i = \sum_{k=1}^{n} \sum_{j \in B_k} a_{ij}(\theta_k) x_j, \;\; i \in V_h, \;\; h = 1,\ldots,S,$$

and bounding the dose delivered to each voxel,

$$LB_i \leq d_i \leq UB_i, \;\; i \in V_h, \;\; h = 1,\ldots S.$$

The variables $y_h^{\max}$ are the maximum dose delivered to voxel h, that is,

$$y_h^{\max} = \max\{d_i : i \in V_h\}, \;\; h \in O_1 \cup C_1,$$

which can be replaced with the linear inequalities

$$y_h^{\max} \geq d_i, \;\; i \in V_h, \;\; h \in O_1 \cup C_1.$$

The same holds for variables $y_h^{\min}$, but as the minimum dose delivered to voxel h,

$$y_h^{\min} \leq d_i, \;\; i \in V_h, \;\; h \in O_2 \cup C_2.$$

The weighted convex combinations used in the objective function for $p = 1$ and $p = 2$ are also bounded by the following constraints:

$$\alpha_h y_h^{\max} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i \leq g_h^u, \; h \in C_1$$

and

$$\alpha_h y_h^{\min} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i \geq g_h^l, \; h \in C_2,$$

respectively, where the values of g_h^u and g_h^l for the target and the OAR doses are provided by the doctors.

Finally, the target dose for VOI h is represented by t_h, and a 'V' shaped function is defined, where each side has its own slope: the left slope, s_h^l, and the right slope, s_h^r. This function is used to limit underdosing to a target value and overdosing to an OAR. Then, the variables z_i^h, $i \in V_h$, are defined by

$$z_i^h \geq s_h^u (d_i - t_h), \; h \in O_3 \cup C_3, \; i \in V_h$$

and

$$z_i^h \geq s_h^l (t_h - d_i), \; h \in O_3 \cup C_3, \; i \in V_h.$$

Such variables are subject to the additional constraints

$$\sum_{k=1}^{n} \frac{z_i^h}{N_h} \leq g_h^m,\ h \in C_3,$$

which bound the mean dose by the parameter g_h^m. The parameters and the slopes in the 'V' function are determined and modulated by the doctors.

In conclusion, the BAO problem can be formulated as follows:

$$\text{minimize}\ \ f(\theta, x) \tag{2.3}$$

$$\text{subject to}\ \ d_i = \sum_{k=1}^{n} \sum_{j \in B_k} a_{ij}(\theta_k) x_j,\ i \in V_h,\ h = 1, \ldots, S \tag{2.4}$$

$$LB_i \leq d_i \leq UB_i,\ i \in V_h,\ h = 1, \ldots, S \tag{2.5}$$

$$x_j \geq 0,\ j \in B_k,\ k = 1, \ldots, n \tag{2.6}$$

$$0 \leq \theta_k \leq 360,\ k = 1, \ldots, n \tag{2.7}$$

$$y_h^{\max} \geq d_i,\ i \in V_h,\ h \in O_1 \cup C_1 \tag{2.8}$$

$$y_h^{\min} \leq d_i,\ i \in V_h,\ h \in O_2 \cup C_2 \tag{2.9}$$

$$\alpha_h y_h^{\max} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i \leq g_h^u,\ h \in C_1 \tag{2.10}$$

$$\alpha_h y_h^{\min} + \frac{1-\alpha_h}{N_h} \sum_{i \in V_h} d_i \geq g_h^l,\ h \in C_2 \tag{2.11}$$

$$z_i^h \geq s_h^u (d_i - t_h),\ h \in O_3 \cup C_3,\ i \in V_h \tag{2.12}$$

$$z_i^h \geq s_h^l (t_h - d_i),\ h \in O_3 \cup C_3,\ i \in V_h \tag{2.13}$$

$$\sum_{k=1}^{n} \frac{z_i^h}{N_h} \leq g_h^m,\ h \in C_3. \tag{2.14}$$

The objective of this formulation is to reduce the weighted convex combination of mean dose and maximum dose delivered to the OARs (O_1), while enhancing the weighted convex combination of the minimum dose and the mean dose delivered to the targets (O_2) and while penalizing underdosing to the target and overdosing to the OARs (O_3). These parameters used in the model are derived from the interaction with doctors, based on their experience.

Given the complexity of the whole surrounding process, an optimal solution for this problem could be very difficult to find. The nonlinearity caused by the constraints (2.4), the restrictive constraints imposed, an expected highly irregular behavior and the uncertainty of being a differentiable function are all factors that lead to a very critical, or even impossible resolution of the problem.

Due to this, several approaches have been studied and submitted to phantom and real-life cases, revealing promising results when applied to this geometric problem. Some of them will be described in the following chapters and their application to this particular problem will be clarified.

Chapter 3

Radiation interaction with matter

In Chapter 2 the formulation proposed by Bertsimas *et al.* in [3] was introduced. The goal of this formulation is to represent the effects caused by a certain set of radiation beams on a patient's body. A feasible solution of this problem ought to accomplish tumor and other organs constraints defined by technicians. However, the behavior of these radiation beams when traversing the body should be discussed in order to complete this formulation. The present chapter will focus on the interaction of the radiation with matter, in order to include this information in the previous formulation.

3.1 Absorption of energy

Much of the research in the field of radiation therapy is aimed at developing new and more accurate techniques for treating cancer patients with radiation. The radiation applied in the patients does not have a uniform behavior and, as such, this will be analyzed with the aim of increasing the accuracy and efficiency of this treatment.

When an X-ray beam (i.e. a beam of photons, with wavelengths ranging from 0.01 to 10 nm) passes through body tissue, part of its energy is deposited at each location along the path of the beam. The energy absorbed per unit of mass is called absorbed dose and it is very useful when predicting the realistic biological effects of the radiation exposure.

The physical processes associated with the radiation exposure through a medium are not straightforward. The first step in this process usually involves a collision between a photon (elementary particle of electromagnetic radiation) and an electron (stable subatomic particle with negative electric charge) in the body, culminating in the scattering of some radiation and in the projection of a high speed electron. After this projection, the high speed electron can form ionized species and existing molecular bonds, if any, can be broken, resulting in biological damage. Most of the energy, however, is converted into heat, which has little or no effect in the biological system. Additionally, a high speed electron may also suffer a collision with an atomic nucleus, which causes the deceleration of the considered electron, producing electromagnetic radiation. This effect is called *Bremsstrahlung*, a German term for "broken radiation" or "deceleration radiation". These physical phenomena can then undergo interactions in the same way as the original photon. In fact, around 30 interactions are required before all the energy of the photon is converted into electronic motion. A diagram that outlines these interactions is shown in Figure 3.1.

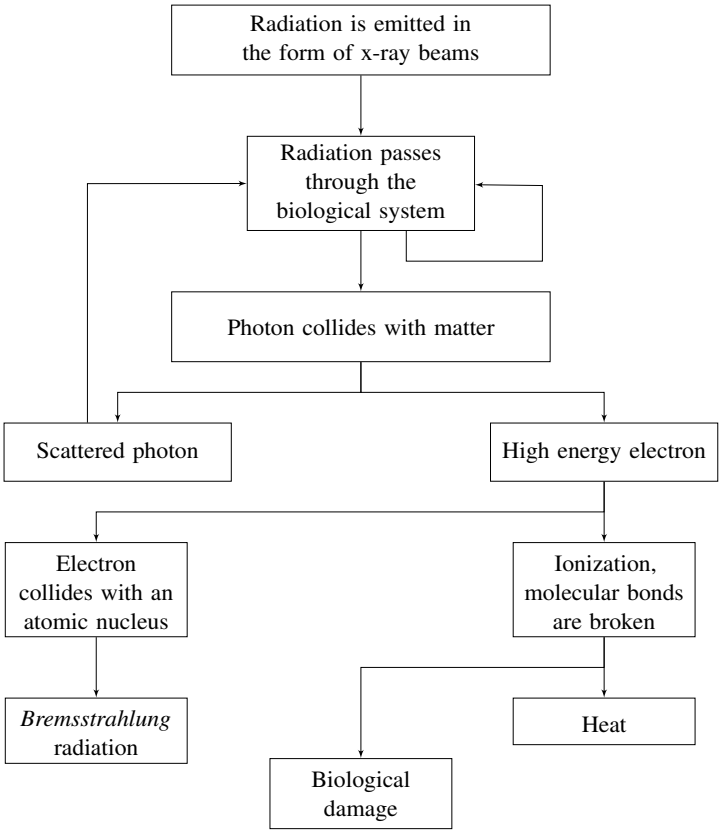


Fig. 3.1 Absorption of energy from radiation resulting in biological damage

X-ray photons may interact with the atoms to produce high speed electrons by five mechanisms known as the photon disintegration, photoelectric process, Compton process, pair production and coherent scattering. Further details on these processes can be found in [14]. In the following chapters, some aspects related to the radiation attenuation and other characteristics of the beam will be analyzed.

3.2 Linear attenuation coefficient

In order to understand the following formulations, let the passage of an X-ray beam through a slab of material of thickness Δx and the existence of a detector P right after, be considered. Let the number of photons recorded by that detector be N, and the number of photons removed from the beam after the interaction with the attenuator be n.

As explained in the previous section, either the photon comes close enough to the atom to interact with it and thus loses the original direction of the beam, or it is not affected at all, thereby maintaining its trajectory. For that reason, n will proportionally depend on the number of photons present in the beam, N. Furthermore, if Δx increases, the number of atoms placed in the beam increases as well with the same ratio, thereby also increasing the chance of interaction. Thus, n varies as the product of N and Δx, and thus can be written as follows

$$n = \mu N \Delta x, \tag{3.1}$$

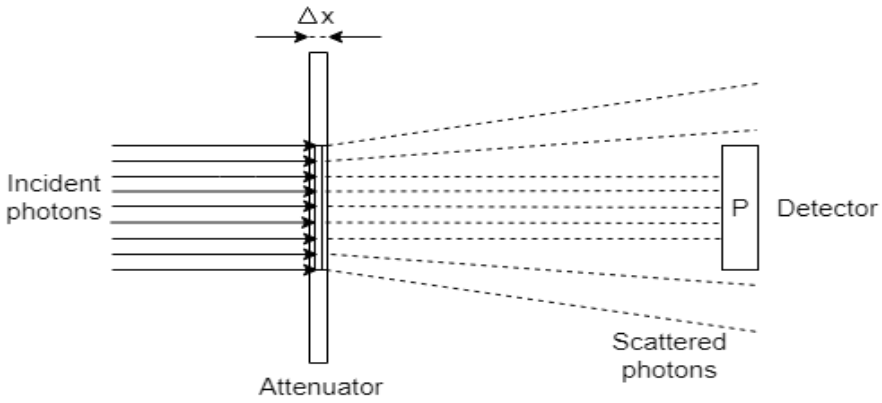


Fig. 3.2 Scattering radiation produced through an attenuator of thickness Δx

where μ denotes the linear attenuation coefficient, and characterizes the gradual decrease in energy of the X-radiation for a certain type of absorbing material. For example, bone absorbs X-rays better than flesh and, consequently, it has a higher value of μ. Lastly, it is necessary to know the dimension to express μ. Given that N and n are integer numbers with no dimensions, $\mu\Delta x$ must also be dimensionless. If μ, is expressed by cm^{-1}, then x must be measured in cm. In order to segregate μ, the equation (3.1) can be rearranged as follows, showing that μ is a fraction of the attenuator

$$\mu = \frac{(n/N)}{\Delta x} = \frac{\text{fraction of photons that interact in } \Delta x}{\text{thickness of the attenuator}}. \tag{3.2}$$

Now, if ΔN represents the change in the number of photons in the beam when passing through Δx, then equation (3.1) yields

$$\Delta N = -\mu N \Delta x. \tag{3.3}$$

The thickness Δx must be small enough so that it is possible to evaluate its instantaneous effect through the layer. Taking the proportionality between ΔN and Δx, and replacing these terms by their corresponding differentials dN and dx, the equation (3.3) becomes a first order linear differential equation:

$$-dN = \mu N \, dx.$$

It can be noticed that the number of photons decreases from N_0 to N and respectively the thickness of the attenuator varies between 0 and x and it yields:

$$\begin{aligned} \int_{N_0}^{N} \frac{dN}{N} &= -\int_0^x \mu \, dx \Leftrightarrow \\ \Leftrightarrow \ln(N) - \ln(N_0) &= -\mu x \Leftrightarrow \\ \Leftrightarrow N &= N_0 \, e^{-\mu x}. \end{aligned} \tag{3.4}$$

This equation may be used to calculate the attenuation for any thickness of material, whereas equation (3.1) is only applicable and reliable when the fractional reduction by a layer of material is considerably small.

To illustrate the difference between equations (3.1) and (3.4), the number of photons absorbed from a beam by 1 cm layers of material will be calculated by applying both equations. Assuming that $\mu = 0.2$ cm^{-1} and the initial number of photons from the beam is $N_0 = 1000$, the Table 3.1 is obtained. It is noticeable that the number of interactions per layer becomes smaller as layers of attenuators are added. The results obtained for both equations are considerably different, since the calculations obtained from equation (3.1) are not accurate.

Table 3.1 Attenuation of a beam considering $\mu = 0.2$ cm^{-1}

Thickness attenuation x (cm)	Number of interactions in 1 cm of slab ($N_0 - N$)		Number of photons left in beam (N)	
	Equation (3.1)	Equation (3.4)	Equation (3.1)	Equation (3.4)
0	0	0	1000	1000
1	200	181.27	800	818.73
2	160	148.41	640	670.32
3	128	121.51	512	548.81
4	102.40	99.48	409.60	449.33
5	81.92	81.45	327.68	367.88
6	65.54	66.69	262.14	301.19
7	52.43	54.60	207.71	246.60
8	41.94	44.70	165.77	201.90
9	33.55	36.60	132.22	165.30
10	26.84	29.96	105.38	135.34

Besides this, the total number of photons, calculated by equation (3.4), is only valid when the beam is narrow. A broad beam, compared to a narrow beam, exhibits a much less regular behavior, since its incidence is much broader, being able to reach heterogeneous media with different compositions.

The need for a narrow beam can be interpreted comparing Figures 3.2 and 3.3. The radiation range from the broad beam covers a wider area from the attenuator, reaching some areas which the narrow beam does not reach. For this reason some photons may be scattered and present an unexpected performance. Although one could still define an attenuation coefficient by equation (3.2), in some cases the attenuation coefficient is no longer a constant. Its value depends on the thickness of the attenuator, on its area and shape, and on the distance between the attenuator and the detector (or the following attenuator), as well as on the photon energy. The new behavior can be described by the equation

$$N = N_0\, e^{-\mu x} B(x, h\nu, A, L) \tag{3.5}$$

where B is a rather complicated factor, sometimes called *photon build-up factor*, which estimates the amount of photons scattered by the attenuator. This factor depends on several parameters: thickness (x), total energy ($h\nu$), area (A) and distance (L). The factor B ranges in magnitude between 1 and 100 and is usually obtained experimentally. Values for B can be found in the literature [9, 18].

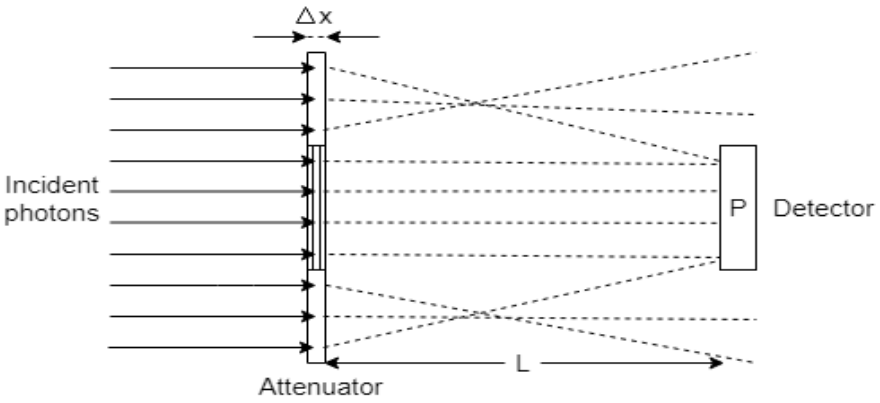


Fig. 3.3 Effect of a broad beam radiation through a heterogeneous attenuator, with different attenuation coefficients

3.3 Transferred and absorbed energy

The beam intensity represents the total amount of energy going through a medium per unit area and per unit time. The total energy, E_{total}, can be calculated from the number of photons of the beam, N, and their individual energy, considering that the photons in the incident beam have the same energy, composing a monoenergetic beam, as:

$$E_{total} = N \times E_{photon},$$

with $E_{photon} = hc/\lambda$, where c is the speed of light in vacuum, λ is the photon's wavelength and h is known as the Planck constant. Since the frequency, that is, the number of occurrences of the considered event per unit of time, is given by $\nu = c/\lambda$, then the photon energy equation can be simplified to

$$E_{photon} = h\nu.$$

This means that, for a constant frequency, when all the photons assume the same energy, the proportionality between beam energy E and frequency ν is considered, while not absorbed by the tissue,

$$E \propto \nu.$$

Therefore, the equation (3.4) can be rewritten as

$$E(x) = E_0\, e^{-\mu x}, \tag{3.6}$$

where $E(x)$ represents the total energy of the beam after passing through a material of thickness x and E_0 represents the initial energy of the beam. This is only valid if the beam attenuation coefficient, μ, is a constant.

As briefly explained in Section 3.1, when a photon interacts with an absorber, a set of complex phenomena takes place (Figure 3.1). In general, part of the energy of the photons is radiated as scattered radiation and another part is converted into kinetic energy of high speed electrons or

positrons [5]. Once these particles have been set in motion, they are at risk of losing energy by collisional losses or by emitting *Bremsstrahlung* radiation. It is impossible to predict exactly what will happen for every specific case; however, after a considerable number of interactions, one can calculate the kinetic energy received by charged particles in the volume, regardless of how they dissipate the energy, $\bar{E}_{tr}$, and the respective kinetic energy absorbed by the volume, $\bar{E}_{ab}$. The difference between these two energies is equal to the energy radiated away, as *Bremsstrahlung*,

$$\bar{E}_{tr} - \bar{E}_{ab} = \bar{E}_{ra}. \tag{3.7}$$

However, most of the interactions of interest in radiology occur in materials with a low number of electrons per atom. The energy emitted as *Bremsstrahlung* is negligible for those cases (up to 10 MeV). Taking into account that, in radiotherapy, the photon energies rarely exceeds that quantity, it is often considered that $\bar{E}_{tr} \approx \bar{E}_{ab}$.

To calculate the energy absorbed by a block of tissue, it is often convenient to use the energy absorption coefficient μ_{ab}, which is a quantity of interest in predicting the biological effects of the absorbed radiation in the tissues. Hereafter, the equation that calculates the absorbed dose, through the length x, is

$$E_{ab}(x) = E_0\, e^{-\mu_{ab}x}. \tag{3.8}$$

The negligence of the secondary interactions not only simplifies this equation, but also the implementation of the radiation attenuation in the formulation of the BAO problem considered in Chapter 2. Given this, the problem reformulation is considered and explained in the following section.

3.4 Beam angle optimization problem formulation

Considering the formulation presented in Section 2.2, the gradual loss of energy flux considered in this chapter is applied directly to the dose deposition matrix A. Thereby, matrix A assumes values in the interval $[0,1]$ according to the attenuation of the radiation in voxel i, from the bixel j, positioned in the angle k, for $i \in V_h$, $j \in B_k$, $k = 1,\ldots,n$.

Additionally, the energy measures should also be adjusted. The energy radiated from the bixels, E_0, is usually measured in MeV (megaelectron volt). By definition, the electron volt is the amount of energy gained (or lost) by the charge of a single electron moving across an electric potential difference of 1 volt, whereas 1 megaelectron volt is equal to $1.60217662 \times 10^{-13}$ J (joule). On the other hand, the energy absorbed by the voxels, E_{ab}, is measured in Gy (gray). Gray is a derived unit of ionizing radiation dose, defined as the absorption of 1 joule of radiation energy per kilogram of matter and, hence, it is equal to 1 J $\cdot$ kg^{-1}. Consequently, we can assume a new requirement to the formulation:

$$a_{ij}(\theta_k) = \begin{cases} \frac{1.60217662\times10^{-13}}{W_h^v} \left(\frac{l^v}{l^b}\right)^2 e^{-\mu_h L_{ij}(\theta_k)}, & \text{if voxel } i \text{ is irradiated by beamlet } j, \text{ from } \theta_k, \\ 0, & \text{otherwise,} \end{cases} \tag{3.9}$$

where W_h^v, $h = 1,\ldots,S$, is the weight (in kg) of the voxel i, $i \in V_h$ considered in the absorption, $(l^b)^2$ and $(l^v)^2$ represent the surface area of the bixels, that is, the area of incidence of the beamlets, and the

surface area of the voxels, respectively. Therefore the value $(l_j^b/l_i^v)^2$ represents the surface portion of radiation projection of a certain beamlet j, $j \in B_k$, $k = 1,\ldots,n$, to the voxel i, $i \in V_h$, $h = 1,\ldots,S$. This fraction is calculated assuming that the energy beamlet is uniform in all its area. Besides this, μ_h denotes the absorption energy coefficient (in cm^{-1}) of the set of voxels associated with a certain VOI h, $h = 1,\ldots,S$. Finally, $L_{ij}(\theta_k)$ represents the distance (in cm) traversed by the radiation through the body until it reaches the voxel i, for $i \in V_h$, $j \in B_k$ and $k = 1,\ldots,n$. Lastly, in the new formulation, the variable x_j is now representing the energy emitted by the beamlet j (in MeV), rather than its intensity, for any $j \in B_k$ and $k = 1,\ldots,n$.

Chapter 4

Heuristic methods

Most of the search methods have the advantage of being generally applicable and flexible. Nevertheless, for complex combinatorial problems, these methods may not be sufficient to achieve good optima values. Many studies and investigations were made in order to obtain search algorithms that are able to find higher quality solutions in lower running times. Following this line of thought, over the past few decades, *heuristic algorithms*, or simply *heuristics*, have emerged and been developed. They can be used to speed up the process by employing a practical unorthodox method, which is not guaranteed to be optimal, but could be sufficient for the required goals, finding a satisfactory solution among a large number of alternative solutions.

In addition to the need of finding good approximated solutions in a reasonable amount of time, there may be other possible reasons that justify using these approximation methods:

1. No exact method to solve optimal problems is known.
2. Although an exact method is known, it is not possible to execute it on the hardware available.
3. The heuristic method is more flexible than the exact method, allowing the implementation of conditions that could make the exact method unsolvable.

Besides this, there are some properties that the heuristic methods should fulfill:

1. A solution has to be obtained with a balanced computational effort.
2. The solution should be close to the optimal.

Many heuristic methods have been designed to solve a specific problem without the possibility of generalization or application to other similar problems. Many others have to be adapted and their parameters and procedures are fitted to a specific problem.

The beam angle optimization problem introduced in Chapter 2, and in particular the formulation (2.3)–(2.14) presented by Bertsimas *et al.* [3], has been reported to be complex to solve due to its nonlinearity and a non-convex objective function. For this reason, several heuristic methods have been presented to reach a feasible solution of the BAO problem in a reasonable amount of time. The proposed methods dynamically explore angles, fixing them iteratively, and allowing for the minimization of the corresponding intensities.

This chapter is dedicated to the heuristic methods that could be implemented to the BAO problem presented in Section 2.2.

4.1 Simulated annealing

Simulated annealing (SA) [1, 10] belongs to the class of heuristic methods and its emergence was based on an analogy with the physical process of annealing used in the metallurgy. In this process, a pure lattice structure of a solid is made by increasing the temperature of the heat bath heating up the solid object until it melts, then reversing this process by cooling the heat bath down carefully, until the particles arrange themselves into a low energy-state. In the early phase, all particles of the solid arrange themselves randomly. In the ground state, the particles are arranged in a highly organized and effective structure in a way that makes the energy of the system minimal. The ground state of the solid is exclusively obtained if the maximum temperature is sufficiently high and the reversing process of cooling is done sufficiently slowly. Otherwise the solid will solidify into a meta-stable state rather than into the ground state.

Heuristic methods, like the SA, can be applied on certain problems with some success, providing an approximated solution. The SA method has already been used for an enormous number of optimization problems such as the traveling salesman problem, the placement of components on microcircuitry, or in astrophysics, among others. SA plays a special role in this field for two reasons. First, it appears to be quite successful when applied to a wide range of practical combinatorial problems, which has given it a large amount of credibility. Second, its stochastic component facilitates the theoretical analysis of its asymptotic convergence, which made it noticeable between mathematicians.

Let (Ω, f) be an instance of a minimization problem with feasible solutions in Ω, which will be called states from now on, and an objective function $f : \Omega \to \mathbb{R}$. The aim of the SA method is to reach an optimal solution $x^* \in \Omega$, minimizing the value of f over Ω. The method is presented in the following.

Step 1 The SA procedure starts by considering an initial state x^0 from Ω as the benchmark state, an initial temperature $T_0 > 0$, a kind of temperature dependent density function to generate new possible points and a corresponding temperature updating function. The objective value $f(x^0)$ is calculated from the first point of the sequence and the minimum variable is set as $x_{\min} = x^0$ and $f_{\min} = f(x^0)$.

Step 2 The strategy continues with the calculation of a random vector z^k by using the generation probability density function. Then z^k is added to the current iteration point x^k, generating y^k, i.e., $y^k = x^k + z^k$. If the generated point y^k is a feasible solution ($y^k \in \Omega$), the algorithm should proceed to Step 3; otherwise, the current step is repeated.

Step 3 A random value η is generated with a uniform distribution over $(0,1)$ and, then, the probability of accepting the trial point y^k as the new iteration point x^{k+1}, given x^k and T_k, is calculated as follows:

$$P(y^k|(x^k, T_k)) = \min(1, \exp\{(f(x^k) - f(y^k))/T_k\}).$$

The acceptance process of y^k is known as the *Metropolis acceptance criterion* and depends on x_k and T_k. The temperature controls the likelihood of an objective function value increase. At the beginning, when T_k is large, there is a bigger probability of an increase of the objective function value, allowing the algorithm to choose a higher value to overcome a lower one, whilst

avoiding it to run a general downhill algorithm and to be stuck at a local minimal. At the end of the method, when T_k gets closer to 0, there is a lower probability of accepting an increase of the objective function value.

Therefore, if $\eta \leq P(y^K|(x^k,T_k))$, the Metropolis acceptance criterion sets

$$x^{k+1} = y^K \text{ and } f(x^{k+1}) = f(y^k),$$

otherwise

$$x^{k+1} = x^k \text{ and } f(x^{k+1}) = f(x^k).$$

Step 4 The objective function value of the new iteration x^{k+1} is compared to the optimal value $f_{\min}$ achieved. So, if $f(x^{k+1}) < f_{\min}$, set $x_{\min} = x^{k+1}$.

Step 5 If the scheduled stopping condition is satisfied, then the algorithm stops; otherwise, it updates the temperature by means of the function referred to in Step 1 and it returns to Step 2.

Algorithm 1 outlines the pseudo-code of the SA method, given an objective function f, the domain interval Ω, the neighborhood function N used to calculate the random vector z^k and the maximum temperature T_0. Moreover, it is assumed that the temperature is updated at every iteration.

Algorithm 1 Simulated annealing algorithm

procedure SA(f, N, Ω, x^0, T_0)
  $k \leftarrow 0$
  $x_{\min} \leftarrow x^k$
  $f_{\min} \leftarrow f(x_{\min})$
  $T_k \leftarrow T_0$
  while stopping criterion is not satisfied **do**
    $z^k \leftarrow rand\left(N(x^k,T_k)\right)$
    $y^k \leftarrow x^k + z^k$
    if $rand(0,1) \leq \min\{1, \exp\{(f(x^k) - f(y^k))/T_k\}\}$ **then**
      $x^{k+1} \leftarrow y^k$
    else
      $x^{k+1} \leftarrow x^k$
    if $f(x^{k+1}) < f_{\min}$ **then**
      $x_{\min} \leftarrow x^{k+1}$
      $f_{\min} \leftarrow f(x_{\min})$
    $k \leftarrow k+1$
    $T_{k+1} \leftarrow$ temperature is updated
  return $x_{\min}$

We illustrate the described SA method with a simple example in $\mathbb{R}^2$. The simulation was implemented in Matlab RS2015a. Even though this is a phantom case, it can recreate the SA behavior and its ability to look for a global minimal avoiding local minima. The parameters of this simulation were selected in order to obtain a suitable example.

Let f be the objective function given by $f(x) = \cos(2x) - \frac{x}{5}$, with $\Omega = [-2,3]$, while the initial temperature is $T_0 = 1$, the stopping criteria $T_k = 0$ and the cooling steps $T_{k+1} = T_k - 0.05$. The

acceptance probability function, $e^{-(f(y)-f(x))/T}$, and the generation of new points both used the uniform distribution.

The Figure 4.1 shows the behavior of the applied SA method in order to find the global minimum of function f with the above parameters, starting from a randomly generated initial point. Each point in the plot represents an iteration that improved the previous minimal point. The red bullet represents the final solution, obtained by the method after 6 improved iterations. The initial point was around $x^0 \approx -1$ and, even though the algorithm has moved away from the global minimum of f in Ω several times, the final solution seems to be very close to the global minimum. The output value from this SA trial was $x_6 \approx 1.3405$, while the global minimum is $x^* = \frac{1}{2}\left(\pi - \arcsin\left(-\frac{1}{10}\right)\right) \approx 1.6209$. Thus, the absolute error of this approximation is around 0.2804, which makes x_6 a reasonable approximation to the global minimum, taking into account the number of iterations. Despite the good performance of the SA method for this case, it should be remarked that this is a probabilistic based method and, therefore, it is not possible to ensure it always behaves the same way.

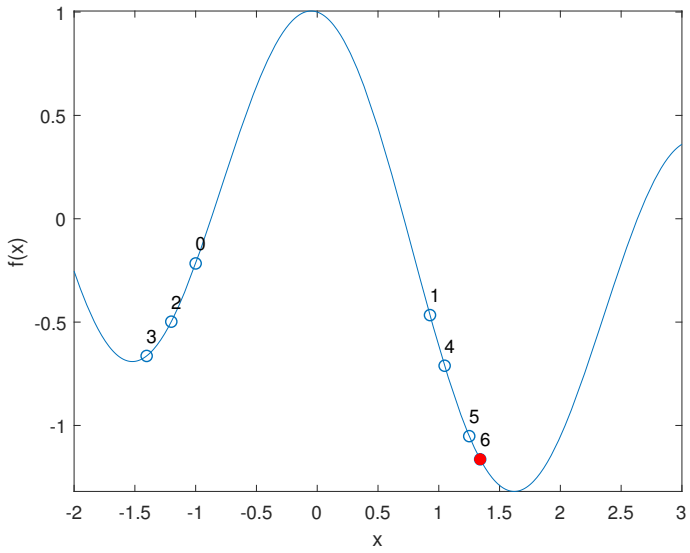


Fig. 4.1 SA iterations and final approximation.

The following theorem gives a global convergence property of the minimum objective value sequence

$$\left\{ \min_{0 \le j \le k} f(x^j);\ k \ge 0 \right\} \tag{4.1}$$

corresponding to the random iteration points induced by the SA algorithm.

Theorem 1 ([23]). *Suppose that $\mu(\Omega_\varepsilon) > 0$ holds for every $\varepsilon > 0$, where*

$$\Omega_\varepsilon = \{x \in \Omega |\ f(x) < f^* + \varepsilon\},$$

where $\mu(\cdot)$ denotes the Lebesgue measure on $\mathbb{R}^n$, and that there exists an integer $k_0 > 0$ such that $p(\cdot, T_k)$, the temperature dependent generation probability density function used to generate the

random vectors, satisfies the following condition:

$$\min_{x,y\in\Omega} p(y-x,T_k) \geq \frac{M}{k},\ \forall k \geq k_0, \tag{4.2}$$

where $M > 0$ is a constant. Then, for any initial state x^0 of f in Ω, the sequence (4.1) of minimum objective values converges in probability to the global minimum f^*.

Proof. For any $k \geq 0$, $x^k \in \Omega$ and $\Omega_1 \in \Omega$, let

$$F_k = \min_{0\leq j\leq k} f(x^j)$$
$$P_{k+1}(x^k,\Omega_1) = P(x^{k+1} \in \Omega_1 |\, x^k).$$

To prove this theorem, it is required to show that, for any $\varepsilon > 0$ and initial state $x^0 \in \Omega \setminus \Omega_\varepsilon$, the minimum objective value sequence $\{F_k, k \geq 0\}$, induced by the SA algorithm, satisfies the following relation:

$$\lim_{k\to\infty} P(F_k - f^* \geq \varepsilon) = 0.$$

For all $\varepsilon > 0$ and initial state $x^0 \in \Omega \setminus \Omega_\varepsilon$,

$$\begin{aligned}
P(F_k - f^* \geq \varepsilon) =& P(f(x^k) - f^* \geq \varepsilon, f(x^{k-1}) - f^* \geq \varepsilon, \ldots, f(x^1) - f^* \geq \varepsilon, f(x^0) - f^* \geq \varepsilon) \\
=& P(x^k \in \Omega \setminus \Omega_\varepsilon, x^{k-1} \in \Omega \setminus \Omega_\varepsilon, \ldots, x^1 \in \Omega \setminus \Omega_\varepsilon | x^0) \\
=& \int_{\Omega\setminus\Omega_\varepsilon} P_1(x^0,dx^1) \int_{\Omega\setminus\Omega_\varepsilon} P_2(x^1,dx^2) \ldots \left[\int_{\Omega\setminus\Omega_\varepsilon} P_{k-1}(x^{k-2},dx^2) \max_{x^{k-1}\in\Omega\setminus\Omega_\varepsilon} P_k(x^{k-1},\Omega\setminus\Omega_\varepsilon) \right] \\
\leq& \prod_{j=0}^{k-1} \max_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega\setminus\Omega_\varepsilon).
\end{aligned}$$

This gives

$$\begin{aligned}
\lim_{k\to\infty} P(F_k - f^* \geq \varepsilon) \leq& \prod_{j=0}^{k-1} \max_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega\setminus\Omega_\varepsilon) \\
=& \exp\left\{ \log \prod_{j=0}^{k-1} \max_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega\setminus\Omega_\varepsilon) \right\} \\
=& \exp\left\{ \sum_{j=0}^{\infty} \log \left[\max_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega\setminus\Omega_\varepsilon) \right] \right\} \\
=& \exp\left\{ \sum_{j=0}^{\infty} \log \left[1 - \min_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega_\varepsilon) \right] \right\}.
\end{aligned}$$

Since

$$\min_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega_\varepsilon) \in [0,1)$$

the inequality could be written as:

$$\lim_{k\to\infty} P(F_k - f^* \geq \varepsilon) \leq \exp\left\{ -\sum_{j=0}^{\infty} \min_{x^j \in \Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j, \Omega_\varepsilon) \right\}. \tag{4.3}$$

For $j > 0$,

$$\begin{aligned}
\min_{x^j\in\Omega\setminus\Omega_\varepsilon} P_{j+1}(x^j,\Omega_\varepsilon) &= \min_{x^j\in\Omega\setminus\Omega_\varepsilon} P(y^j \in \Omega_\varepsilon |\, x^j) \\
&= \min_{x^j\in\Omega\setminus\Omega_\varepsilon} P(x^j + z^j \in \Omega_\varepsilon |\, x^j) \\
&= \min_{x^j\in\Omega\setminus\Omega_\varepsilon} \int_{\Omega_\varepsilon} p(y - x^j, T_j)\, dy \\
&\geq \mu(\Omega_\varepsilon) \min_{\substack{x^j\in\Omega\setminus\Omega_\varepsilon \\ y\in\Omega_\varepsilon}} p(y - x^j, T_j) \\
&\geq \mu(\Omega_\varepsilon) \min_{x,y\in\Omega} p(y - x, T_j) \\
&\geq \mu(\Omega_\varepsilon) M/j.
\end{aligned} \tag{4.4}$$

Hence, applying the inequalities (4.3) and (4.4), it follows that

$$\lim_{k\to\infty} P(F_k - f^* \geq \varepsilon) \leq \exp\left\{ -\sum_{j=0}^{\infty} \mu(\Omega_\varepsilon) M/j \right\}$$

, which completes the proof. □

4.2 Steepest descendent method

The steepest descent (SD) method is one of the simplest gradient methods used to find a local minimal of a function (or a global minimal in case of a convex function). Often, it is undesirable to accurately minimize the function due to its computing cost associated to its complexity. Because of this and also its simplicity and effectiveness, the SD is well-known. This method could be combined with other heuristic methods in order to improve their optimization, as will be seen in the following.

Let (Ω, f) be an instance of an optimization problem with feasible solutions set in Ω, and $f : \Omega \to \mathbb{R}$ an objective function. The SD method can be summed up in the following steps.

Step 1 The SD method starts by considering a state x^0 from Ω, given in advance. Its objective function value, $f(x^0)$, is calculated.

Step 2 The method goes on calculating the symmetric of the gradient of f at point x^k, stored as ρ_k. The value $x^k + \alpha_k \rho_k$ is calculated as well, where $\alpha_k > 0$ (if iteration k has not reached the optimal point) represents the step length of the iteration and ρ_k represents the search direction. If the generated point $x^k + \alpha_k \rho_k$ is in Ω, the algorithm should proceed to Step 3, otherwise repeat the current step with a reduced α_k.

Step 3 A good performance of a unidirectional search method depends on the choices of the ρ_t direction and the α_t step length. The choice of α_t should lead to a substantial and inexpensive decrease of the objective function. Thus, two conditions are employed to require a significant decrease condition to the new iteration $x^k + \alpha_k \rho_k$. These conditions, known as the *Wolfe conditions*, are an inexact line search stipulation that requires α_k to decrease the objective function by a significant amount. The Wolfe conditions encompass the *sufficient decrease condition* (SDC) and the *curvature condition* (CC). The SDC is also known as the *Armijo condition* and is given by the inequality:

$$f(x^k + \alpha_k \rho_k) < f(x^k) + c_1 \alpha_1 \nabla f(x^k)^\top \rho_k, \tag{4.5}$$

where $c_1 \in (0,1)$. The SDC must be paired with CC:

$$\nabla f(x^k + \alpha_k \rho_k)^\top \rho_k \geq c_2 \nabla f(x^k)^\top \rho_k, \tag{4.6}$$

where $c_2 \in (c_1, 1)$. In case these two condition are not satisfied, the value of α_k should be reduced until then.

Step 4 After having a value of α_k that fulfills the Wolfe conditions, the new point x^{k+1} is fixed as $x^k + \alpha_k \rho_k$ and its objective value is calculated.

Step 5 If a preset stopping condition is satisfied, then the algorithm stops, otherwise it returns to Step 2.

Algorithm 2 outlines the pseudo-code of the SD method, providing in advance an objective function f, an initial point x_0, the domain interval Ω, the neighborhood function N and the stopping criterion ε used to control $\| \rho_t \|$. Moreover, it is also assumed that every time a reduction of α_k is required, α_k is divided by $R \in \mathbb{R}_{>1}$.

Algorithm 2 Steepest descent algorithm

```
1:  procedure SD(f, N, Ω, x^0, ε, α, R)
2:      k ← 0
3:      ρ_k ← −∇f(x^k)
4:      while ‖ ρ_k ‖> ε do
5:          y^k ← x^k + α_k ρ_k
6:          if Wolfe conditions are satisfied then
7:              x^{k+1} ← y^k
8:              k ← k + 1
9:              ρ_k ← −∇f(x^k)
10:         α_k ← α_k / R
11:     return x^k
```

We now illustrate the SD method for the example considered in the last section. The constants ε and α have the values 0.001 and 0.1, respectively. The value attributed to R is 2. For the Wolfe constants, $c_1 = 0.3$ and $c_2 = 0.7$. The results of this application are summarized in Figure 4.2. For this case, the SD simulation shows a local performance that reached the local minimum of the concavity.

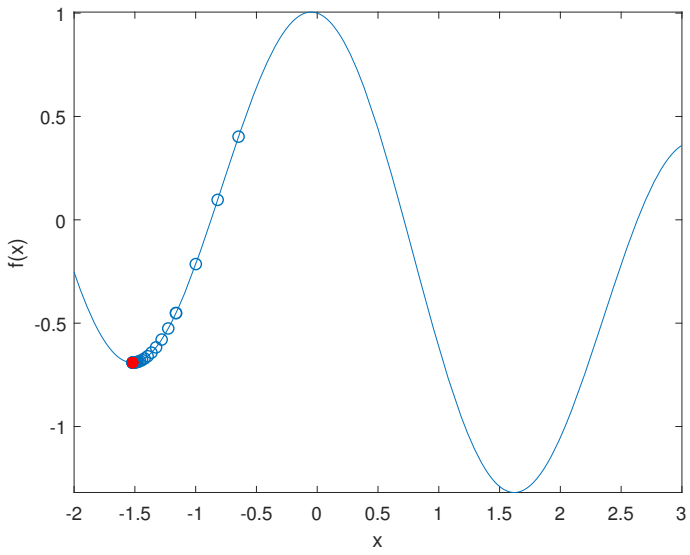


Fig. 4.2 SD iterations and final approximation.

The blue circles in Figure 4.2 represent the first point and the following iterations. Finally, the red bullet represents the output of the method (≈ -1.5205), while the local minimal is $x \approx -1.5207$, so the absolute error on this approximation is around 2.0×10^{-3}. This approximation could be improved by decreasing the threshold ε.

The following theorem claims the global convergence to SD method when the Wolfe conditions are respected.

Theorem 2 ([19]). *Consider that $x^{k+1} = x^k + \alpha_k \rho_k$, with $\rho_k = -\nabla f(x^k)$ and $\alpha_k > 0$, satisfies the Wolfe conditions. Suppose that ∇f is Lipschitz continuous (with $\nu > 0$ as a constant) in an open set, containing the level set $L(x^0) = \{x : f(x) \leq f(x^0)\}$. Further assume that f is lower bounded in $L(x^0)$. Then,*

$$\nabla f(x^k) \longrightarrow 0.$$

Proof. From SDC we have

$$f(x^{k+1}) \leq f(x^k) - (c_1 \parallel \nabla f(x^k) \parallel^2)\alpha_k.$$

An upper bound for α_k is required. From CC

$$\begin{aligned} &-(\nabla f(x^{k+1}) - \nabla f(x^k))^\top \nabla f(x^k) \geq (1-c_2) \parallel \nabla f(x^k) \parallel^2 \\ \Rightarrow &(1-c_2) \parallel \nabla f(x^k) \parallel^2 \leq \nu \alpha_k \parallel \nabla f(x^k) \parallel^2 \\ \Rightarrow &\frac{c_2 - 1}{\nu} \geq -\alpha_k. \end{aligned}$$

Then we have:

1. $f(x^{k+1}) \leq f(x^k) - c_1 \alpha_k \parallel \nabla f(x^k) \parallel^2$,
2. $\frac{c_2-1}{\nu} \geq -\alpha_k$.

Therefore, if $c = c_1 \frac{1-c_2}{\nu}$,

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - c_1 \frac{1-c_2}{\nu} \parallel \nabla f(x^k) \parallel^2 \\ &= f(x^k) - c \parallel \nabla f(x^k) \parallel^2 . \end{aligned}$$

Now, it is possible to conclude that

$$f(x^{k+1}) \leq f(x^0) - c \sum_{j=0}^{k} \parallel \nabla f(x^j) \parallel^2 .$$

Given that, f is lower bounded in $L(x^0)$ and f^* is the lower limit,

$$\begin{aligned} c \sum_{j=0}^{k} \parallel \nabla f(x^j) \parallel^2 &\leq f(x^0) - f(x^{k+1}) \\ &\leq f(x^0) - f(x^*) \end{aligned}$$

and, thus,

$$\left\{ \sum_{j=0}^{k} \parallel \nabla f(x^j) \parallel^2 \right\}$$

is a converging sequence, since it is upper bounded. Therefore,

$$\nabla f(x^k) \longrightarrow 0$$

and, so, it is finally concluded that, under the required conditions, the method will converge. □

4.3 Hybrid method

The hybrid method (HM) was originally developed for a short-term hydrothermal combinatorial problem [21] and was later applied by Bertsimas *et al.* in [3] to the BAO problem. It results from a combination of the SA method and the SD method. While the SD method allows a quick calculation of a local minimal of an objective function, the SA method is used to prevent the method from being trapped on some local minimum with its probabilistic acceptance function. The combination of both methods looks for complementing them, bringing together the best features of each of them. As such, the HM consists on the alternation of the two methods after a certain number of iterations.

The general setting of the HM is similar to the SA method. An objective function, denoted by f, is to be minimized within the domain Ω.

Step 1 The HM procedure starts by considering an initial state x^0 from Ω, an initial temperature $T_0 > 0$, a kind of temperature dependent density function to generate new possible points and a corresponding temperature updating function.

Step 2 The method runs Algorithm 1 for a fixed number of iterations with the initial parameters and functions defined in Step 1. After a certain number of iterations, the output x^k is returned.

Step 3 The method continues to run the Algorithm 2 fully, in order to obtain a value that is as close as requested to the local minimum of the concavity that contains x^k. This way, the comparison with other points, in the next step of the loop, will always be made from a nearby local minimal value.

Step 4 After the SD performance, the method returns to Step 2, in order to give continuity to the SA method. The alternation of these steps is performed until the stopping criteria of HM is reached.

The solution found with this method is expected to be at least as close to the global minimum of the objective function as those obtained when applying simply the SA algorithm, since the approximated solutions are not worsened when using the SD method. To conclude this section, the pseudo-code of the HM method is outlined in Algorithm 3. The parameter β represents the number of iterations of the SA method before applying a new iteration of the SD method.

Algorithm 3 Hybrid method algorithm

1: **procedure** HM(f, N, Ω, x^0, T_0, ε, α, β, R)
2: $k \leftarrow 0$
3: $T_k \leftarrow T_0$
4: **while** stopping criterion is not satisfied **do**
5: $x^{k+1} \leftarrow$ apply β iterations of SA(f, N, Ω, x^k, T_k)
6: $x^{k+1} \leftarrow$ SD(f, N, Ω, x^k, ε, α, R)
7: $k \leftarrow k+1$
8: **return** x^k

We now illustrate the HM method for the example given in the first section of this chapter. The results of this application are shown in Figure 4.3.

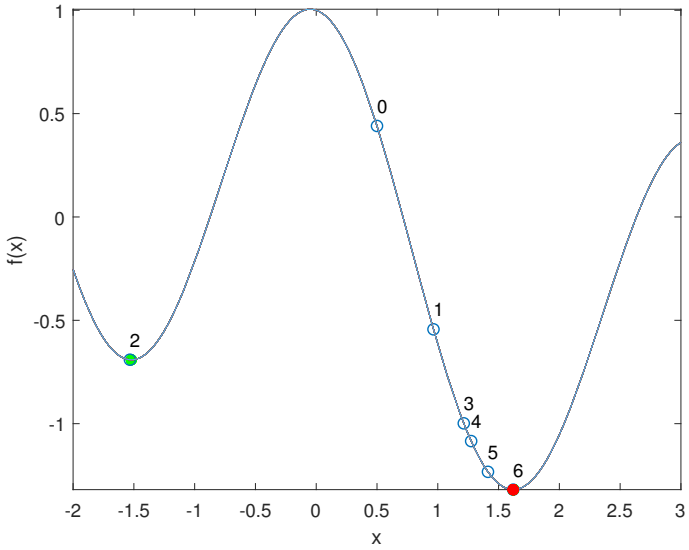


Fig. 4.3 HM iterations and final approximation.

For this case, the HM simulation performed better than the SA in calculating the global minimum. Still, similarly to the SA method, the probabilistic approach of this heuristic method makes it fallible.

The blue circles in Figure 4.3 represent the SA results and the corresponding improved iterations, while the green bullet on the left represents the SD results (at bullet number 2). Finally, the red bullet (overlapping a green one) represents the algorithm's output.

Like for the SA simulation, an approximated value of the global minimum of function f was computed, even though the approximations have sometimes moved away from the desired minimal point, heading towards other local minimal. The output value of this algorithm was $x_6 \approx 1.6204$. Once again, the exact global minimum is $x \approx 1.6209$; therefore, the absolute error on this approximation is around 5×10^{-3}, which is smaller than that obtained by applying the SA method only.

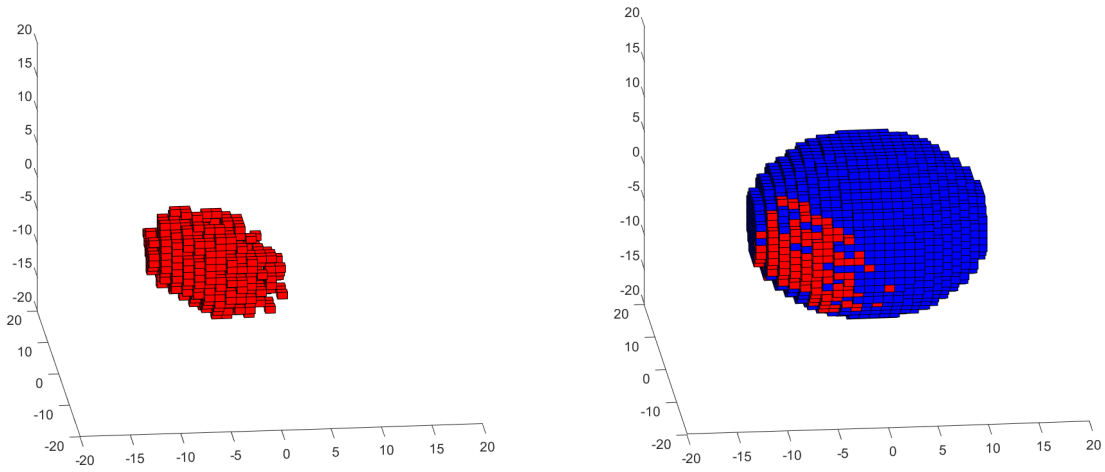
Chapter 5

Computational experiments

In this chapter, computational experiments is presented based on the heuristic methods described on Chapter 4. For this purpose, the formulation introduced in Chapters 2 and 3 is considered and adapted to an artificial case study. The different methods are tested and applied to finding sets of 1, 3, 5 and 7 angles for the MLC to deliver the radiation over the patient's body. The obtained results are compared in terms of the output objective function values and of their run times.

5.1 Case study

A phantom structure was built to support the computational experiments. This structure is composed by a phantom organ and a mechanical structure. The structure assumes an almost spherical shape, as illustrated in Figure 5.1. The considered organ is formed by red voxels, representing the malignant cells, and blue voxels, representing the non-cancerous cells.

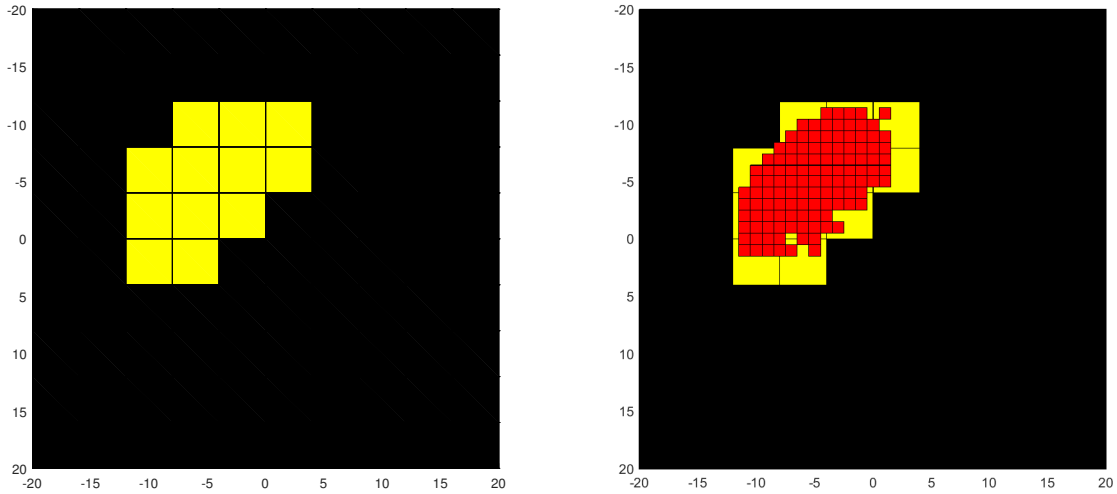


(a) Tumor

(b) Surrounding vital tissues

Fig. 5.1 Voxels distribution representing the phantom organ

The mechanical structure represents the radiation source through the MLC, which is discretized by square shaped bixels. The radiated beamlets are defined by the state of these bixels, which can be opened or closed. As shown in Figure 5.2a, the bixels in yellow represent an open area of the MLC, whereas the bixels in black represent a closed area of the MLC. In Figure 5.2b, it is exemplified how the open bixels fit with the red voxels.



(a) Structure of bixels representing the MLC

(b) Shaping of the MLC to the red voxels

Fig. 5.2 The MLC for $\theta = 0°$

For every possible angle in the continuous range $[0, 360[$, the projection calculations are made in order to find the bixels required to open and to achieve the optimal shape. Thus, for every angle from the considered set, every bixel is defined and the projection of its inner space is compared, step by step, with the geometrical center of the red voxels. If the projection of a certain bixel matches with the center of any tumor voxel, the bixel is activated, and the calculations proceed to the next bixel.

The dimensions assigned to the MLC enable the structure to radiate all the voxels, regardless of its position. When an angle of incidence θ is attributed to the MLC, an inverse rotation matrix is applied to the voxels [6, pp. 4-11], rotating the overall position of the human phantom structure by $-\theta$ degrees. The inverse rotation of the voxels by θ degrees has the same effect as rotating the MLC by θ degrees. However, keeping the bixel areas static makes the calculations easier, since the values of the bixel edges remain unchanged. Considering the voxel i and supposing that its position in the three-dimensional reference is $[px_i \; py_i \; pz_i]^T$, its position after the rotation is calculated by the

following matrix calculations:

$$\begin{bmatrix} px_i^* \\ py_i^* \\ pz_i^* \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}^{-1} \begin{bmatrix} px_i \\ py_i \\ pz_i \end{bmatrix}$$

$$= \begin{bmatrix} \cos(-\theta) & 0 & \sin(-\theta) \\ 0 & 1 & 0 \\ -\sin(-\theta) & 0 & \cos(-\theta) \end{bmatrix} \begin{bmatrix} px_i \\ py_i \\ pz_i \end{bmatrix}$$

for any $i \in V_h$, $h = 1, \ldots, S$. Once this process is completed, the MLC has finally defined which bixels should be open or closed, shaping the incidence on the tumor and obtaining an analogous output to the one shown in Figure 5.3.

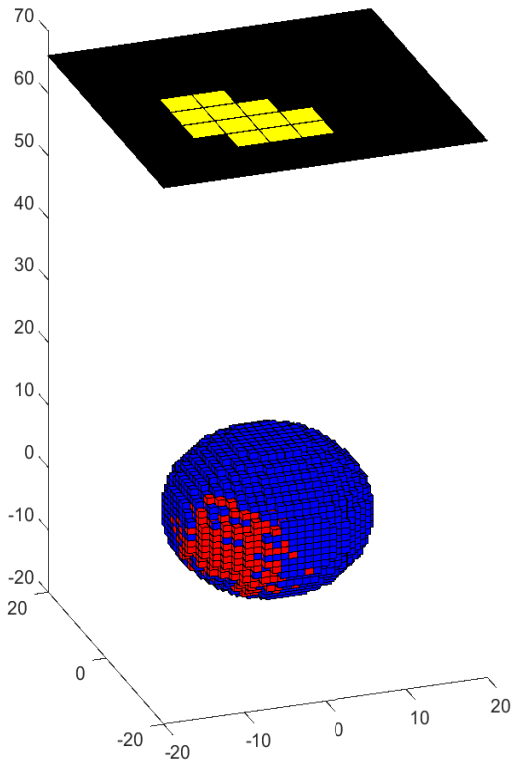


Fig. 5.3 Output display from MATLAB with $\theta = 0°$

For the current case study, combinations of $n = 1, 3, 5, 7$ incident angles and $|B_k| = 100$ beamlets per angle will be considered, $k = 1, \ldots, n$. The structure constructed is composed by two VOIs: a tumor structure (V_2) that contains $N_2 = 592$ voxels and a set of the normal surrounding tissue (V_1) that contains $N_1 = 6531$ voxels. The voxels and bixels side lengths are $l^V = 1$ cm and $l^b = 4$ cm

respectively. The weight of every voxel is considered to be $W = 1.07 \times 10^{-3}$ Kg. The distance between the MLC and the center of the voxel structure is 40 cm.

To simulate the BAO problem for this case study, the formulation presented in Chapter 2, adapted from formulation (2.3)–(2.14), together with constraint (3.9), was considered. Some simplifications were assumed. This concrete case study is composed of two structures and, therefore, the objective function term and the constraints involving $p = 3$ were not included. Furthermore, lower and upper bounds over the convex combination of the maximum (minimum) and the mean doses delivered to the voxels in O_1 (O_2) were also not considered in the tests. Additionally, also the variables z_i^h were not included. Thus, the used formulation is

$$
\begin{aligned}
\min \quad & f(\theta, x) = w_1^+ \left(\alpha_1 y_1^{\max} + \frac{1-\alpha_1}{N_1} \sum_{i \in V_1} d_i \right) - w_2^- \left(\alpha_2 y_2^{\min} + \frac{1-\alpha_2}{N_2} \sum_{i \in V_2} d_i \right) \\
\text{subject to } & d_i = \sum_{k=1}^{n} \sum_{j \in B_k} a_{ij}(\theta_k) x_j, \; i \in V_h, \; h = 1,2 \\
& LB_h \leq d_i \leq UB_h, \; i \in V_h \\
& x_j \geq 0, \; j \in B_k, \; k = 1, \ldots, n \\
& 0 \leq \theta_k \leq 360, \; k = 1, \ldots, n \\
& y_1^{\max} \geq d_i, \; i \in V_1 \\
& y_2^{\min} \leq d_i, \; i \in V_2.
\end{aligned}
\tag{5.1}
$$

It is important to recall that if voxel i is irradiated by beamlet j positioned in θ_k, then $a_{ij}(\theta_k)$ denotes the ratio of dose delivery to the voxel i by the beamlet j, from the beam k for any $i = 1, \ldots, m$, $j = 1, \ldots, p$, $k = 1, \ldots, n$. If that is not the case, then $a_{ij}(\theta_k)$ is 0. Taking this into account,

$$
a_{ij}(\theta_k) = \begin{cases} \frac{1.60217662 \times 10^{-13}}{16 \times 1.07 \times 10^{-3}} e^{-0.01 L_{ij}(\theta_k)}, & \text{if voxel } i \text{ is irradiated by beamlet } j \text{ from } \theta_k \\ 0, & \text{otherwise,} \end{cases}
$$

where $L_{ij}(\theta_k)$ represents the distance (in cm) traversed by the radiation through the body until it reaches the voxel i, for $i \in V_h$, $h = 1,2$, $j \in B_k$ and $k = 1, \ldots, n$.

Given this reformulation, which simplifies some medical constraints while including the absorbed energy attenuation, some other parameters need to be established. Namely, the minimum dose for each voxel in the tumor is set to 50 Gy (and 0 Gy for voxels in the normal tissue) and the maximum dose for each malignant voxel is set to 70 Gy. The importance of the vital organ for the objective function is given by $w_1^+ = 1$ and for the tumor is $w_2^- = -1$. The weighted convex combination of the maximum/minimum dose and the mean dose for vital/tumoral structures is $\alpha_1 = \alpha_2 = 0.25$. Finally, the absorption coefficient is considered to be $\mu = 0.01$ cm^{-1}. The parameters used in the problem are summarized in Table 5.1.

Having stated that, the heuristic methods are applied where the variable θ is dynamically explored and fixed. With this, for every iteration, the dose deposition matrix A is calculated, relatively to the fixed set of angles θ, making it a linear problem dependent on the vector x. Finally, x is an optimal solution calculated by means of an auxiliary minimization function.

Table 5.1 Formulation parameters

	$h=1$	$h=2$
α_h	0.25	0.25
w_h^+	1	—
w_h^-	—	-1
N_h	6 531	592
LB_h	0	50
UB_h	$+\infty$	70

Lastly, it is finally possible to apply the formulation to the phantom case. However, to obtain the experimental results, the heuristic methods have to be outlined, which will be done in the next section.

5.2 Heuristic methods for the IMRT

As mentioned earlier, beam angle optimization for IMRT treatments is known to be a very challenging problem: it is nonlinear, highly non-convex, and with a large number of local minima. These will hamper the approaches to the problem considered and, thus, some different methods and variations, based on Chapter 4, will be tested and compared.

Firstly, to show how challenging this problem can be, let us consider the objective function values in (5.1) for every angle multiple of 0.1º in [0º, 360º[, using the parameters previously defined. That is, the minimization problem is solved at $360/0.1$ angles. The remaining values are obtained by linear interpolation, giving rise to the plot in Figure 5.4. The image obtained reinforces the idea of an ill-behaved function, which makes the good performance of any method applied more difficult.

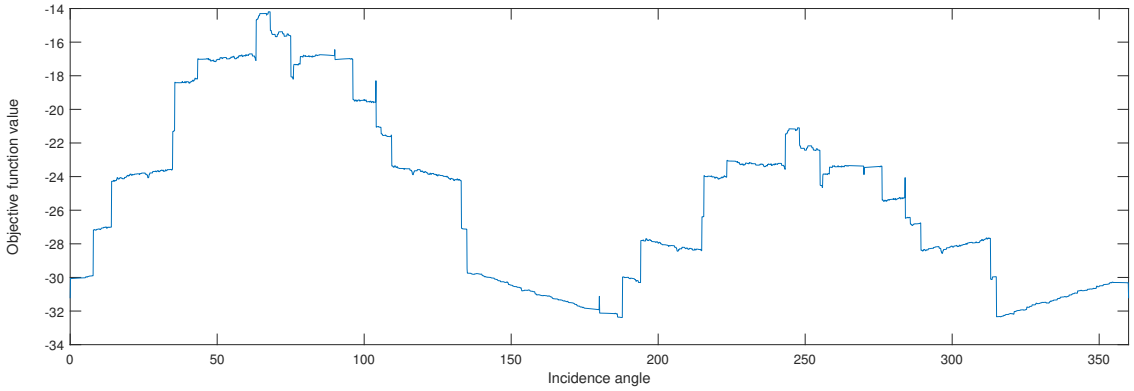


Fig. 5.4 Objective function values for incident angles between 0º and 360º with intervals of 0.1º

The methods start with a vector θ, composed by n randomly generated angles. The matrix A, associated with the vector θ, is arranged by projecting the beamlets through the voxels, activating the ones that focus on tumor voxels and calculating the dose deposition of these. Once the values of θ and of A are known, the linear problem defined by $f(x)$ is solved optimally using the package CPLEX

through a MATLAB auxiliary function, called `cplexlp`. Finally, the program returns a vector x which is optimal.

In these experiments, the different implementations were based on the methods described in Chapter 4. The heuristic methods are performed and, in every single iteration, the linear problem is solved using the updated vector of angles. Taking this into account, the following methods were implemented:

- Simulated Annealing (SA) - This method corresponds to Algorithm 1. The SA method was implemented using two probability distributions to generate the new set of angles based on the previous one, the normal distribution with mean 0 and a given standard deviation σ, with $\sigma = 45^\circ, 90^\circ$, or the uniform distribution in $(-r, r)$, with $r = 45^\circ, 90^\circ$.

- Adaptive Simulated Annealing (ASA) - This method has the same specifications as the SA method. However, the generation probability density function, which defines the range of new random values, decreases with the number of iterations, that is, as the temperature T decreases. In particular, if the normal distribution is considered, the standard deviation is updated as $\sigma \times T$. Otherwise, if the uniform distribution is considered, the interval is defined by $r \times T$.

- Hybrid Method (HM) - This method corresponds to Algorithm 3 and is based on the combination of a SA procedure and a SD method. This particular version of the HM runs the SD method with a time limit of 10 seconds and the SA method with time limit of 20 seconds, alternately. The SD method considered runs with an initial decrease step of $\alpha = 2$. The gradient value is calculated by the finite difference method, with $\gamma = 10^{-4}$. The CC from Wolfe conditions are not verified for the SD method, due to the highly non-convexity of the objective function. However, the SDC was considered with $c_1 = 0.1$.

- Adaptive Hybrid Method (AHM) - This method consists on a combination of the ASA and the HM methods. It performs the iterative process of HM, while the angles interval is adjusted as in the ASA.

- Simulated Annealing with final Steepest Descendent (SASD) - This method performs the SA procedure normally, finishing with the application of the SD method for at most 20 seconds and with the same parameters the SD method used in the HM.

- Adaptive Simulated Annealing with final Steepest Descendent (ASASD) - This method performs the ASA method until the end, finishing with one iteration of the SD method for 20 seconds.

The computational tests are divided into three groups and will be described in detail in the following section. The first set of experiments simulates the manual and intuitive choice of angles made by physicians. The second set of experiments focuses on the assessment of the SA and the ASA methods. The most reliable methods will be transposed to the third set of experiments and compared with methods that apply the SD method.

5.3 Computational results

The computational tests presented in this section were run on an Intel® Core™ i5-3230M, 2.6 GHz processor, 8 Gb of RAM. The codes were implemented in Matlab R2018a, which used CPLEX 12.8.0 as the linear program solver.[1]

Experiments I The first part of this section aims to recreate the trial and error approach used by technicians. This 'traditional' process may present some vulnerabilities, since it is entirely dependent on human experience and intuition. The obtained results are used for comparison with the results achieved by the heuristic methods. The simulation started by considering the objective function values, using the parameters in Table 5.1 for every integer angle in $[0^\circ, 360^\circ[$. After obtaining the 360 results for $n = 1$, the 5 angles with the best objective function values are chosen and equidistant angles are added for the cases $n = 3, 5, 7$. The best results achieved for the different sets of angles considered are presented in Table 5.2.

Table 5.2 Results for Experiments I

Number of angles	Angles	Objective value
1	187º	-32.3716
3	62º, 182º, 302º	-35.2831
5	38º, 110º, 182º, 254º, 326º	-36.2241
7	33º, 84º, 136º, 187º, 238º, 290º, 341º	-36.2532

Although it is impossible to simulate, with such a simplistic method, the whole experience of physicians, the results achieved can still be used as a reference for the analysis of the following computational results. The most important conclusion that we can draw from these reference values is that the number of angles increases as the objective function values decrease. Furthermore, it is also noticeable that the decrease seems to be slower as the number of angles increases. Lastly, this process ran in a overall time of 29.250 seconds.

Experiments II For the second set of experiments, the following conditions were considered:

- For each instance, 30 runs of each code were considered.
- Two different heuristic methods were applied: the SA and the ASA.
- The initial temperature was set to $T_0 = 5$, with cooling steps given by $T_{k+1} = T_k - 0.05$, and the stopping criterion is defined by $T_{\min} = 0$.
- For these particular experiments, the lines 13 to 15 of Algorithm 1 were not considered. This means that, for a fixed number of iterations, the position of the last iteration of each method is clearer and the selection of the most advantageous convergence cases is easier.

[1] The codes developed to support the computational experiments are available at the url https://www.mat.uc.pt/~mat12102/Codigo_IMRT.

Both the average in the objective function values and the average run times (in seconds) are presented in Table 5.3. In this table, and in the following, each code is identified by a sequence of letters (which stand for the implemented method), the letter `N` or `U`, for the used probability distribution (the first for normal and the second for uniform), and the standard deviation or the limit of the interval, for each of the two previous cases, respectively. The best 4 results for each set of angles in Table 5.3 are highlighted in green, where, the darker the shade is, the more accurate is the result or the faster is the method.

Considering Table 5.3, it is possible to observe a similar behavior as in Table 5.2, with the average results improving as the number of angles increases. This was already expected, since the increase of n allows for a more balanced radiation distribution along the body, avoiding a more intensive incidence of radiation on specific non-tumoral organ tissues. As the value of n increases in Table 5.3, gradual improvements of 11.91%, 0.26% and 0.10% are verified, respectively, which reinforce the decay in the improvement of the objective values as n increases, as initially observed in Table 5.2.

Table 5.3 Average results for Experiments II

	Objective function				CPU times (seconds)			
	$n=1$	$n=3$	$n=5$	$n=7$	$n=1$	$n=3$	$n=5$	$n=7$
SA_N90	-32.2575	-36.1597	-36.2382	-36.2668	68.610	204.855	355.298	494.307
SA_N45	-32.3001	-36.1671	-36.2369	-36.2672	70.816	231.412	376.222	493.927
SA_U90	-32.2884	-36.1499	-36.2236	-36.2686	67.268	207.033	366.575	487.471
SA_U45	-32.3323	-36.1259	-36.2360	-36.2792	65.870	205.344	362.955	495.313
ASA_N90	-32.3528	-36.1867	-36.2372	-36.2932	67.009	201.357	353.643	492.775
ASA_N45	-32.3624	-36.1853	-36.2591	-36.2544	67.079	200.719	360.110	498.526
ASA_U90	-32.3574	-36.1110	-36.2519	-36.2979	66.483	206.502	368.280	495.738
ASA_U45	-32.3425	-36.0474	-36.2042	-36.2463	66.649	206.047	363.984	496.002

Observing the values of Table 5.3 in greater depth, it is possible to note that, in most cases, the ASA methods lead to better objective function values when compared to the SA methods with the same parameters. In addition, there is a predominance of better results when the normal distribution is applied, rather than with the uniform distribution. Finally, it is important to highlight the consistency of the methods, shown by the small standard deviation values, between 0.0213 and 0.3372.

We now compare the results of the 'traditional' method and the best average results obtained with the different SA methods. For $n=1$, a decay of -0.03% was registered, which means that, for a single angle, the 'traditional' case not only obtained a better result but also did it faster. Yet, for $n=3,5,7$, improvements of 2.56%, 0.10% and 0.12% were achieved with the heuristic methods, respectively. Still comparing the Tables 5.2 and 5.3, it is possible to conclude that using a single angle only pays off if it is important to solve the problem quickly, since it provides very poor objective function values in comparison to other numbers of angles. Moreover, in that case it is preferable to choose the angle using the 'traditional' method. Analogously, considering 7 angles yields a very small improvement of the objective function (less than 0.01%) and, on top of that, the run time increases by 36.51%. Therefore, using 7 angles would only make sense if the calculation and treatment times were unimportant.

Table 5.3 also reveals that the run time increases as the number of angles increases. This is mainly due to the increase of the number of columns of the dose deposition matrix A as the number of angles increases, which leads to an increase in the time needed to solve the linear program as well.

Based on the previous conclusions, the third set of computational tests considers $n = 3,5$ angles, which have been proven to be the two sets with the best balance between objective function values and run times. Besides this, and taking Figure 5.5 into account, the four methods selected for the next round of tests are: `SA_N90`, `ASA_N90`, `ASA_N45` and `ASA_U90`.

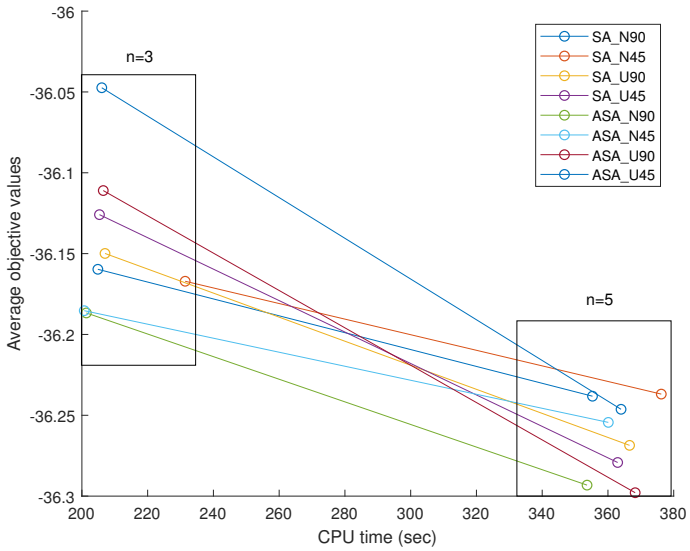


Fig. 5.5 Average results for Experiments II

Experiments III In the final phase of these computational tests, the four SA selected codes are tested together with their variants SASD, ASASD, HM and AHM. These will result in a comparison of a total of 12 codes, fulfilling the following conditions:

- 30 runs are performed for each method.
- Each trial for the various methods was performed for 5 minutes (300 seconds), in order to impose a time restriction to find the solution. This restriction, not only matches all methods at run time level, but can also simulate a case where planning time is limited.
- The initial temperature is set to $T_{\max} = 5$, with cooling steps of $T_{int} = \frac{199}{200}$, with $T = T \times T_{int}$ as the temperature updated function.
- SA includes a variable that stores the minimum value of each run, as described in lines 13 to 15 of Algorithm 1. Given this, it is possible to evaluate the best value obtained for each trial.

According to these conditions, once the methods are performed, the average results and the best result of each set of 30 trials are summarized in Table 5.4. Again, the best 4 results for each set of angles are highlighted in different shades of green.

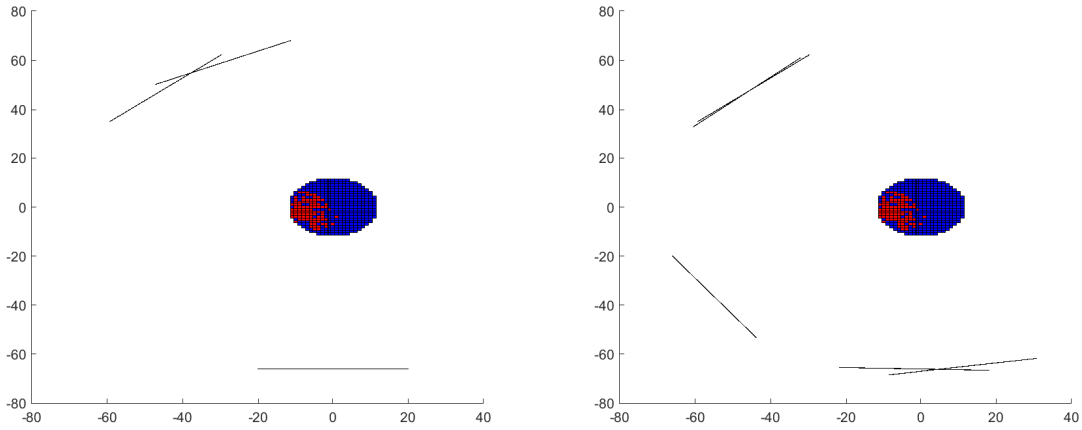
Table 5.4 Average and optimal results for Experiments III

	Average results		Best results	
	$n = 3$	$n = 5$	$n = 3$	$n = 5$
SA_N90	-36.2177	-36.2821	-36.2889	-36.3493
ASA_N90	-36.2484	-36.2949	-36.3289	-36.3430
ASA_N45	-36.2409	-36.3026	-36.3291	-36.3424
ASA_U90	-36.2385	-36.2960	-36.3286	-36.3455
SASD_N90	-36.2210	-36.2870	-36.2786	-36.3490
ASASD_N90	-36.2510	-36.2976	-36.3260	-36.3567
ASASD_N45	-36.2577	-36.3200	-36.3287	-36.4164
ASASD_U90	-36.2506	-36.3033	-36.3270	-36.3846
HM_N90	-36.2135	-36.2757	-36.2663	-36.3379
AHM_N90	-36.2491	-36.2720	-36.3207	-36.3155
AHM_N45	-36.2464	-36.2778	-36.3190	-36.3239
AHM_U90	-36.2377	-36.2766	-36.2809	-36.3253

Taking into account the results in Table 5.4, the event that stands out the most is the high prevalence of green markers in the methods that implement the final SD. In fact, taking the average results into account, for $n = 3$, the final SD methods presented improvements between 0.007% and 0.046% regarding simple SA methods. Nonetheless, the hybrid methods revealed very poor average results that did not stand out from the SA methods. As a matter of fact, these methods presented a performance between -0.002% and 0.015% when compared with the SA methods with the same parameters and probabilistic distribution. Additionally, the average results for $n = 5$ exhibit very identical improvements for $n = 3$, concerning the same methods. While methods with a final application of the SD method showed improvements between 0.007% and 0.048%, the hybrid method presented a worsening of results ranging from -0.018% to -0.082% relative to the respective SA methods. It is also important to enhance the consistency of the samples average results, with standard deviations between 0.0185 and 0.0398.

With respect to the best result found for each instance, it is possible to note that these are distributed between the first two thirds of the table. This remark differs from the average results observations, given that the best average results are concentrated in the second set of methods, due to the strong influence of the stochastic pattern present in the methods. Finally, it should be noted that the best result for $n = 3$ was achieved by code `ASA_N45`, while the best result for $n = 5$ was achieved by code `ASASD_N45`. The solutions found in those cases are illustrated in Figure 5.6.

Based on Table 5.4 and on the plots in Figure 5.7, it is possible to conclude that the methods that implement ASA with final SD method were those that stood out from the set of all the tested methods. In addition, it is also possible to conclude that, out of these 3 methods, `ASASD_N45` was the one that presented, in general, the best results and, therefore, this would be the most suitable to look for an optimal solution in the considered spectrum of parameters. Regarding the number of angles, the combination of 5 angles seems to be the best hypothesis. However, the application of this number of angles implies a longer treatment time and, for this reason, this decision will depend on the specifications of a particular clinical case.



(a) $n = 3$ $(180.01°, 317.51°, 333.57°)$ (b) $n = 5$ $(170.31°, 181.62°, 236.28°, 315.44°, 317.53°)$

Fig. 5.6 Solution with the best objective value

In the literature, the hybrid method has been shown to give better results than using the SA method only, for some cases [3]. Here the performance of the SD method seems to have compromised the application of the SA method due to the short time limit imposed to the combination of the two methods (300s). While the final SD method ran only for 20 seconds, out of the 300 seconds (6.66% of the total time), the HM could ran for 100 seconds (33.33%), which may have compromised the application of the SA method. In principle, by extending the available time window the HM could bring more promising results.

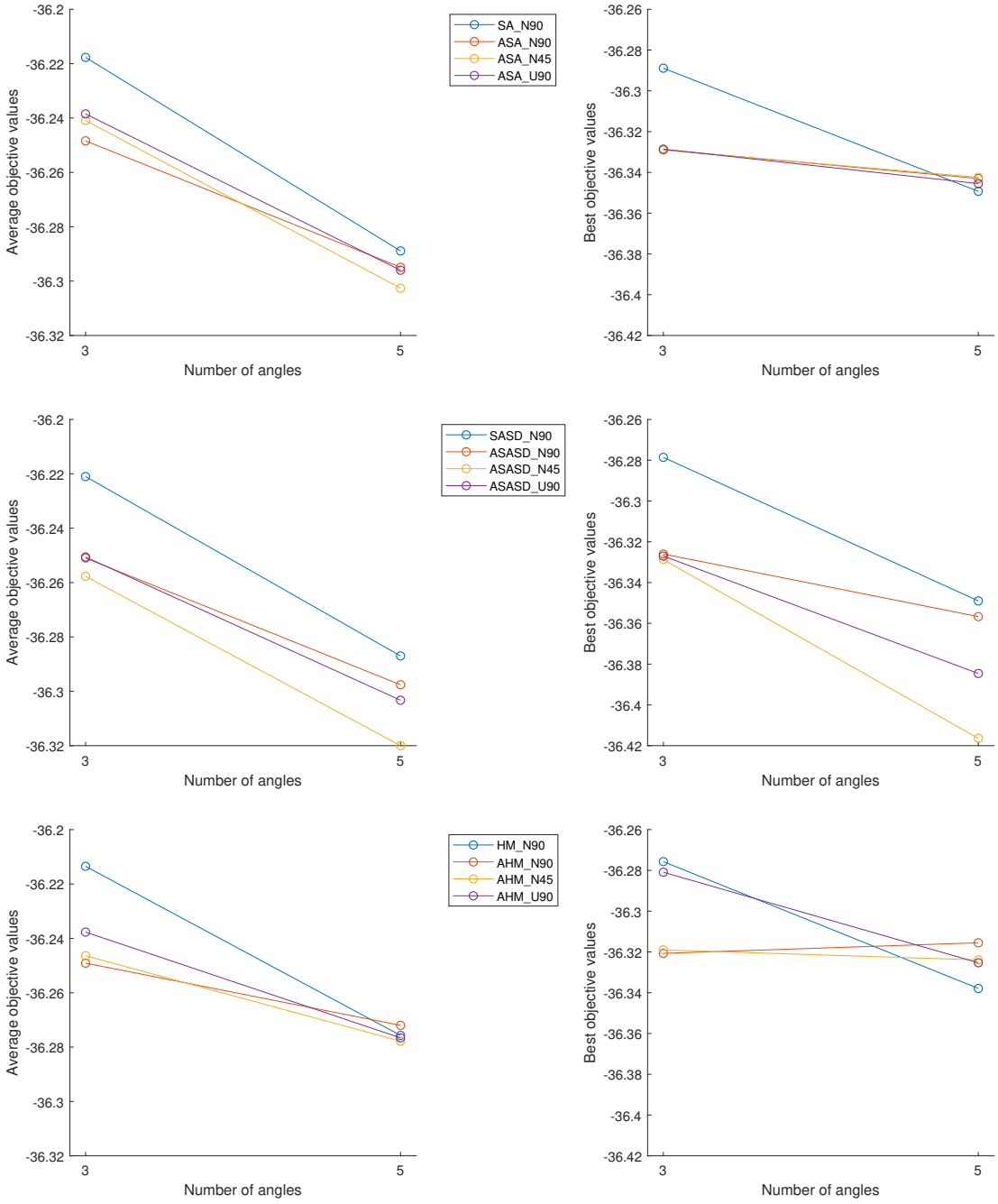


Fig. 5.7 Average and best results for Experiments III

Chapter 6

Concluding remarks

In this thesis, the beam angle optimization problem was studied. The goal of the BAO problem is to find a set of angles that is able to deliver radiation to cancer structures, complemented with the quantity of energy emitted by the beam for building a treatment plan which is as effective as possible. The aim is bifold, as it consists of sparing the vital structures, while reaching the malignant cells with a dose as close as possible to the prescribed dose. Furthermore, the dose of radiation delivered to each voxel does not depend on the incidence angle of the beam over the organs alone: it is also affected by the attenuation of the beam as it passes through the tissue. This is a difficult continuous optimization problem since it is highly non-convex and with many local minima. Considering this, the thesis was divided in two phases: firstly, a nonlinear formulation to the problem was introduced, rewarding the incidence of radiation through the tumor structure and penalizing the infringement of vital medical requirements; secondly, heuristic methods were described to find solutions for the formulation. These two phases allowed the intensities of the beams to be optimized in an automated way, for every combination of angles dynamically explored in the continuous space, by the search methods.

The studied methods were the simulated annealing method, the steepest descent method and a combination of the first two that alternates them. For the computational experiments, a phantom case was considered and the discretization of the structure into voxels and the discretization of the MLC into bixels, as well as the alignment of the beamlets with the malignant voxels, played a crucial role on the calculations. The functionality of the remaining code was carefully designed in order to keep it simple and time profitable. This was one of the main concerns in the construction of the code, since both the development of the dose deposition matrix, which tends to have large dimensions, and the resolution of the linear problem face a problem of execution time, which is reflected more expressively during the heuristic methods, where these processes are iteratively invoked a significant number of times. Several variants of the studied methods were coded in MATLAB. The linear programs presented in the codes were solved using the software CPLEX.

The implemented codes were tested for three different sets of experiments. The adaptive simulated annealing was the method that, throughout computational tests, presented the most promising results. In the last set of results, it was also possible to conclude that the adaptive simulated annealing with final steepest descent method (ASASD) with points random generated following a narrower normal distribution was the best method to obtain good solutions to the BAO problem quickly, for the case

considered. However, it is important to recall that several factors can be changed in order to find better results. For instance, the run time of each method, increasing the number of angles, or even a more refined description of the voxels.

To conclude it is important to emphasize that the implemented codes are ready to include other representations for more structures of different types besides the studied phantom case. Additionally, it can be stressed that these optimization and computational methods are able to provide solutions that improve the planning of radiotherapy treatments.

References

[1] Aarts, E. and Korst, J. (1988). *Simulated annealing and Boltzmann machines*. John Wiley and Sons Inc., New York, NY.

[2] Bahr, G., Kereiakes, J., Horwitz, H., Finney, R., Galvin, J., and Goode, K. (1968). The method of linear programming applied to radiation treatment planning. *Radiology*, 91(4):686–693.

[3] Bertsimas, D., Cacchiani, V., Craft, D., and Nohadani, O. (2013). A hybrid approach to beam angle optimization in intensity-modulated radiation therapy. *Computers & Operations Research*, 40(9):2187–2197.

[4] Bortfeld, T., Bürkelbach, J., Boesecke, R., and Schlegel, W. (1990). Methods of image reconstruction from projections applied to conformation radiotherapy. *Physics in Medicine & Biology*, 35(10):1423.

[5] Chang, R. and Goldsby, K. A. (2013). *Principles of Chemistry*. Langara College.

[6] Chirgwin, B. H. and Plumpton, C. (2013). *Advanced Theoretical Mechanics: A Course of Mathematics for Engineers and Scientists*, volume 6. Elsevier.

[7] Dias, J., Rocha, H., Ferreira, B., and do Carmo Lopes, M. (2015). Simulated annealing applied to imrt beam angle optimization: A computational study. *Physica Medica: European Journal of Medical Physics*, 31(7):747–756.

[8] Ehrgott, M., Güler, Ç., Hamacher, H. W., and Shao, L. (2010). Mathematical optimization in intensity modulated radiation therapy. *Annals of Operations Research*, 175(1):309–365.

[9] Goldstein, H., Wilkins Jr, J. E., and Spencer, L. (1953). Systematic calculations of gamma-ray penetration. *Physical Review*, 89(5):1150.

[10] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

[11] Langer, M. and Leong, J. (1987). Optimization of beam weights under dose-volume restrictions. *International Journal of Radiation Oncology* Biology* Physics*, 13(8):1255–1260.

[12] Li, Y. and Lei, J. (2010). A feasible solution to the beam-angle-optimization problem in radiotherapy planning with a DNA-based genetic algorithm. *IEEE Transactions on Biomedical Engineering*, 57(3):499–508.

[13] Mageras, G. and Mohan, R. (1993). Application of fast simulated annealing to optimization of conformal radiation treatments. *Medical physics*, 20(3):639–647.

[14] Meredith, W. J. and Massey, J. B. (2013). *Fundamental physics of radiology*. Butterworth-Heinemann.

[15] Morrill, S. M., Lane, R. G., Wong, J. A., and Rosen, I. I. (1991). Dose-volume considerations with linear programming optimization. *Medical Physics*, 18(6):1201–1210.

[16] Rosen, I. I., Lane, R. G., Morrill, S. M., and Belli, J. A. (1991). Treatment plan optimization using linear programming. *Medical Physics*, 18(2):141–152.

[17] Shepard, D. M., Ferris, M. C., Olivera, G. H., and Mackie, T. R. (1999). Optimizing the delivery of radiation therapy to cancer patients. *Siam Review*, 41(4):721–744.

[18] Taylor, L. S. (1960). Report of the international commission on radiological units and measurements (ICRU).

[19] Vicente, L. N. (2015). *Aulas de Optimização Numérica*. Departamento de matemática da Universidade de Coimbra.

[20] Webb, S. (1989). Optimisation of conformal radiotherapy dose distribution by simulated annealing. *Physics in Medicine and Biology*, 34(10):1349.

[21] Wong, K. P. and Wong, Y. W. (1995). Thermal generator scheduling using hybrid genetic/simulated-annealing approach. *IEE Proceedings-Generation, Transmission and Distribution*, 142(4):372–380.

[22] Xing, L. and Chen, G. T. (1996). Iterative methods for inverse treatment planning. *Physics in Medicine & Biology*, 41(10):2107.

[23] Yang, R. (2000). Convergence of the simulated annealing algorithm for continuous global optimization. *Journal of optimization theory and applications*, 104(3):691–716.

[24] Yu, C. X. (1995). Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Physics in Medicine & Biology*, 40(9):1435.

[25] Zhang, Y. and Merritt, M. (2006). Fluence map optimization in imrt cancer treatment planning and a geometric approach. In *Multiscale Optimization Methods and Applications*, pages 205–227. Springer.