UNIVERSITÄT ZU LÜBECK

An Interactive Approach to Localised Dose Planning for Radiotherapy

Tiberiu Viuleţ

Dissertation

Universität zu Lübeck
Institut für Robotik und Kognitive Systeme

IM FOCUS DAS LEBEN

Aus dem Institut für Robotik und Kognitive Systeme
der Universität zu Lübeck

Direktor:
Prof. Dr.-Ing. Achim Schweikard

An Interactive Approach to Localised Dose Planning for Radiotherapy

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck
— Aus der Sektion Informatik —

Vorgelegt von Tiberiu Viuleţ
geboren in Bukarest, Rumänien

Lübeck 2015

Tiberiu Viuleţ
Institut für Robotik and Kognitive Systeme
Universität zu Lübeck
Ratzeburger Allee 160
23562 Lübeck, Germany

Prüfungskommission

Vorsitzender	Prof. Dr.-Ing. Erik Maehle
1. Berichterstatter	Prof. Dr.-Ing. Alexander Schlaefer
2. Berichterstatter	Prof. Dr. rer. nat. Heinz Handels

Tag der mündlichen Prüfung: 30.11.2015

Zum Druck genehmigt. Lübeck, den 02.01.2016

Zusammenfassung

Einleitung In der Strahlentherapieplanung werden Ziele verfolgt, die zueinander in Konflikt stehen: einerseits ist eine hohe Dosis nötig, um Krebszellen zu töten; anderseits muss gesundes Gewebe so gut wie möglich vor schädigenden Nebenwirkungen der Strahlung geschont werden. Dementsprechend wird während der Planung eine komplexe dreidimensionale Dosisverteilung gestaltet, die durch die Tumor- und Organengeometrie eingeschränkt ist. Typischerweise werden mehrere mathematische Kriterien auf den jeweiligen Volumina definiert, die der Planer in einem iterativen Prozess mittels mehrerer Statistiken gegeneinander abwägt. Jedoch blenden sowohl die Kriterien als auch die Statistiken die örtliche Information aus: der Benutzer kann weder spezifizieren wo genau in einem Volumen Dosisänderung stattfinden soll, noch welchen Effekt eine lokale Änderung auf eine Statistik hat. Der Fokus dieser Arbeit war es, ein Planungstool (*Interactive3D*) zu entwickeln, das das Problem angeht, indem der Benutzer die Dosisverteilung lokal und präzise manipulieren kann. Dadurch können diverse lokale Dosis-Trade-offs analysisert werden, was wiederum neue Möglichkeiten zur Planung eröffnet. Ein Beispiel eines lokalen Trade-offs wäre eine Region des Tumors vorzugsweise mit höherer Dosis abzudecken, wobei auch eine höhere Dosierung einer weniger sensitiven Region eines Organs hingenommen wird.

Methodik Zunächst wurde ein existierendes mathematisches Modell erweitert, um auch lokale Dosis-Kriterien neben den konventionellen Kriterien zu optimieren. Für die Planung wurde daher ein Top-down Ansatz gewählt: eine initiale Lösung wird durch ein konventionelles Verfahren ermittelt und anschließend durch das Interactive3D Tool weiter verfeinert.

Als nächstes wurde eine Benutzeroberfläche zur Manipulation des mathemathischen Modells erstellt. Um das Optimierungsproblem zu visualisieren werden sowohl die Formen der aktuellen als auch die der angestrebten Dosisverteilung, sowie die einschränkenden Strukturen dargestellt. Diese einschränkenden Strukturen bestehen aus den unteren und oberen Dosisgrenzen, die für jedes Volumen definiert wurden. Sie sorgen dafür, dass die Dosis sich nicht über den von ihnen begrenzten Raum hinaus verteilt. Der Optimie-

rungsprozess wird durch die Manipulierung der graphischen Objekte gesteuert. Um das Optimierungsziel zu definieren, muss der Benutzer die Form der aktuellen Dosis zur gewünschten Zieldosis deformieren. Da das Ziel üblicherweise aufgrund physikalischer Einschränkungen nicht erreicht werden kann, ist der Planer gezwungen ein oder mehrere Trade-offs durchzuführen. Um zu spezifizieren wo und wie stark Dosis-Grenzen überschritten werden dürfen, muss der Benutzer die einschränkenden Strukturen anpassen, sodass sich die Dosis weiter ausbreiten kann.

Nach jedem Interaktionsschritt folgt unmittelbar ein Optimierungsschritt, dessen Laufzeit von der Größe des Optimierungsproblems abhängt. Mit Hinblick auf die Benutzbarkeit wird das Optimierungsproblem so klein wie möglich gehalten. Zum einen wird die Anatomie des Patienten mit einem adaptiven Verfahren diskretisiert, sodass nur die kritischen Bereiche in höherer Auflösung repräsentiert werden; zum anderen wird die Anzahl der Strahlen für Optimierung limitiert, um den Suchraum zu begrenzen. Die Qualität der Lösung wird nachträglich durch einen automatischen Algorithmus verbessert, indem zufällig weitere Strahlen generiert werden.

Ergebnisse Die Methoden dieser Arbeit wurden auf realen und auf synthetischen Patientendaten evaluiert. In einem ersten Experiment wurde ein vereinfachter Fall anhand konventioneller Verfahren und durch Interactive3D geplant. Obwohl die Interactive3D Lösung von den anderen Lösungen im Pareto Raum der nicht-lokalisierten Kriterien dominiert wurde, konnte die Überdosierung in einem weniger sensitiven Areal eingegrenzt werden.

Als nächstes wurden mehrere Szenarien von realen Fällen mit Interactive3D geplant, um das Potential der lokalen Dosismanipulierung aufzuzeigen. In einem Szenario wurde ein besonders empfindlicher Bereich eines Organs geschont, indem Dosis in Richtung eines anderen, weniger sensitiven Organs verlagert wurde; trotzdem hat sich die Abdeckung des Tumors nur minimal verschlechtert. In einem weiteren Szenario wurde ein Teilvolumen des Tumors, welches ein höheres Rezidivrisiko aufwies, vorzugsweise bestrahlt. Dafür wurde sowohl eine geringere Dosishomogenität im Tumorzentrum, als auch eine überhöhte Dosis in den proximalen Organen toleriert.

Weitere Experimente hatten die Reduzierung des Optimierungsproblems als Ziel. Wie erwartet führten die gröberen Diskretisierungen der Anatomie und des Strahlenraumes zu Lösungen mit geringerer Qualität. Allerdings wurde gezeigt, dass Hotspot-Fehler durch die adaptive Anpassung der Dosisgrenzen minimiert werden konnten. Zudem konnte die Qualität der Lösung nach der interaktiven Planung durch die Generierung von zusätzlichen Strahlen teilweise wiederhergestellt werden.

Schlussfolgerungen Das Interaktive3D Tool ermöglicht die direkte und präzise Kontrolle über die Form der Dosisverteilung. Dadurch können lokale Kriterien gegeneinander abgewogen werden, um akzeptablere Trade-offs zu erreichen. Obwohl die interaktive Planung ein komplexes Verfahren darstellt, kann sie in Kombination mit etablierten Ansätzen das Endergebnis der Strahlentherapieplanung verbessern.

Abstract

Introduction Radiotherapy treatment planning is a process of balancing conflicting goals: on the one hand, a high enough dose is needed to kill cancer cells; on the other hand, exposure of healthy tissue needs to be kept to a minimum. Consequently, planning boils down to shaping the three-dimensional dose distribution to an intricate shape, as constrained by geometries of the tumour and of the anatomy. Typically, a number of mathematical criteria are defined on the entire volumes of interest, which the human planner weighs them against each other in an iterative optimisation process. At each step, statistics are used to quickly evaluate solutions. However, as both statistics and criteria obfuscate locality information, the planner cannot specify where within the volumes dose change should happen. The focus of this work was to address this issue by developing a planning tool, the *Interactive3D*, that allows the user to locally and precisely manipulate the dose shape. This opens up the new possibility of studying various local dose trade-offs. For example, such a trade-off would be to preferentially cover a region of the tumour by accepting a higher dose exposure to a less sensitive part of an organ at risk.

Methods First, an existing mathematical model was extended to allow for the optimisation of local dose shape criteria along side conventional criteria. As such, a top-down approach to planning was established: an initial solution of adequate quality is found by established means; afterwards, this solution is further refined using the interactive planning tool.

Next, a graphical user interface was built on top of the mathematical model. The optimisation problem is visualised, by displaying the current and target dose shapes, as well as the limiting structures. These limiting structures consist of the upper and lower dose bounds defined for each of the volumes of interest and prevent dose from either shrinking into, or extending beyond them. The user interactions, guiding the planning process, consist of manipulating these virtual structures. Thus, to set an optimisation target, the user deforms the current dose shape to a desired target dose shape. Typically, due to physical constraints, the target dose is not be achievable and the planner is forced

to perform a series trade-offs. To specify where and by what amount the original bounds may be violated, the user can sculpt out the limiting structures, allowing the dose to evolve further in those areas.

Each interaction step is followed by a reoptimisation step. To improve usability, computation time was kept as low as possible by limiting the size of the optimisation problem. The patient anatomy is sparsely sampled, while only the critical areas are represented with a high resolution. Furthermore, only a small number of beams are used during optimisation, keeping the search space small. Solution quality is re-established posthoc, by automatically resampling the beamset, after interactive planning is finished.

Results The methods presented were evaluated on synthetic and real patient data. In a first experiment, a simplified case was planned by conventional tools and by Interactive3D. While the Interative3D solution was shown to be dominated by the other solutions in the Pareto space of the non-localised goals, it contained overdosage to a less sensitive area; this was not possible by conventional means.

Next, the potential of Interactive3D was shown in a number of scenarios in which multiple local dose trade-offs were studied. In one scenario, a highly sensitive region of an organ was spared, by shifting dose towards another, less sensitive organ; tumour coverage was only slightly affected. In another scenario, a subvolume of the original tumour was preferentially covered with dose, as it presented a high risk of recurrence. To this end, dose homogeneity deep within the tumour was sacrificed while adjacent organs were also exposed to slightly more dose.

Further experiments aimed at limiting the size of the optimisation problem. As expected, reduced anatomical and beam space resolutions had a negative impact on the solution quality. However, it was shown that hotspots appearing due to coarse volume sampling could be compensated by adapting the dose bounds. Furthermore, solution quality could be recovered after interactive planning, by resampling the optimisation beamset.

Conclusions The Interactive3D tool gives the planner direct and precise control over the shape of the dose distribution. This opens up new planning possibilities, as dose regions can be balanced against each other in order to achieve new, potentially more acceptable trade-offs. While Interactive3D also adds a layer of complexity, it can be used alongside conventional tools to improve the planning outcome.

Contents

1 Introduction

Medical advances have pushed life expectancy ever further and, with it, cancer has emerged as one of the leading causes of death worldwide [1], with one in two adults expected to be diagnosed with cancer during the course of their life [2]. Scientific advances in therapy options, such as surgery, radiation therapy, chemotherapy, or immunotherapy, have improved survival rates for some cancer types, while for others, such as pancreatic cancer, patient survival has hardly changed [3]. Of all treatment options, radiation therapy (or radiotherapy) is one of the most important, with about $50\,\%$ of the patients receiving some kind of radiation treatment during their illness [4]. Radiation therapy uses high doses of ionising radiation to damage tumour cells [5, p. 16]. Dose is usually delivered by multiple beams originating outside the patient body and converging on the tumour site. As beams pass through the body, they invariably also affect healthy tissue. Because beams converge from multiple directions, the dose burden is spread out on a much larger area, making radiation therapy tolerable [6, p. 1–2]. Through powerful physical models and simulations, the three-dimensional dose distribution can be accurately computed and, based on it, the effects of the therapy can be estimated [7–9].

Treatment planning is the process of determining a beam configuration that generates dose distribution which maximises tumour coverage (i.e. how much of the tumour is covered by dose) while minimising dose exposure of healthy tissue [10, p. 21]. This is a tall order due to the fact that tumours are usually situated in the proximity of many sensitive structures and, consequently, there is a large number of contradicting criteria to be taken into account. The obvious approach is to model this as an optimisation problem for which the human planner still has to direct the execution of the solver [11, 12]. Classical software tools simplify planning by presenting the user with a statistical view of the dose distribution corresponding to the current solution – typically, scalar metrics (percentage of organ covered by dose, minimum, mean and maximum doses, etc.) or one-dimensional dose histograms [10, 13, 14]. Based on these, the optimisation parameters are tweaked until a satisfactory solution is found, in what is known as the *human iteration loop* [15, 16]. For example, the importance weight of one of planning goals can be increased relative to the others and a subsequent optimisation step finds a new solution

for which the dose statistics are recomputed [17]. The planner then decides whether the new plan is superior to the old on or not. While such tools are very effective in finding a broadly satisfactory solution, tweaking it is time costly, as the planner has no direct control over the shape of the dose distribution.

1.1 Motivation and Purpose of this Work

Typical planning criteria are either completely non-localised (e.g. dose homogeneity, tumour control probability, etc.) or grossly localised in space (defined on entire volumes or on volume fractions) [12]. Furthermore, the statistics used to assess the quality of the plan hide all locality information, as the spatial dose distribution is flattened to a single scalar value or one-dimensional graph [18, 19]. The human planner, tasked with modelling the three-dimensional dose distribution to very intricate shapes, can do so only in an indirect manner, without direct local control. Consequently, the following planning questions cannot be readily addressed:

1. Can dose be improved in a specific area (such reducing dose to a very sensitive region of an organ)?

2. Does compromising dose within a specific location improve a certain planning goal (such as allowing more dose through a less sensitive region to improve overall target coverage)?

This problem has been recognised in a number of publications [20–22] and, while optimisation models that incorporate local information have been presented by a number of authors [17, 22, 23], to our knowledge, no complete planning framework has been proposed that offers optimised local manipulation of the dose distribution.

The purpose of this work is to develop such a tool, which allows the user to interactively form the three-dimensional dose distribution, and to embed it into our existing planning system. We call this tool *Interactive3D* (I3D), and in its development we tackle a number of (from our perspective) scientifically relevant topics:

1. Complete graphical representation of the optimisation problem: optimisation goal, current solution, and limiting structures.

2. Tweaking the optimisation parameters by natural interactions with the graphical elements. This would allow the user to set the optimisation goal, or to perform constraint relaxations, by interacting with the virtual scene.

3. Fast reoptimisation for rapid feedback to the user.

Interactive3D is intended to augment the existing planning workflow. As such, we position this tool as an additional link in the planning toolchain. Therefore, another objective of this work is to show possible use-cases of Interactive3D, as it is applied to planning scenarios that were derived from real clinical cases.

1.2 Organisation

The present work is organised as follows. The remaining sections of this chapter give the reader an insight into the field of radiation therapy. We discuss the interaction between radiation and matter, the biological effects of radiation, and the various delivery systems, with a focus on external beam radiotherapy.

Chapter 2 (Treatment Planning) The main planning steps are detailed: dose computation as it is deposited by beams of radiation, definition of planning goals, and approaches to multicriteria optimisation. An overview of sampling strategies is given, as well as of the visualisation and of the interaction methods used treatment planning.

Chapter 3 (Mathematical Model) The existing planning framework within which we develop our methods is presented. Starting with the mathematical formulation of a volume based optimisation problem, we derive a method for localised dose optimisation. Next, the different planning interfaces for the classical planning approach are presented. Ending this chapter, the requirements for an interface for localised dose planning are formulated. The next three chapters tackle the implementation of such an interface.

Chapter 4 (Visualising the Optimisation Problem) In this chapter the various approaches taken to fully represent the optimisation problem and to highlight areas of interest are presented. The current solution, optimisation constraints, and the various organs are represented as closed surfaces within a virtual space.

Chapter 5 (Direct Optimisation Interactions) This chapter shows how interactions with the virtual scene can be translated to meaningful optimisation operations: setting local optimisation targets and relaxing local constraints.

Chapter 6 (Sampling Issues) The optimisation problem is the principal limiting factor in achieving interactive rates. This chapter details our approaches to sampling the search and anatomical spaces, with the explicit goal of reducing the size of the optimisation problem.

Chapter 7 (Experimental Results) This chapter presents the experiments for showcasing the Interactive3D tool on real cancer cases and for benchmarking the sampling methods.

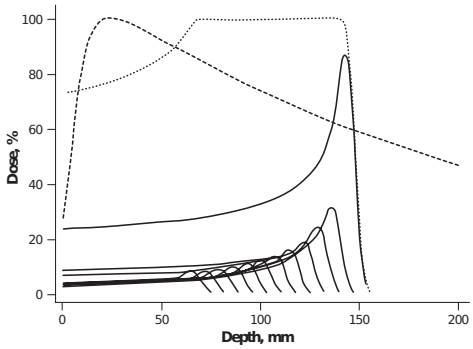
Chapter 8 (Conclusion) This chapter summarises the work presented in earlier chapters and emphasises contributions to the state of the art. Finally, indications for future research are given, with the goals of improving the planning tool and of answering the questions that were left open.

1.3 Biological Effects of Radiation

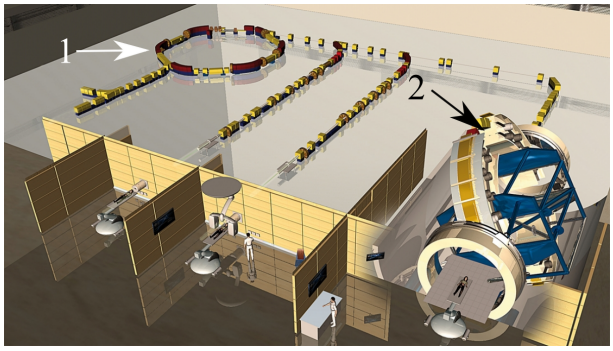
As ionising radiation interacts with matter, it ejects electrons from their orbit around the nucleus which results in the localised release of a large amount of energy. This energy can readily damage the DNA strands in cells and, if left unrepaired through mitosis, it will likely lead to cell death [5, Ch. 2]. Dose is defined as the energy imparted by radiation per unit of mass of material and is measured in Gray (Gy) [24, p. 10]. Because tumour cells often have short life cycles, they also have a limited capability to repair radiation damage. This implies that, for certain cancers, it may be advantageous to divide the treatment into a number of fractions, which, combined, deliver a large amount of dose. Each fraction has the opportunity to kill the remaining cancer cells, while, in between fractions, healthy cells have the chance to repair radiation damage [5, Ch. 3–4] [25].

Depending on the mechanism that leads to ionisation, radiation can be categorised as directly or indirectly ionising. Directly ionising radiation results from the interaction between charged particles (e.g. protons, electrons, atomic nuclei, etc.) moving at relativistic speeds and matter. On the one hand, indirect ionising radiation carries no electric charge, with the bulk of ionisation effects being the result of secondary ionisation events [5, Ch. 1]. Ionising radiation can be produced by the interaction of beams of photons or particles with matter. A photon travelling at relativistic speed may interact with an atom to eject an electron lying on the outer shell, which, in turn, will ionise other nearby atoms [24, Ch. 5]. For photon radiation, the maximum dose is deposited just below the surface of the target and drops exponentially as depth increases. On the other hand, particle

beams exhibit very different interaction with matter [24, Ch. 6]. Particles, such as protons, have a much better penetration profile and impart most of the dose shortly before coming to a full stop. The amount of dose imparted by particle radiation as it passes through the target is called the *Bragg curve* and it's maximum is called the *Bragg peak*. By modulating the energy of the scanning beam, one can spread out the Bragg peak to deposit most of the dose over the range of depths in which the tumour is located, with minimal spillage beyond that. Figure 1.1a shows the dose profile resulting from modulating the Bragg peak of proton beams, compared to the dose profile of photon radiation. Clearly, proton radiation can be used to deliver the same amount of dose to the tumour as photon therapy, but with improved sparing of proximal tissue. Indeed, the use for particle therapy has been growing steadily [26], but the technological costs associated with this technology [27] mean that photon radiation will remain the standard for years to come (Figure 1.1b shows a rendering of a particle therapy centre; in comparison, most photon therapy systems are comparable in size to a CT scanner).



(a) Bragg Peak modulation



(b) Particle therapy centre

Figure 1.1: (a) By adding the contributions of twelve proton beams (solid lines) the resulting Bragg peak (dotted line) is spread out to cover an extended depth range. For comparison, the dose distribution of a $10\,\mathrm{MV}$ photon beam is shown (dashed line). The dose delivered by the proton beam is lower than the peak just under the surface and drops to zero shortly after $150\,\mathrm{mm}$. In comparison, the photon beam delivers its maximum dose just under the surface; some dose is also delivered well beyond $150\,\mathrm{mm}$. Reproduced with permission from [27]. (b) Rendering of the Heidelberg Ion-Beam Therapy Center. The synchotron (1) accelerates the ions to $75\,\%$ of the speed of light and the resulting beam is directed to the three therapy rooms. The targeting device (2) bends the beam for improved delivery angle. Annotated from [28].

1.4 External Beam Radiation Therapy

In *external beam radiation therapy* (EBRT) the therapeutic dose is delivered from outside the body, by beams of photons or particles. Early systems used radioactive material, such as cobalt and caesium, as the source for radiation. With the advent of the *linear accelerator* (LINAC), conformal radiation therapy (CRT) became possible, as beams were formed by fixed or multileaf collimators and delivered from different directions towards the target [6, Ch. 1]. Next, *computed tomography* (CT) and, later, *magnetic resonance imaging* (MRI) were incorporated into clinical routine for both diagnosis and for directing therapy, in what is known as image guided radiotherapy (IGRT). Another important advance was the introduction of inverse planning, which would determine the beam configuration given a set of required dose targets. This led to the wide adoption of *intensity modulated radiotherapy* (IMRT), which could deliver highly conformal and nonconvex dose distributions [29]. If precision is required, IMRT is dosimetrically superior to CRT [30, 31], as less healthy tissue is exposed while achieving the same or better target coverage. Still, it is important to note that inverse planning is more complex and more time consuming than forward planning, as the user has to take into account a large number of criteria. Despite this, IMRT has quickly replaced CRT for most interventions, even though, for some types of cancers, no improvement in outcome could be proven [32].

Presently, research is concentrated around some promising therapy options. Adaptive radiotherapy (ART) tracks morphological and positional changes of the target between fractions and adjusts the treatment plan accordingly. Furthermore, recognising that tumours are not homogeneous masses, but made up of subvolumes with different concentration of clonogens, *dose painting* techniques use functional and molecular imaging to increase dose exposure in areas with greater tumour infiltration [6, pp. 6–9]. Figure 1.2 gives an overview of the key technologies used in the different therapy types we presented.

Stereotactic body radiation therapy (SBRT) and *stereotactic radiosurgery* (SRS) deliver a high amount of dose in either a single session or in a small number of fractions, with the aim of inflicting the maximum amount of damage to the target. SRS is used to treat tumours of the brain and spine, while SBRT is targeted at malignancies throughout the body, excluding the brain and spine. Because of the high treatment dose, very good delivery precision is required. This is usually achieved through the use of a stereotactic frame, which fixates the patient and provides a coordinate system to locate the tumour [33]. An equivalent treatment system, is the CyberKnife®, developed by Accuray™[34–36]. On the CyberKnife, the LINAC is mounted on a six axis robotic arm, while patient position-

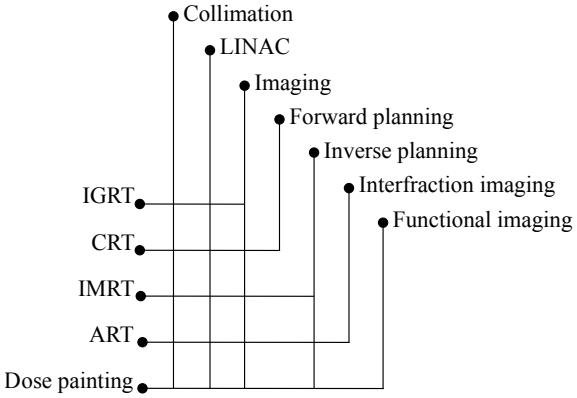


Figure 1.2: The role of key technologies leading to some of the most important types of EBRT.

ing is assisted by a five axis robotic couch (see an overview in Figure 1.3). The system is thus able to target tumours anywhere within the body, with the added advantage that beams do not have to be generated around an isocentre. Stereoscopic X-ray imaging and an optical tracker allow the system to predict the target position in real time and to compensate patient and organ motion; consequently, a stereotactic frame is not required. In contrast to classical IMRT machines, where beams are formed by rectangular leaflets, the CyberKnife beams are shaped by circular collimators of varying radii.

Figure 1.3: The CyberKnife® radiosurgery system showing the LINAC mounted on the industrial robot, the robotic couch, and the stereoscopic X-ray sources and floor detector. Image courtesy of Accuray Incorporated - © [2015] Accuray Incorporated. All Rights Reserved.

The methods in this work were developed for the CyberKnife system, yet they can easily be generalised for most IMRT planning systems. The only exception is the beam resampling algorithm from Section 6.2, as beam generation takes into account particularities of the delivery setup.

2 Treatment Planning

Treatment planning is the process of determining machine parameters that lead to the delivery of the therapeutic dose distribution. In the case of the CyberKnife, these parameters define a set of beams; each beam is specified by source position, direction, shape, and intensity [34]. As depicted in Figure 2.1, planning starts with a view of the patient anatomy, showing the organs and the tumour area. Current planning approaches can be grouped in one of two categories, *forward* or *inverse* planning, for which an overview is given in the next section.

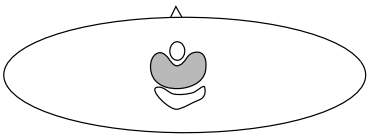
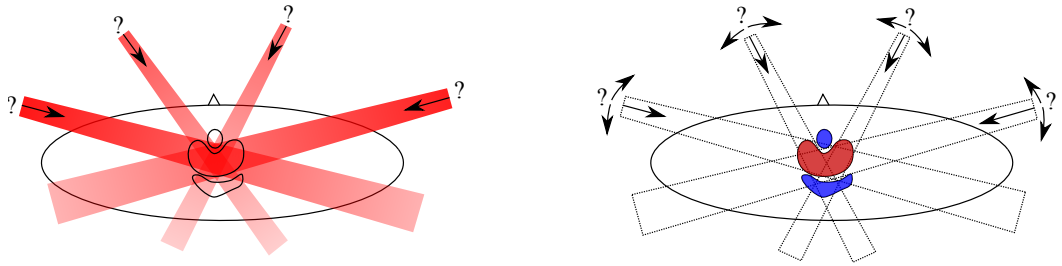


Figure 2.1: Axial sketch of a patient anatomy containing a concave tumour (gray) surrounded by two healthy organs.

2.1 Forward and Inverse Planning

A typical *forward planning* approach is as follows. First, the medical physicist determines the entry direction and the shape of each beam, by manually placing them in the planning software. Next, attenuating wedges and multileaf collimators may be added to modulate the intensity and the transversal shape of the beam respectively. Finally, the intensity of each beam is computed such that the therapeutic dose is delivered, usually by solving an optimisation problem. This is an iterative process, as the physicist has to repeatedly adjust the beam parameters until an acceptable plan is found. An example of forward planning is shown in Figure 2.2a. Forward planning is usually employed for simpler cases, where the tumour is well defined and not very close to critical organs [37].



(a) Forward planning

(b) Inverse planning

Figure 2.2: (a) Forward planning method in which the beam shapes and directions are determined beforehand by the user. The subsequent optimisation adjusts the beam intensities (red) such that the treatment dose is delivered. (b) Inverse planning method. The user specifies the dose constraints for both the target (typically, maximum and minimum dose bounds) and for the healthy organs (typically, maximum dose bounds) and the ensuing optimisation process fully determines the treatment beam arrangement: source, direction, shape, and intensity. The minimum and maximum bounds are depicted in red and blue respectively.

As the precision of the delivery and imaging improved, the search space size also grew, making forward planning inadequate for such systems [38]. This led to the development and application of *inverse planning* methods. In this approach, the clinician first contours the different volumes, tumours and organs. She then defines the planning goals (criteria), such as prescription dose and dose tolerance values (see Figure 2.2). Finally, an optimisation step is used to find the machine parameters that give rise to a dose distribution that optimally fulfils these criteria.

For a CyberKnife treatment, 100 to 200 beams are used; each beam originates in one of 100 predefined nodes, it can intersect the tumour in any point, and it is shaped by one of 12 circular collimators [10, 39]. It is then easy to see why, given such a large parameter search space, inverse planning is used for the CyberKnife.

2.2 The Inverse Planning Problem

Generally, inverse planning consists of the following steps:

1. Patient anatomy is imaged.
2. The various *volumes of interest* (VOIs), organs and tumour sites, are contoured.
3. Dosimetric goals, such as prescription dose, minimum and maximum doses, are defined and mathematically modelled as optimisation criteria.

4. An optimisation problem is solved to yield the machine parameters that deliver the dose distribution, corresponding to an acceptable trade-off between the goals.

2.2.1 Volumes of Interest

Computed Tomography (CT) is universally used in radiation therapy for contouring the various organs and targets [40] – the *volumes of interest* (VOIs). However, CT limitations, such as low soft tissue contrast and the lack of functional information, mean that various other imaging modalities are increasingly being incorporated into this process. Magnetic resonance imaging (MRI) offers excellent contrast in soft tissue, despite having a relatively low spatial resolution [41]. Consequently, MRI can improve the contouring of both organs and tumours. Functional MRI (fMRI) and diffusion tensor imaging (DTI) have been investigated for their ability to reveal previously invisible structures. Preliminary studies have shown that dosage to sensitive areas, such as the primary motor cortex or brain fascicles, can be substantially reduced by simply taking into account these volumes during planing [42–44]. Furthermore, nuclear imaging techniques, such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT), have the potential to expose both tumour inhomogeneities as well as previously invisible cancer sites [45–47]. Using these technologies, target contouring can be adjusted to include malign tissue, which previously would have been missed in conventional scans, or to exclude healthy tissue, which would have been treated as part of the tumour.

We adopt the following VOI definitions from [48]:

Gross tumour volume (GTV) The gross demonstrable extent of the tumour. Additional imaging, such as PET or fMRI, may influence the size and shape of the GTV.

Clinical target volume (CTV) The GTV plus the areas that may contain subclinical malignant disease with a certain probability of occurrence (typically higher than 5%). The subclinical malignant disease denotes microscopic infiltration beyond the boundary of the primary tumour and the potential metastatic extension to other tissue and/or to the lymph nodes. Delineation of the CTV depends on the experience of the clinician.

Planning target volume (PTV) The CTV plus margins to account for uncertainties due to organ and patient motion, internal and setup inaccuracies (such as patient positioning and alignment). These margins are extended to account for the maximum variations along each direction.

Organ at risk (OAR) Normal tissue that is especially vulnerable to radiation during treatment. Dose to OARs is usually restricted to remain below certain values. Similar to the PTV, a *planning organ at risk volume* (*PRV*) may be defined to account for uncertainties in the localisation of the OARs. Throughout this work we do not discriminate between PRVs and OARs, and only use the latter concept.

Shell Artificial structure formed by extending the surface of the PTV by a given distance. Shells are used to restrict the dose outside the PTV to below a certain value.

2.2.2 Dose Deposition

The accumulation of dose in the tissue can be computed by a number of algorithms. The most accurate ones are based on Monte Carlo (MC) simulations of a large number of photons passing through matter [49]. MC simulations show very good agreement with empirical measurements, yet are computationally expensive [50]. Consequently, model based algorithms are used during planning, as they trade complexity and correctness for speed [49]. Of these, convolution algorithms are the most accurate, being able to better account for scatter or inhomogeneities than, for example, pencil beam algorithms. Typically, the final planning step is to verify the quality of the plan by computing the dose distribution through an MC algorithm [50, 51].

2.2.3 Plan Evaluation

The dose distribution realised by a plan is subjected to a number of benchmarks. First, it is visually inspected, either as a three-dimensional rendering or overimposed on top of two-dimensional axial slices, as isodose curves, or colourwash (see example in Figure 2.3). Next, cumulative *dose-volume histograms* (DVHs) summarise the dose distribution on a per VOI basis. The DVH plots the percentage of a VOI receiving at least a certain amount of dose against dose values. For example, a dose of $0\,\mathrm{Gy}$ will correspond to $100\,\%$ of the VOI (meaning that the entire VOI receives a dose of at least $0\,\mathrm{Gy}$).

In this work, the following metrics are used to characterise a plan:

Coverage (CO) The percentage of the PTV receiving a dose greater or equal to the prescription dose (T_x).

Dose volume (V_x^{VOI}) Similar to CO, the percentage of a VOI receiving a dose greater or equal to x.

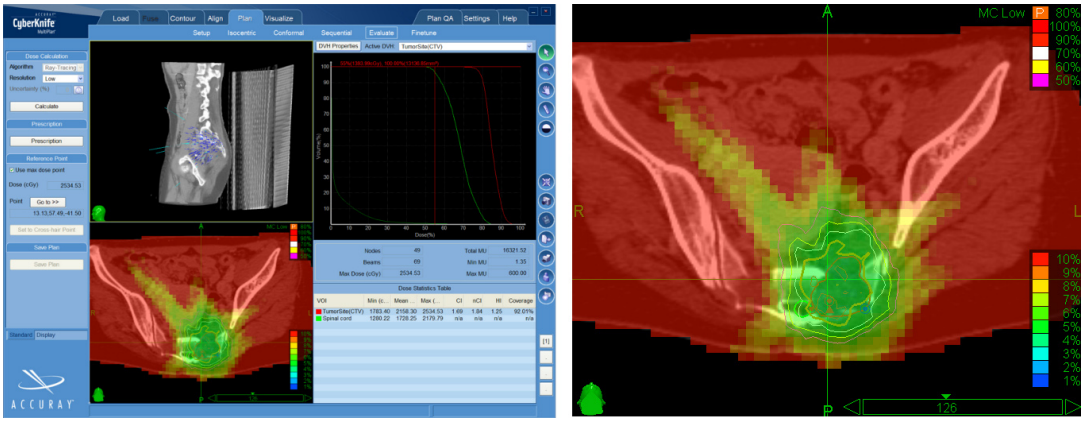


Figure 2.3: Plan evaluation for a spine case using the MultiPlan® software. Right half is a detail view of the bottom-left tab. (top-left tab) Sagittal view of the treatment beams. (top-right tab) Dose-volume histograms for different VOIs. (bottom-left tab) Dose colourwash and isodose curves overimposed onto axial CT slice. (bottom-right tab) View of numerical dose statistics. Source: Accuray [52].

Volume dose (D_y^{VOI}) The dose value for which y % of the VOI is covered.

Conformality index (CI) The volume of the tissue receiving a dose that is greater or equal to the prescription dose, relative to the volume of the PTV receiving the same doses. $CI = V_{T_x}^{patient}/V_{T_x}^{PTV}$.

Homogeneity index (HI) Maximum PTV dose over prescription dose: $HI = D_{MAX}^{PTV}/T_x$, where D_{MAX}^{PTV} is the maximum dose measured within the PTV.

2.3 Plan Optimisation

One must first define the criteria by which a dose distribution is considered to be optimal. In practice, a large number of often conflicting criteria are taken into account. Consequently, there rarely exists a solution that simultaneously satisfies all goals [53]. The task of the human planner is to identify those plans that strike the *optimal* trade-offs between all criteria.

2.3.1 Planning Criteria

Some of the most used planning criteria are: coverage, conformality, homogeneity, mean and maximum doses to OARs and SHELLs, VOI dose-volumes (percentage of the VOI

covered subjected to a given dose), total treatment time, and total *monitor units* (MU). Monitor units express the output of the linear accelerator, which is usually calibrated such that 1 MU corresponds to 1 cGy measured at a fixed reference point [54, p. 433].

Optimisation criteria are modelled as mathematical functions for which an extreme value (maximum or minimum) represents the optimum. A widely used function is the L_p-norm of the dose deviation from a given reference dose. Explicitly, for a given VOI, the L_p norm is a summation of deviations from a given reference dose across all voxels. For example, to optimise for PTV coverage, one often minimises the L_1-norm of all PTV voxels receiving a dose lower than the prescription dose (note that the L_1-norm in this case, is simply a linear approximation of a non-convex objective). Other functions, such as the tumour control probability (TCP) or the normal tissue complication probability (NTCP) model the biological response of the tissue to dose [5].

2.3.2 Multicriteria Optimisation

As a large number of goals are taken into account, it is natural to use multicriteria optimisation to search for a solution. In this section we present the general concepts of multicriteria optimisation and the established methods to solve this class of problems.

Consider the following criterion function $f : \mathbb{R}^n \to \mathbb{R}$ for which a *global minimum* $\mathbf{x}_0$ exists: $f(\mathbf{x}_0) < f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{x}_0\}$ [1]. Suppose we regard l criteria, $f_1, ..., f_l$, all similarly defined as f, for which the global minimum is not necessarily the same.

Definition 2.1: Vector Dominance [55, Ch. 4]

A vector $\mathbf{x} \in \mathbb{R}^n$ *is said to dominate another vector* $\mathbf{y} \in \mathbb{R}^n \Leftrightarrow x_i \leq y_i, \forall i \in \{1, ..., n\}$ *and* $\exists j \in \{1, ..., n\}$ *for which* $x_j < y_j$.

Definition 2.2: Multicriteria Optimisation Problem (MOP) [55, Ch. 4]

An MOP minimises the following problem/program:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_l(\mathbf{x})) \colon X \to T \\ s.t. \quad & \mathbf{u}(\mathbf{x}) \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{2.1}$$

X is referred to as the *search (decision) space* and T is the *objective space*; X and T are subspaces of $\mathbb{R}^n$ and $\mathbb{R}^l$ respectively. The *feasible set*, X_{feas}, is defined as:

[1] Convention: we use bold capital letters to denote matrices ($\mathbf{A}$), and bold small letters to denote vectors ($\mathbf{x}$); for sets we use capital letters (S), while for scalar variables we use small letters (a).

$$X_{feas} = \{\mathbf{x} \mid \mathbf{u}(\mathbf{x}) \leq \mathbf{b} \text{ and } \mathbf{x} \geq \mathbf{0}\} \subseteq X$$

Definition 2.3: Pareto Efficient Solution [55, Ch. 2, 4]

A solution $\mathbf{x}_0 \in \mathbb{R}^n$ *is said to be Pareto efficient (nondominated)* $\Leftrightarrow \nexists \mathbf{x} \in \mathbb{R}^n$ *such that* $\mathbf{F}(\mathbf{x})$ *dominates* $\mathbf{F}(\mathbf{x}_0)$.

In other words, a Pareto efficient solution cannot be improved without worsening at least one criterion. Conversely, non-Pareto solutions are said to be dominated (inefficient) and are generally undesirable as they can be improved without any additional cost [56]. The set of all Pareto efficient solutions is known as the *nondominated set*; the *Pareto Front (PF)* is the representation of the nondominated set in the objective space.

Definition 2.4: Pareto Front [13, 57]

For a given MOP, the Pareto Front PF = $\{F(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_l(\mathbf{x})) \mid \mathbf{x}$ *is Pareto efficient*$\}$

The discussion above motivates Pareto efficiency as a necessary condition for a solution to be used therapeutically; it is, however, not sufficient, as some criteria may well be outside acceptance parameters [58]. This work defines the concept of **Pareto optimality** as the necessary and sufficient condition to accept a solution. In fact, the process of planning is complete once such a solution is identified. Figure 2.4 sketches the Pareto front for a bi-criteria optimisation problem. If we consider that the clinical requirements are: (1) the PTV under-coverage should not be larger than 25% and, (2) OAR V_{10} should not be larger than 75%, only the continuous portion of the curve is clinically acceptable and, thus, Pareto optimal.

Definition 2.5: Pareto Optimality

A plan/solution is Pareto optimal if it is Pareto efficient with respect to all optimisation criteria and it satisfies all clinical requirements.

When considering non-convex criteria and/or non-convex constraints, a number of issues make the MOP hard to solve [59]. The main difficulty consists in the existence of local optima. As such, the stopping condition for non-convex algorithms is to some degree arbitrary. Convex optimisation problems do not suffer from such problems: any local optimum is necessarily the global optimum and can be found by well established algorithms [55, Ch. 4]. Moreover, the Pareto front of MOPs with convex criteria and constraints is itself convex [56]. Two classes for solving such MOPs have been widely employed in radiotherapy planning: Pareto front exploration [13, 60–67] and lexicographic

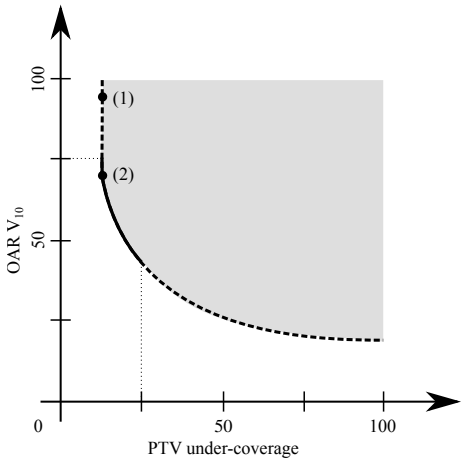


Figure 2.4: Pareto front for a two criteria MOP: PTV under-coverage (percentage of PTV volume receiving less than the prescription dose) and OAR V_{10} (percentage of OAR volume covered by the 10% isodose). The red segments represent areas of the PF that are not clinically acceptable; Pareto optimal plans are only on the black segment. The objective space is grayed out. Solutions (1) and (2)—black dots—exemplify a lexicographic ordering approach: (1) is the solution after optimising the first criterion; (2) is the solution after optimising the second criterion and preserving the previous result.

ordering [14, 58, 68, 69].

2.3.2.1 Pareto Front Exploration (PFE)

One approach to identifying a Pareto optimal plan is to search for it on the Pareto front [13, 65–67]. The original MOP is solved by combining the multiple criteria into a single objective scalar function [70]:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i}^{l} w_i f_i(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{u}(\mathbf{x}) \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{2.2}$$

If the weights are chosen such that they are positive, then it is trivial to show that the minimiser of the above program is a Pareto efficient solution of the original MOP [71]. Furthermore, for any such solution $\mathbf{x}_0$, there exists a set of positive weights w_i such that $\mathbf{x}_0$ is the global minimiser of the optimisation problem 2.2 [72, 73]. By solving the optimisation problem for different weight values, the shape of the PF can be reconstructed (e.g. through the sandwiching approach [63, 74, 75]). For a finer analysis, new points (plans)

can be generated on the fly by interpolating the solutions that have already been solved. Such points are approximations of Pareto solutions and their quality can be assessed [63].

In practice, Pareto front exploration faces several difficulties. First, the number of plans that have to be optimised is relatively large and increases exponentially with the number of objectives that are considered [73, 76]. Consequently, there is a substantial computational overhead associated with this method. Furthermore, the choice of the individual weights is not straight forward [61, 63, 71], as areas of the PF may be incorrectly reconstructed if undersampled.

2.3.2.2 Lexicographic Ordering (LO)

Lexicographic Ordering solves the MOP by sequentially optimising one criterion at a time. At each step, the optimum with respect to the current criterion is obtained. Before advancing to the next step, the current solution is saved by transforming the optimised criterion into a constraint. This guarantees that subsequent steps will not worsen previously optimised criteria. With each step, the problem becomes more and more constrained and, as such, later criteria are harder to satisfy [68]. This implicit prioritisation means that important goals should be optimised first.

Suppose the criteria functions $(f_1, ..., f_l)$ are ordered in decreasing priority. The first step is to optimise the following program, yielding $\mathbf{x}_1$ as the global minimiser:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_1(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{u}(\mathbf{x}) \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{2.3}$$

The subsequent step is to optimise the program below, for which a constraint preserving the previous result has been added:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_2(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{u}(\mathbf{x}) \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \\ & f_1(\mathbf{x}) \leq f_1(\mathbf{x}_1) \end{aligned} \tag{2.4}$$

Let $\mathbf{x}_i$ be global optimum of the i-th program, where $i \in \{1, ..., l\}$. Then, the final solution $\mathbf{x}_l$ must satisfy the conditions $f_i(\mathbf{x}_l) \leq f_i(\mathbf{x}_i)$; because $\mathbf{x}_i$ was already found to be the global optimiser of f_i for less strict conditions, it follows that $f_i(\mathbf{x}_l) = f_i(\mathbf{x}_i)$. It is trivial to show that there exists no $\mathbf{y} \in \mathbb{R}^n$ such that $F(\mathbf{y})$ dominates $F(\mathbf{x}_l)$. Supposing such a

$\mathbf{y}$ exists, then, by definition, $\exists j$ for which $f_j(\mathbf{y}) < f_j(\mathbf{x}_l)$ and $f_i(\mathbf{y}) \leq f_i(\mathbf{x}_i), \forall i \in \{1, .., l\}$. Obviously, $j \neq 1$, as $f(\mathbf{x}_1)$ is the global minimiser for the least constrained problem. If $j = 2$, then $f_1(y) \leq f_1(\mathbf{x}_1)$; this is the same constraint that was added to find $\mathbf{x}_2$ as the global optimiser of the extended program and thus $f_2(\mathbf{x}_2) < f_2(\mathbf{y})$, which invalidates our assumption. Continuing by mathematical induction, one shows that $F(\mathbf{x}_l)$ is non-dominated and, thus, $\mathbf{x}_l$ is a Pareto efficient solution of the original MOP. Figure 2.4 shows the solutions of a two criteria MOP obtained through an LP approach. First, the PTV criterion is optimised, then the OAR criterion, keeping the previous goal intact (2).

As more and more constraints are added, criteria that are optimised later become harder and harder to satisfy. It is then important not to overachieve the former optimisation steps in order to allow enough slack for these latter criteria [68]. Once the prioritisation list has been determined, the planning process is automatic and a single Pareto efficient solution is found [68, 69]. If this solution is deemed to be unsatisfactory[2], the planner may modify the optimisation parameters, such as dose bounds or target doses, and then rerun the LO process.

Several variations on the basic LO method have been studied. Wilkens et al. [14] set equally prioritised goals to be optimised at the same time, while slip variables allow for implicit relaxation of the previous criteria. Schlaefer et al. [58, 77] tackle the limitations of automatic LO planning by a stepwise approach. Between each optimisation step, the planner is allowed to relax some constraints and the same objective is reoptimised. Only after this objective has been satisfactorily improved, does the planner move on to the next one. This process can be seen as exploring the solution space by moving along one dimension at a time and, after optimising all criteria, arriving at a point that lies on the PF. Movements on the PF can be circular: one can return to a previous point in the objective space by a series of dual relaxation and optimisation steps. Long et al. [69] also propose an interactive approach which combines LO with PF exploration. At each step, a bi-criteria problem is solved and the planner can select the desired relative trade-off between the two.

2.3.2.3 Limitations of Multicriteria Optimisation

As it is relatively easy to find the global optimum of real valued convex functions, it is desirable to model planning criteria as such. However, one limitation of using these function is that they are only partial models of the actual biological criteria (e.g. the irradiation of tumorous tissue with enough dose to cause cell death). Consider two different

[2] Due to one or more criteria not being sufficiently satisfied.

dose distributions $\mathbf{d_1}$ and $\mathbf{d_2}$ and a criterion f. It is not uncommon for $f(\mathbf{d_1}) = f(\mathbf{d_2})$ even though the dose distributions and the implied therapeutic effects are different.

Such an example is presented in Figure 2.5: two different dose distributions resulting in the same PTV coverage, of less than 100% at the prescription dose d_{Tx}. Suppose that in one case, the uncovered part of the PTV is in a low risk area, while the other one is in a high risk area of tumour infiltration. All other things being equal, one would regard the plan delivering $\mathbf{d_2}$ to be superior to the one delivering $\mathbf{d_1}$ even though, mathematically, both dose distributions result in the same criterion-value. Coverage is clearly not sufficient to describe the plan.

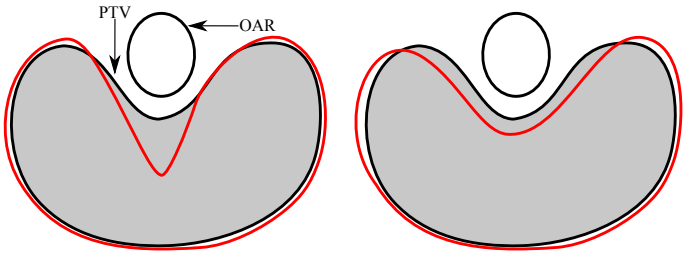


Figure 2.5: Two different dose distributions (red) achieving the same coverage of the PTV (gray filled) and avoiding the OAR (white filled). Assuming the probability of tumour infiltration is higher in the centre of the PTV than at the periphery, the dose distribution in (a) is less advantageous than the one in (b).

In a VOI-centric approach[3], a further limitation is that one only has limited control over the shape of the dose distribution. Take for example the classical scalarization method (Equation 2.2). It is obvious that changing the weighting factors of an objective defined over an entire VOI will remodel the dose distribution towards or away that volume. Yet it is not possible to say how much or where change will happen within the volume. This problem has been addressed by varying the weights [17, 20, 21, 23] or the reference (prescription) doses [22, 23] at voxel level for penalty functions. For example, to increase the dose in a set of voxels, one could increase either the corresponding weights or one could modify the corresponding reference doses, thus raising the relative importance of those voxels. Unfortunately, the selection of these parameters is not straight forward and some authors set them empirically [17, 20], by using heuristics [21], or iteratively [23]. Furthermore, any improvement comes at a cost to voxels whose parameters

[3]Each criterion is defined over an entire volume.

did not change and, consequently, it is difficult to spatially control the trade-offs.

Taking into account the above limitations and noting that the shape of the three-dimensional dose distribution is a central theme of the treatment planning process, we formulate our first hypothesis:

Hypothesis 1. *The local shape of the three-dimensional dose distribution can be explicitly formulated as a criterion and can be optimised alongside the other planning criteria.*

Note that such a problem, of locally shaping the dose distribution, is typical for dose painting techniques, where dose is prescribed either per voxel (known as *dose painting by numbers*—DPBN) or per region/sub-volume (known as *dose painting by contours*—DPBC). This fine grained dose prescription is usually based on morpho-biological information obtained through various imaging modalities [46].

Hypothesis 1 summarises the main the effort of this work: to construct a planning tool that allows the user to precisely form the dose distribution. Consequently, the following requirements are formulated for this tool:

1. Goals and trade-offs must be precisely definable in space.

2. Favourable trade-offs must be made explicit, to allow their efficient exploration.

3. The final solution must be Pareto optimal.

Some obvious design constraints follow directly from these requirements. First, this tool must make adequate use of visualisation methods for representing the optimisation problem. Second, the optimisation problem must be modifiable through interaction with the virtual scene. Third, computation time must be kept to a minimum, to ensure quality of interaction. Fourth, a multicriteria approach is needed. Next sections give an overview in the state of the art of the sampling, visualisation and interaction approaches in radiation therapy planning.

2.4 Sampling Strategies

The optimisation problem is a discretized mathematical model of the continuous planning problem. Such a problem would be intractable with current hardware if one were to take into account every voxel of the CT scan, or every possible beam orientation and size. In keeping the optimisation program to a manageable size, careful consideration must be given to the sampling of the anatomical and beam spaces.

2.4.1 Anatomical Space Sampling

The sampling of the patient anatomy determines the dosage errors resulting from planning. With decreasing resolution, hotspots and coldspots appear throughout the VOIs, as the coarse sampling cannot account for the fast dose variation [78–82]. For example, if the shape of the PTV is not accurately reconstructed, areas to the edge of the PTV may not be covered by the prescription dose. Conversely, an inadequately reconstructed OAR may lead to overdosing the actual organ, especially on its outer surface. Such effects are known as aliasing errors and are due to improper sampling [83].

Niemierko and Goitein use adaptive grid spacing [84] and random sampling [85] to reduce the number of samples whilst limiting dosage errors within acceptable limits. Dempsey et al. [79] elegantly use Fourier analysis to estimate that a uniform sampling rate of $2.5\,\text{mm}$ would limit dose errors to under $2\,\%$ for IMRT (considering a minimum leaflet width of $5\,\text{mm}$). Compare this to the much finer rate of $0.8\,\text{mm}$ needed to ensure the same error bounds for proton therapy [80]. Both papers demonstrate that adaptive grids, in which areas of high dose variation are sampled on a finer grid, can be employed to reduce the total number of voxels while maintaining errors within bounds.

Scherrer et al [86] introduce a two step adaptive clustering technique: first, a cluster hierarchy is constructed of voxels based on their dose-volume information and, second, clusters leading to dose shape errors during subsequent the optimisation are locally refined. The first step is common for all plan computations of a single case, while the second step must be ran every time the plan changes. Similarly, Martin et al. [87] approximate the solution of the full optimisation problem by random sampling, and they apply this to a gradient-based optimisation algorithm.

2.4.2 Beam Space Sampling

The planning problem is intrinsically combinatorial, and the cardinality of the sampled search space influences the quality of the solution: the higher the cardinality the better the solution [11, 88].

Recognising this limitation, Schweikard et al [39] proposed an iterative random beam resampling algorithm (RBR). At each step, a complete and optimal solution is found with respect to the current beam set. The beam set is thus partitioned into non-zero weighted beams, B_+, and zero weighted beams, B_0. Beams in B_0 correspond to non-basis variables, and having no contribution to the current solution, they can be safely replaced with newly generated beams. The optimisation in the next iteration may improve on the previous solution, or, at worst, deliver the same solution. Thus, one can take into

account a large number of beam configuration, while keeping the problem dimensions within very reasonable limits. The complete algorithm is given in Listing 1. Expanding on this, Schlaefer et al. [10] developed their iterative beam resampling algorithm (IBR), which uses heuristics to rank the quality of the beams that are generated. Least useful beams are removed and new beams are introduced based on the quality of the previously determined solution. Compared to RBR, IBR is deterministic and also results in better quality solutions in less iterations. A disadvantage of IBR is that it is controlled by a large number of parameters, which are difficult to set.

Algorithm 1 The classic RBR algorithm [39]. Following optimisation, the beamset is partitioned into weighted (B_+) and unweighted (B_0) beamsets: $B = B_+ \cup B_0$. N_r random beams from B_0 are replaced by B_n, new randomly generated beams. N_t is the total number of beams for one optimisation run. Δ_{obj} is the objective improvement over the previous iteration. T_{obj} is the objective improvement threshold - when the improvement is less than this threshold the algorithm ends. $\mathbf{generateRandom}(N)$ randomly generates a beam set of size N. $\mathbf{selectRandom}(B, N)$ randomly selects N beams from beam set B.

1: **function** RBR(N_t, N_r)
2: $B \leftarrow generateRandom(N_t)$ ▷ Initialise beam set
3: $\Delta_{obj} \leftarrow \infty$ ▷ Initialise objective improvement
4: $o_{prev} = \infty$ ▷ Initialise previous objective value
5: **while** $\Delta_{obj} > T_{obj}$ **do**
6: $(B_+, B_0, o_{curr}) \leftarrow \mathbf{solve}(B)$ ▷ Optimise with current beam set
7: $\Delta_{obj} \leftarrow o_{prev} - o_{curr}$
8: $o_{prev} = o_{curr}$
9: $B_n \leftarrow \mathbf{generateRandom}(N_r)$ ▷ Randomly generate new beams
10: $B_r \leftarrow \mathbf{selectRandom}(B_0, N_r)$ ▷ Randomly select beams to remove
11: $B = B_+ \cup B_n \cup (B_0 \setminus B_r)$
12: **end while**
13: **return** B_+, o_{curr}
14: **end function**

2.5 Visualisation in Radiation Therapy

In radiotherapy, as in most other fields of medicine, huge volumes of data are generated during a single treatment. The bulk of the data is consists of scans of the patient anatomy [89] and computed dose distributions. One then has to use visualisation techniques, to

gain insight from this data, through pictures [90]. The usable information is either directly available, or data may require further processing, such as filtering or segmentation, to reveal salient features [91].

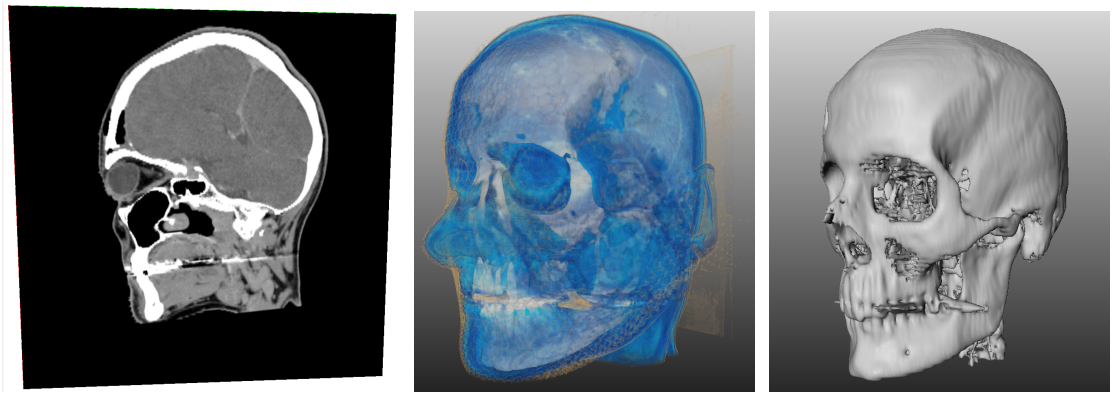
Typically, imaging is used to gain insight into disease extent, such as tumour localisation and morphology, and to analyse the computed dose distribution [92]. Furthermore, imaging is often used as a real-time assessment tool, for tracking tumour and organ motion, and even, more recently, for visualising the beams during treatment [93, 94]. However, it is the former use that is relevant for this work.

Disease and dosage information are usually extracted from three-dimensional scalar data sets of the patient anatomy and of the dose distribution. The most simple approach is to visualise these data as two dimensional (2D) slices, usually parallel to one of the three coordinate axes (as axial, sagittal, and coronal slices). For inspecting the treatment plan, the dose information is overlayed on slices of the patient anatomy, as either *isodose curves* or as *colourwash*. Isodose curves are closed lines connecting points that receive the same amount of dose. Points inside a curve with a given threshold d receive a dose greater than d, while points outside the curve receive a dose less than d. Colourwashing is simply the process of mapping the dose value to a pixel colour. An example of both visualisation techniques was shown in Section 2.2.3 Figure 2.3.

However, viewing scalar fields in three dimensions seems a more natural fit to the nature of the data. Consequently, research has been aimed at developing algorithms for 3D rendering. It is important to mention that, despite numerous advances in this direction, slice-by-slice viewing is a mainstay of clinical practice. This is because 3D visualisation is not universally superior to slice-by-slice viewing. For example, slices are better for detecting and analysing small features, while 3D rendering offers improved spatial context [95, 96]. For 3D visualisation, the most used methods are *cutting plane projection*, *direct volume rendering*, and *indirect volume rendering* [91].

The cutting plane approach is similar to slice visualisation, with the only difference being that the projection plane may be arbitrarily placed within the dataset. Figure 2.6a shows an example of cutting plane visualisation, where the scalar values are mapped to gray levels.

In the direct rendering approach, a transfer function is used to map each sample to an RGBA Tuple (colour and opacity); the entire scene is then projected back on the screen plane [97]. An optical model treats the samples as a distribution of particles with possible interactions between them. For example, particles can be *emitting*-only, in which case each sample emits light but does not absorb any. Complex models may also consider simple and multiple *scattering* and *shading* properties of particles. As such, a particle



(a) (b) (c)

Figure 2.6: Visualisation of a patient CT data by (a) cutting plane projection, (b) direct volume rendering, and (c) isosurface extraction using MeVisLab. Modifying the transfer function leads to different information to be displayed by direct rendering. In comparison, plane cutting and isosurface visualisation are a lot more simplistic.

may scatter external light rays impinging on it and it may also shade other particles. The standard algorithm for direct volume rendering is ray casting [98, Ch. 39]. For each pixel on screen, at least one ray is projected through the data set and the contributions of regular samples along the ray are combined to determine the final pixel value (see Figure 2.7). As the ray travels a volume block, not a series of objects, most ray casting algorithms do not consider scattering or shading effects [91]. Central to direct rendering is the use of a suitable transfer function that maps the scalar values to image attributes, such as colour and opacity [99]. Often, there is overlap between the value ranges of different structures (such as blood, muscle, gray, white matter, etc.), making observation of these structures in isolation particularly difficult (see Figure 2.6b). Multidimensional transfer functions can solve such problems, yet are difficult to define and incur additional overhead [100]. Complexity in the case of direct rendering is an issue to the planner, as he is confronted with an overload of visual information.

Indirect volume rendering extracts low level features, such as surfaces and curves, from the data set and displays them on screen. One of the most used techniques in this category, isosurface extraction, partitions the scalar field into areas containing values less than and areas containing values greater than a given threshold (outside and inside the surface, respectively). The surface itself connects points of value equal to the threshold. As this work makes heavy use of isosurface visualisation, the next section presents the *marching cubes* algorithm for extracting isosurfaces.

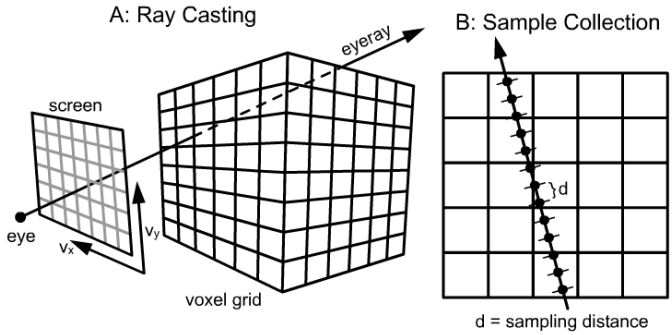


Figure 2.7: Ray casting through a volume. (a) From the point of view of the viewer rays are cast through the centre of each pixel on screen and pass through the data set. (b) The ray is sampled at regular intervals and the pixel value is computed from each of these contributions (source: `http://johnrichie.com/V2/richie/isosurface/volume.html`)

2.5.1 The Marching Cubes Algorithm

Marching cubes (MC) is the classical algorithm patented by Lorensen and Cline [101] for approximating an isosurface through a scalar field sampled on a regular grid. Given an iso-value v_0, the algorithm decides how the isosurface may intersect each cell on a grid. A cell is a cuboid, with each vertex representing a sample. For each cell, some of its vertices may be on the inside and some may be on the outside of the surface; in other words, the surface will pass between vertices of different classification (see Figure 2.8). Building the isosurface means triangulating the locations of the cutting points, as the surface passes through each cell. As there are 8 vertices per cell, there are 256 possible combinations for a surface to cut a cell. However, if symmetries are accounted for, there are only 15 unique combinations. So to extract the isosurface, one would compute the triangulation case and offset each vertex by linearly interpolating along the cut edge.

An important flaw of marching cubes is that some cases are ambiguous, leading to incorrect triangulation and, thus, holes in the surface (see Figure 2.8(c, d)). Fortunately, this issue can be mitigated [90] and the Visualisation Toolkit library, which was used in the implementation of this work, includes such a solution [102, Ch. 6].

2.6 Tools for Interactive Planning

Next, the state of the art of interactive approaches used in treatment planning is presented. Although it is a general concept, the meaning for *interactivity* shall be constrained to cover only the areas of interest for this work. Virtually all treatment planning is done

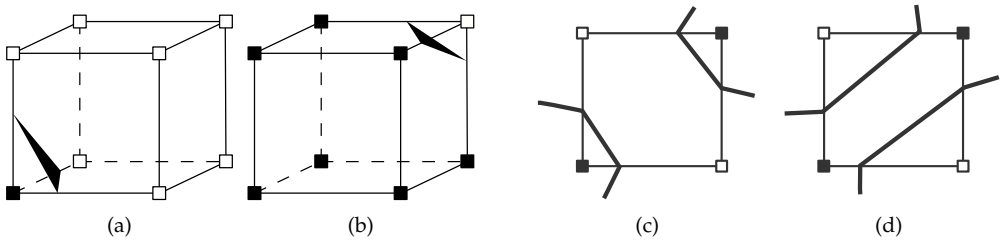


Figure 2.8: (a, b) Isosurface cutting a cell, in which one vertex is differently classified to the other seven (e.g. the value of the filled vertex is less than the threshold value, which in turn is less than the values of the empty vertices). The case showed in (b) can be easily reduced to (a). (c, d) Two-dimensional cell showing am ambiguous case: given the vertex classifications, both cases are plausible.

interactively, with the user having to at least delineate the VOIs, define the dosimetric goals, and tweak plan parameters [15] [6, Ch. 17]. In practice, the degree of interactivity varies greatly across tools. In some, the user sets the optimisation parameters and the search for the solution is automatic [13, 68, 75, 103]. In others, a stepwise approach is employed, whereby the user may interactively modify the optimisation parameters at each step [58, 64, 65]. Additionally, the granularity of the individual steps is also an aspect that sets interactive methods apart.

A first class of interactive planning tools allows the user to direct the search for a solution by repeatedly tweaking optimisation parameters and solving the modified problem. Generally, the user interacts with the DVH graphs to set optimisation goals, to modify constraints (e.g. dose-volume or simple dose constraints) [10, 16, 104] or to change importance weights of voxels conforming to selection criteria [20]. Although the problem of locality is well understood when using DVH curves, only a number of authors propose tools for direct interaction with the three-dimensional dose distribution. Schlaefer et al. [77] render voxels as spheres, which are colour coded to indicate the amount of dose they receive. Through direct voxel selection, the user may set goals and relax constraints. In [105] we extend this approach, whereby the user interacts with isodose surfaces. Neither of the methods operates at interactive rates (real-time), as reoptimisation is necessary to obtain an updated plan. Kamerling et al. [106] also propose interacting with isodose curves in order to remodel the dose distribution. They achieve real-time interaction through a two step process. First, the beam fluence is adjusted to satisfy the local dose target. Next, the beam fluence is adjusted again, to recover dose changes introduced outside the target area by the initial step. Their approach is not inverse planning, and as such, the solution is not optimal. Furthermore, despite the recovery step, localised

isodose manipulations give rise to non-local effects. The author of [107] takes a similar approach, whereby the *achievable dose estimation* is quickly computed and the bulk of the interaction takes place on the DVH.

The second class of tools enables planners to navigate extensive databases of pre-solved plans. The most advantageous plan is then selected based on a number of clinical criteria [63, 64, 66, 73]. These tools provide the user with a fast overview of each plan, through statistics, DVH curves, and dose renderings.

The third class is that of tools supporting the planner in the execution of discrete planning tasks. Christopher et al. [108] use an autostereoscopic display for organ delineation and plan evaluation in conformal radiotherapy. Olofsson et al. [109] combine three-dimensional visualisation and haptic feedback to support planners in placing treatment iso-centres. Haptic feedback is used to indicate dose constraints in space. Schlaefer et al. [110] use a three-dimensional display to devise heuristics that help them select the most advantageous treatment beams. Geng [111] makes use of a novel three-dimensional display for placing beam isocentres. He posits that such a display increases understandability of the scene, which would ultimately improve plan safety and effectiveness.

Quite a few of the tools we mentioned involve the user interacting with a virtual environment. Research in this field has given birth to a number of complex systems, either for creating virtual reality worlds or for mimicking natural interactions. For example, head mounted displays are used to completely transpose the user into the virtual space. Furthermore, novel input devices, such as six-dimensional mice or gesture recognition systems, allow the user to manipulate the world around him. Most of these systems exhibit a number of drawbacks, such as high complexity and price, low availability or, even, simulator sickness [112]. It is unsurprisingly, then, that despite the promise of improved usability [109, 110], most commercial systems still use conventional displays along with mouse and keyboard input.

3 Mathematical Model

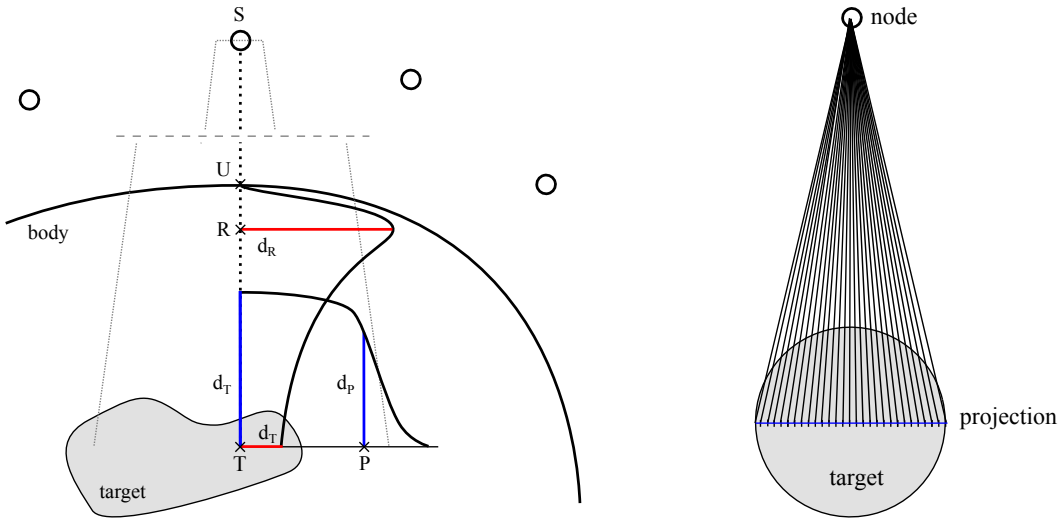
Hypothesis 1 formulated the main objective of this work, that of precisely shaping the distribution. This chapter presents the mathematical model used for spatial dose optimisation, and shows how it embeds within a conventional planning framework. In the first section, an overview of this previously existing framework is given and the optimisation model is detailed. Next, this model is extended to support localised dose planning. Finally, the various user interfaces for the conventional optimisation model are presented and the requirements for a novel planning interface are formulated.

3.1 The Planning Framework

The treatment planning framework is based on the work of Schlaefer et al. [10] and was specifically designed for ACCURAY's CyberKnife delivery system. Architecturally, it is made up of multiple modules, with the most relevant being the optimisation, dose computation, graphical representation, and user interface modules. New modules can be added and plugged into the planning workflow.

3.1.1 The Dose Computation Model

The CyberKnife shapes the radiation beams using either fixed diameter tungsten collimators, or the newer variable diameter *Iris* collimator [113], both with sizes between $5\,\text{mm}$ to $60\,\text{mm}$. Dose deposition is computed from dosimetric measurements in a water phantom. Tissue density inhomogeneities are inferred from CT scans and provide correcting factors for the phantom measurements. The process of computing the dose delivered by a beam to a point P within the patient is sketched out in Figure 3.1a. Beams originate in predefined nodes around the patient and are subject to a number of attenuation effects. The dose is highest along the beam axis, has a peak a few centimetres under the surface of the patient and falls off quickly as the depth increases. This effect is accounted for by the *tissue phantom ratio* (TPR), which is measured in a water phantom: the dose along the beam axis at different depths, relative to the dose at the reference depth (d_R).



(a) Dose computation (b) Beam orientations

Figure 3.1: (a) Computation of the dose deposited by a beam in a point P. The tissue phantom ratio is given by $TPR(T) = \frac{d_R}{d_T}$, while the off-centre ratio is $OCR = \frac{d_P}{d_T}$. The beam originates at a predefined node, S, and enters the body at U. T is a point on the beam axis such that $\overrightarrow{TP}$ is normal to the beam axis. R is the point on the beam axis at the reference depth at which d_R is measured. (b) Simplified sketch of all possible beam orientations starting from one node and directed at a spherical target. Beams are oriented so as to hit any of the voxels belonging to the projection of the target (as seen from the node eye view).

On a plane normal to the beam axis, the dose drops to about $50\,\%$ at a distance equal to half the nominal beam diameter from the intersection of the plane with the beam axis (beam centre, T). To account for the lateral attenuation, phantom measurements are taken on an axial plane at set distances from the beam centre (these measurements are made at multiple depths). The *off-centre ratio* is then defined as the dose measured on the plane, relative to the dose at the beam centre.

The dose deposited in a point P by a beam b shaped by a collimator $\mathfrak{C}$ in one monitor unit is calculated as:

$$\rho_{b,P} = K_{\mathfrak{C}} \cdot TPR_{\mathfrak{C}}(T) \cdot OCR_{\mathfrak{C}}(P)$$

where $K_{\mathfrak{C}}$ is a compensating factor taking into account other effects, such as the inverse square relationship between dose and distance to source [10, Ch. 4]. Both $TPR_{\mathfrak{C}}(T)$ and $OCR_{\mathfrak{C}}(P)$ are interpolated from the tabulated phantom measurements. Finally, the dose

deposited at a point P by N beams is:

$$d(P) = \sum_{i=1}^{N} w_i \cdot \rho_{b_i,P}$$

where each beam is weighted by w_i expressed in monitor units (MUs).

3.1.2 Beam Generation

The optimisation problem operates within a finite search space to find a solution—the larger this space, the longer the optimisation time. In IMRT, a low number of beam orientations, or fields (typically 3 to 10) are used. Fields are divided into rectangular unit subfields, called beamlets; the set of all beamlets represents the search space. The dose deposited by each beamlet into each voxel is computed and the optimisation process resolves the weight of each beamlet [114]. Unlike IMRT, the CyberKnife system delivers its beams from more than 100 nodes sampled around an isocentre. Starting from any node, beams can be directed to any point on the tumour and can be shaped by any one of 12 collimators. Consider the simple case presented in Figure 3.1b, where beams are generated to hit circular target of $50\,\mathrm{cm}^3$: starting from the node, beams are oriented to hit any of pixels the making up the target projection on a plane perpendicular to the line connecting the node and the target centre. This projection covers an area of $16.5\,\mathrm{cm}^2$, which can be coarsely sampled by 64 voxels at a $5\,\mathrm{mm}$ sampling rate. Considering that there are more than 100 nodes and 12 possible collimator sizes, the complete search space would consist of roughly $78\,000$ beams. An optimisation problem of this size is intractable on present hardware or, at best, very slow to find a solution (indeed, even a generous memory availability of $24\,\mathrm{GB}$ limits the beam size to about $10\,000$, using the present framework).

To tackle this issue, one can employ adaptive sampling methods, which can adequately cover the search space with only a fraction of the beams. This is analogous to adaptively sampling the anatomical space to reduce the total number of voxels. The beam generation module of the planning framework uses the methods presented by Schlaefer in [10]. Beams are generated according to a two-dimensional probability distribution functions (PDF), which specifies the probability of a beam being directed at a pixel of the target projection. Different PDFs can be used, depending on the main planning goal: if a good conformality is desired, the PDF would increase the likelyhood of beams hitting the centre of the PTV, rather than the surface; conversely, if homogeneity is more important, beams that target the surface of the PTV would be have to be more likely than beams targeting the centre.

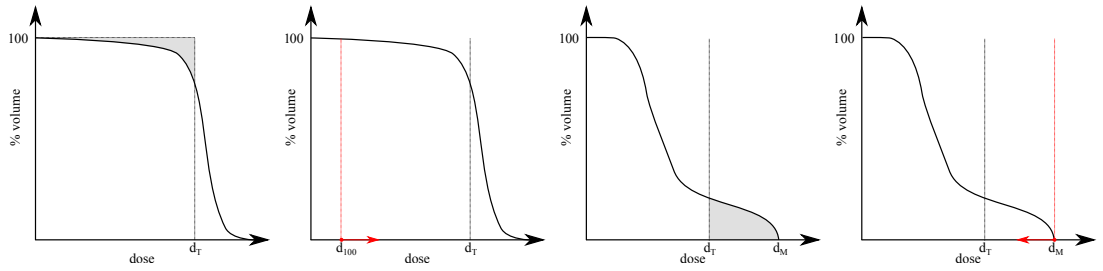
3.1.3 Dose Optimisation

We utilise the optimisation model introduced by Schlaefer et al [58, 77] which solves the linear program described below:

$$
\begin{array}{lllll}
min & \mathbf{c_x^T x}+\mathbf{c_{\hat{s}}^T \hat{s}}+\mathbf{c_{\check{s}}^T \check{s}}+\mathbf{c_{\hat{t}}^T \hat{t}}+\mathbf{c_{\check{t}}^T \check{t}} & & & \\
s.t. & \mathbf{Ax} -\mathbf{\hat{s}} & & -\mathbf{\hat{t}} & \leq \mathbf{b_u} \\
& \mathbf{Ax} & +\mathbf{\check{s}} & +\mathbf{\check{t}} & \geq \mathbf{b_l} \\
& \mathbf{x} & & & \leq \mathbf{b_x} \\
& \mathbf{\hat{s}} & & & \leq \mathbf{b_{\hat{s}}} \\
& \mathbf{\check{s}} & & & \leq \mathbf{b_{\check{s}}} \\
& \mathbf{\hat{t}} & & & \leq \mathbf{b_{\hat{t}}} \\
& \mathbf{\check{t}} & & & \leq \mathbf{b_{\check{t}}} \\
& \mathbf{M\hat{s}} & & & \leq \mathbf{b_{\hat{v}}} \\
& \mathbf{M\check{s}} & & & \leq \mathbf{b_{\check{v}}} \\
& \mathbf{1^T x} & & & \leq b_w \\
& \mathbf{x}, \mathbf{\hat{s}}, \mathbf{\check{s}}, \mathbf{\hat{t}}, \mathbf{\check{t}} & & & \geq \mathbf{0}
\end{array}
\tag{3.1}
$$

where $\mathbf{A} = [a_{ij}], i \in \{1..m\}, j \in \{1..n\}$ is the dose deposition matrix for all n beams in the combined m voxels of all p VOIs. a_{ij} is the amount of dose deposited by beam j in voxel i in one unit of time; $\mathbf{x}$ is the vector of beam weights. As such, $\mathbf{d} = \mathbf{Ax}$ is the dose vector for all voxels and it is lower and upper bounded by $\mathbf{b_l}$ and $\mathbf{b_u}$ respectively. The vectors of slack variables $\hat{\mathbf{s}}$, $\check{\mathbf{s}}$, $\hat{\mathbf{t}}$, and $\check{\mathbf{t}}$ allow for deviation from the bounds on a per voxel, or on a per VOI basis: there are two variables for each voxel ($\hat{s}_i$ and $\check{s}_i, i \in \{1..m\}$) and two variables for each VOI ($\hat{t}_k$ and $\check{t}_k, k \in \{1..p\}$). The slack variables are upper bounded by $\mathbf{b_{\hat{s}}}$, $\mathbf{b_{\check{s}}}$, $\mathbf{b_{\hat{t}}}$ and $\mathbf{b_{\check{t}}}$. The sum of the slack variables is bounded for each VOI by $\mathbf{b_{\hat{v}}}$ and $\mathbf{b_{\check{v}}}$, as $\mathbf{M}$ is a $m \times p$ matrix mapping the voxel indices to VOI indices. The individual beam weights and the total beam weight (total MU) are bounded by $\mathbf{b_x}$ and b_w respectively. The objective function is a sum of simple objectives, where each objective is multiplied by a coefficient: $\mathbf{c_x}$, $\mathbf{c_{\hat{s}}}$, $\mathbf{c_{\check{s}}}$, $\mathbf{c_{\hat{t}}}$, and $\mathbf{c_{\check{t}}}$. This formulation supports the following simple objectives:

- *PTV coverage* through the minimisation of the mean (**OCO**) or of the maximum (**OMI**) deviation from a prescription dose.
- *OAR sparing* through the minimisation of the mean (**OME**) or of the maximum (**OMA**) deviation from a certain dose.
- *Treatment time* (exposure) through the minimisation of the sum of weights for all active beams (**OBW**).



(a) OCO (b) OMI (c) OME (d) OMA

Figure 3.2: DVH illustration of the simple optimisation goals. (a) OCO minimises the area between the DVH curve, the $100\,\%$ volume limit and the target dose d_T limit. The resulting DVH must take shape in the grayed-out area. (b) OMI minimises the difference between the PTV minimum dose (d_{100}—maximum dose covering $100\,\%$ of the PTV), and the target dose, d_T. The red arrow shows the direction in which the minimum dose can change after optimisation. (c) OME minimises the area under the DVH, between the target dose, d_T, and the OAR maximum dose, d_M. The resulting DVH must take shape in the grayed-out area. (d) OMA minimises the difference between the OAR maximum dose, d_M, and the target dose, d_T. The red arrow indicates the direction in which the maximum dose may change after optimisation.

OCO Minimises the total dose deviation from a prescription dose d_T. $T = \{i \in PTV \mid d_i < d_T\}$ is the set of PTV voxels receiving a dose lower than d_T. The upper bound of the slack variables is set to the difference between target dose and current dose, $b_{\check{s}_i} = d_T - b_{l_i}$ and $b_{l_i} = d_T, \forall i \in T$. To minimise the total dose deviation, the sum of all slack variables is optimised by setting $c_{\check{s}_i} = 1$ and all other coefficients to zero. Figure 3.2a shows an example of OCO optimisation.

OMI Minimises the maximum dose deviation from a prescription dose d_T. The minimum dose to a target can thus be increased. With the same definition of T, one sets $b_{\check{t}} = \max_i(d_T - b_{l_i})$ and $b_{l_i} = d_T, \forall i \in T$. The value of the slack variable $\check{t}$ is minimised by setting $c_{\check{t}} = 1$ and all other coefficients to zero. Figure 3.2b shows an example of OMI optimisation.

OME Dual to OCO, decreases the mean dose to a VOI (typically an OAR) by minimising the total deviation from a reference dose d_T. Let T be the set of voxels of an OAR receiving a dose higher than d_T, $T = \{i \in OAR \mid d_i > d_T\}$. The upper bound of the slack variables is set to the difference between the target dose and the current dose, $b_{\hat{s}_i} = b_{u_i} - d_T$ and $b_{u_i} = d_T \quad , \forall i \in T$. Similarly to OCO, the sum of the slack variables

is minimised by setting $c_{\hat{s}_i} = 1$ and all other coefficients to zero. Figure 3.2c shows an example of OME optimisation.

OMA Dual to OMI, decreases the maximum dose to an OAR by minimising the maximum deviation from a reference dose, d_T. Keeping the same meaning of T, one sets $b_{\hat{t}} = \max_i(b_{u_i} - d_T)$ and $b_{u_i} = d_T \quad , \forall i \in T$. The value of the slack variable $\hat{t}$ is minimised by setting $c_{\hat{t}} = 1$ and all other coefficients to zero. Figure 3.2d shows an example of OMI optimisation.

OBW Minimises the sum of all active beams (total MU), by simply setting $\mathbf{c_x} = \mathbf{1}$ and all other coefficients to zero.

3.1.4 Stepwise Planning

In stepwise planning, a single objective is optimised at any one time, by setting the coefficients corresponding to the objective to 1 and the other coefficients to 0. Optimising multiple objectives at the same time is fraught with difficulties. First, having a different number of voxels for each VOI leads to an implicit weighting of each goal. Second, the objective values taken by different goals are not comparable to one another, and are dependent on the number of voxels and on the target doses. Consequently, it is not straight forward to balance the values of the objective coefficients, such that a particular trade-off is optimised. A complete discussion on the topic can be found in [58, 77]. An easier approach is to use a lexicographic ordering approach, whereby a single objective is optimised at any given time. With the previous formulation of the optimisation problem, lexicographic ordering can readily be attempted, as shown in Listing 2.

Algorithm 2 Lexicographic Ordering approach to multicriteria optimisation.

1: $O = \{(f_1, o_1), .., (f_m, o_m)\}$ ▷ Define objectives and target values, sorted from highest to lowest priority
2: $C = \{\mathbf{c}_1, .., \mathbf{c}_n\}$ ▷ Define initial constraint set
3: $i \leftarrow 0$
4: **while** $i < m$ **do**
5: $\quad optimise(f_i)$
6: $\quad$ Convert f_i to constraint and add to C
7: $\quad i \leftarrow i + 1$
8: **end while**

The first optimisation step has to establish an initial dose distribution, and is thus either an OCO, or an OMI step. After defining the objectives, O, and the constraints, C, lexicographic ordering is automatic and results in a single Pareto efficient solution. From a clinical point of view though, such a solution is rarely optimal and multiple trial and error runs are required to arrive at an acceptable plan.

The stepwise approach differs from lexicographic ordering in the fact that the planner has the opportunity to modify optimisation parameters before moving to the next step. If the current objective is not suitably satisfied, relaxing some constraints may improve the objective. Typical relaxations are: PTV lower bound, when the goal is to spare an OAR, or, conversely, OAR upper bound, when the goal is to increase the dose to the PTV. This succession of optimisation and relaxation steps allows the planner to relatively quickly explore the solution space. However, relaxing any constraint destroys the Pareto efficiency of the final solution, as previously optimised criteria may also benefit from the change. Pareto efficiency can be reestablished by reoptimising all criteria in order. In practice however, this results in limited gains to the previously optimised criteria, as typical relaxations are small and most of the slack awarded by relaxation would have already been used up. Stepwise planning is summarised in Listing 3.

Algorithm 3 Stepwise approach to multicriteria optimisation.

1: $O = \{(f_1, o_1), .., (f_m, o_m)\}$ ▷ Define objectives and target values, sorted from highest to lowest priority
2: $C = \{\mathbf{c}_1, .., \mathbf{c}_n\}$ ▷ Define initial constraint set
3: $i \leftarrow 0$
4: **while** $i < m$ **do**
5: **while** $optimise(f_i) > o_i$ **do** ▷ Relax and optimise until satisfied.
6: Select and relax acceptable $\mathbf{c}_k \in C$
7: **end while**
8: Convert f_i to constraint and add to C ▷ Fix solution
9: $i \leftarrow i + 1$
10: **end while**
11: $i \leftarrow 0$ ▷ Reoptimise in order to establish Pareto efficiency
12: **while** $i < m$ **do**
13: $optimise(f_i) > o_i$
14: Convert f_i to constraint and add to C ▷ Fix solution
15: $i \leftarrow i + 1$
16: **end while**

3.1.4.1 Objective Fixation

Goals are fixed by constraining the objective value to be less than, or equal to the optimised value. Within present framework, the following fixations are made:

- **OCO:** $b_{\check{V}}$ is set to the current optimisation value. Alternatively, the lower bound for all target voxels is set to the minimum between target dose and current dose: $b_{l_i} = min(d_i, d_T)$; i runs over all target voxel indices.
- **OMI:** $b_{\check{t}}$ is set to the maximum deviation from the target dose, over all target voxels. Alternatively, the lower bound for all target voxels is set to the same maximum deviation from the target dose.
- **OME:** $b_{\hat{V}}$ is set the current optimisation value. Alternatively, the upper bound for all target voxels is set to the maximum between target dose and current dose: $b_{u_i} = max(d_i, d_T)$.
- **OMA:** $b_{\hat{t}}$ is set to the maximum deviation from the target dose, over all target voxels. Alternatively, the upper bound for all target voxels is set to the same maximum deviation from the target dose.
- **OBW:** b_w is set to the current total MU.

3.1.4.2 Constraint Relaxation

Although any constraints may be changed, typically, only VOI and voxel bounds are relaxed during planning. Beam, node and total MU bounds are set once, prior to the first optimisation step, and kept constant. To relax the lower or upper bounds of a VOI V by a certain amount ρ, we set $b'_{l_i} = b_{l_i} - \rho$ or $b'_{u_i} = b_{u_i} + \rho$ respectively, for all voxels $i \in V$.

For example, typical trade-offs are: relaxation of the PTV lower bound, when the goal is to spare an OAR, or, conversely, the relaxation of the upper bounds for an OAR, when the goal is to increase the dose to the PTV. In some cases non obvious trade-offs also bring an improvement, such as relaxing the upper bound of an organ, when the goal is to spare another organ.

Trade-offs can be partially or fully undone as follows: a new goal is set up to recover the relaxation and the constraints that had been tightened to fixate the previous goal, are relaxed to their original values. Take for example the improvement in PTV coverage achieved by relaxing the upper dose bound of an OAR from b_{u_1} to b_{u_2}, with $b_{u_2} > b_{u_1}$. To once again decrease this upper bound to a value $b_{u_3} \in [b_{u_1}, b_{u_2})$ we simply need to

perform an OME step on the OAR with the target dose $d_T = b_{u_3}$. Very probable, it is necessary to also relax the PTV lower bound close to its previous value, in order to fully satisfy the OME goal. Finally, PTV coverage is reestablished by an OCO step.

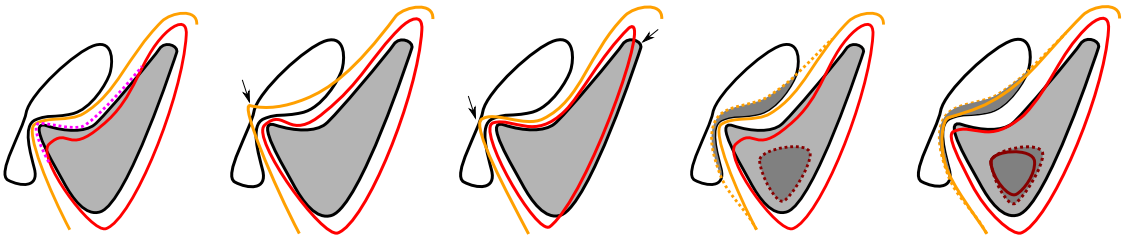
3.2 Modelling the Spatial Dose Optimisation

The optimisation goals used in stepwise planning are defined at the VOI level and consequently suffer from the same limitations discussed in Section 2.3.2.3. Consider the scenario depicted in Figure 3.3a, which studies the trade-off between PTV coverage and OAR sparing. The initial dose distribution (Figure 3.3a) does not adequately cover the PTV; the OAR, however, is very well spared, as the isodose curve representing its upper bound avoids it completely. The first approach to increase PTV coverage consists in relaxing the OAR upper bound and reoptimising for coverage (Figure 3.3b). To avoid major changes to the dose distribution, one could employ *dose-volume constraints* (DVC)[1], instead of relaxing the entire OAR. A DVC specifies which percentage of a VOI is allowed to deviate from a given bound. Yet, this relaxation is also not localised and could still result in a solution similar to the one in Figure 3.3b, in which the upper bound isodose curve cuts through the OAR. Alternatively, one could use a scalarization MCO method with VOI or voxel-wise weights to increase the relative importance of the entire PTV, or of the underdosed PTV voxels, respectively. Since this change in importance factors affects all VOIs/voxels, one could end up with areas of the PTV being uncovered, as their relative importance decreases (see Figure 3.3c).

Neither of these options are intuitive for a human planner, who may have a clear idea of the local dose trade-offs he would be ready to make. Consider then the case presented in Figure 3.3d where the user can directly relax a subset of the OAR voxels by moving the upper bound isodose. Additionally, he may relax the upper bound in a subvolume of the PTV, thus allowing a hotspot to develop. The subsequent optimisation leads to the solution from Figure 3.3e, in which the PTV coverage improves and the VOI and PTV bounds are violated only in the areas of relaxation. In practice, the objective improvement will often be less than the one obtained by any of the previous two methods and one may need to further relax some constraints. At some point, the planner may decide that no other trade-offs are possible, or worth the cost, and the solution will represent the optimal balance between the goals (obviously, one would need to re-establish Pareto-efficiency, yet in practice this changes the solution only slightly).

As previously stated in Hypothesis 1, the precise locality of the goals and trade-offs

[1] DVCs are difficult to use, as there is no direct convex formulation for them [115, Ch. 1].



(a) Initial solution (b) VOI-wide relaxation (c) Goal weight adjustment (d) Local relaxation (e) Result after local relaxation

Figure 3.3: Example of exploring the trade-off between PTV (gray) coverage and OAR (white) sparing. (a) The initial state shows the prescription isodose curve (red) only partially covering the PTV and the isodose curve corresponding to the OAR upper bound value (orange). It would be desirable to achieve a complete coverage of the PTV (dotted magenta curve). (b) Possible solution after relaxing the OAR upper bound (VOI relaxation). As a result, the OAR isodose curve cuts through a sensitive area (arrow). (c) Possible solution after increasing the weight of the PTV coverage objective. Without a careful balancing of the weights, problem areas such as those indicated by the arrows may appear. (d) Our proposed method is to explicitly allow for dose escalation in limited areas within the OAR and within the PTV (dark gray patches, delimited by dotted curves). The dotted curves show the maximum theoretical extent of dose, as a result of relaxation. (e) Subsequent optimisation moves the isodose curves in the areas indicated by the relaxation. Deep the PTV a hot-spot appears (dark red curve). The prescription isodose curve almost fully conforms to the shape of the PTV.

(relaxations) is an additional criteria that we will take into account during planning. To allow for this, next section details how the present framework is extended, by refining the interaction unit from VOI, down to voxel level.

3.2.1 Voxel-wise Optimisation Steps

Consider a voxel discretized space S containing all VOIs. Any voxel $i \in S$ can either be a free voxel or it can belong to one or multiple VOIs. For any set of voxels T representing a subvolume of S, we can pursue one of two goals. First goal is to increase the dose in T to at least d_T. This can be expressed as minimising either the total dose deviation, or the maximum dose deviation of voxels in T. We call these steps *selective coverage* (**SCO**) and *selective maximisation* (**SMI**). Second goal is to reduce the dose in T to at most a value of d_T. Again, this is expressed as minimising either the total dose deviation, or the maximum dose deviation of voxels in T. These steps are called *selective sparing* (**SSP**) and *selective minimisation* (**SMA**). Obviously, SCO, SMI, SSP, and SMA are the voxel-wise analogues of OCO, OMI, OME, and OMA respectively.

Following successful optimisation of any step, the current result must be saved by setting the voxel bounds accordingly. Multiple goal fixation types are possible, depending on the criteria taken into account by the human planner.

3.2.1.1 Selective Coverage (SCO)

Consider an area represented by the target voxel set T, where each voxel receives a dose less than a target dose d_T, $T = \{k \in S \mid d_k < d_T\}$. We initialise the bounds of all slack variables and all objective coefficients to zero. We then set the bounds of the lower slack variables equal to the difference between the target dose and the bound value: $b_{\check{s}_i} = d_T - b_i, \forall i \in T$. We also set the lower bounds of target voxels equal to the target dose: $b_{l_i} = d_T, \forall i \in T$. The optimisation goal is to minimise the sum of all slack variables corresponding to the target voxels, which we achieve by setting $c_{\check{s}_i} = 1$. This results in the following linear program to be solved:

$$\begin{array}{lllll} min & \sum\limits_{i \in T} \check{s}_i & & & \\ s.t. & \mathbf{Ax} & & \leq \mathbf{b_u} \\ & \mathbf{Ax} & +\check{\mathbf{s}} & \geq \mathbf{b_l} \\ & \mathbf{x} & & \leq \mathbf{b_w} \\ & \check{\mathbf{s}} & & \leq \mathbf{b_{\check{s}}} \\ & \mathbf{x}, \check{\mathbf{s}} & & \geq \mathbf{0} \end{array} \tag{3.2}$$

3.2.1.2 Selective Maximisation (SMI)

Considering the same definition for T as in Section 3.2.1.1, we attempt to increase the minimum dose in T to d_T as follows. We initialise the bounds of all slack variables and all objective coefficients to zero. We set the bound of the slack variable $\check{t}$ to the maximum deviation from the target dose $b_{\check{t}} = \max\limits_{i \in T}(d_T - b_{l_i}), \forall i \in T$ and we update the target voxel bounds equal to the target dose, $b_{l_i} = d_T, \forall i \in T$. The optimisation goal is to minimise the value of the variable $\check{t}$ by setting $c_{\check{t}} = 1$.

$$\begin{array}{lllll} min & \check{t} & & & \\ s.t. & \mathbf{Ax} & & \leq \mathbf{b_u} \\ & \mathbf{Ax} & +\check{\mathbf{t}} & \geq \mathbf{b_l} \\ & \mathbf{x} & & \leq \mathbf{b_w} \\ & \check{\mathbf{t}} & & \leq \mathbf{b_{\check{t}}} \\ & \mathbf{x}, \check{t} & & \geq \mathbf{0} \end{array} \tag{3.3}$$

3.2.1.3 Selective Sparing (SSP)

We redefine T to represent an area in which any voxel receives a dose greater than a target dose d_T: $T = \{k \in S \mid d_k > d_T\}$. We initialise the bounds of all slack variables and all objective coefficients to zero. We then set the bounds of the upper slack variables equal to the difference between the bound value and the target dose: $b_{\hat{s}_i} = b_{u_i} - d_T, \forall i \in T$. We also set the upper bounds of target voxels equal to the target dose: $b_{u_i} = d_T, \forall i \in T$. The optimisation goal is to minimise the sum of all slack variables corresponding to the target voxels, and we achieve this by setting $c_{\hat{s}_i} = 1$. This results in the following linear program to be solved:

$$\begin{array}{lllll} min & \sum\limits_{i \in T} \hat{s}_i & & & \\ s.t. & \mathbf{Ax} & -\hat{\mathbf{s}} & \leq \mathbf{b_u} \\ & \mathbf{Ax} & & \geq \mathbf{b_l} \\ & \mathbf{x} & & \leq \mathbf{b_w} \\ & \hat{\mathbf{s}} & & \leq \mathbf{b_{\hat{s}}} \\ & \mathbf{x}, \hat{\mathbf{s}} & & \geq \mathbf{0} \end{array} \tag{3.4}$$

3.2.1.4 Selective Minimisation (SMA)

Using the same definition for T as in Section 3.2.1.3, we attempt to decrease the maximum dose in T to d_T as follows. We initialise the bounds of all slack variables and all objective coefficients to zero. We then set the bound of the slack variable $\hat{t}$ to the maximum deviation from the target dose $b_{\hat{t}} = \max\limits_{i \in T}(b_{u_i} - d_T), \forall i \in T$ and we update the target voxel bounds equal to the target dose, $b_{u_i} = d_T, \forall i \in T$. The optimisation goal is to minimise the value of the variable $\hat{t}$ by setting $c_{\hat{t}} = 1$.

$$\begin{array}{lllll} min & \hat{t} & & & \\ s.t. & \mathbf{Ax} & -\hat{\mathbf{t}} & \leq \mathbf{b_u} \\ & \mathbf{Ax} & & \geq \mathbf{b_l} \\ & \mathbf{x} & & \leq \mathbf{b_w} \\ & \hat{\mathbf{t}} & & \leq \mathbf{b_{\hat{t}}} \\ & \mathbf{x}, \hat{t} & & \geq \mathbf{0} \end{array} \tag{3.5}$$

3.2.2 Constraint Relaxation

Similarly to the previous approach, we will be coerced to perform some trade-offs in order to suitably satisfy the optimisation goals. Indeed, this is all the more likely, as

previously initialised optimisation problems leave little to no slack for the new goals (see discussion in Section 2.3.2.2). Consider a set of voxels R. To relax the upper bound, we set $b'_{u_i} = b_{u_i} + \rho_i, \forall i \in R$, where $\rho > 0$ is the relaxation amount. Equivalently, to relax the lower bounds, set $b'_{l_i} = b_{l_i} - \rho_i, \forall i \in R$ and $\rho > 0$.

3.2.3 Steps for Optimising the Spatial Dose Distribution

The optimisation steps presented in previous sections can be chained into a workflow to achieve a Pareto efficient solution, similar to the one presented in Listing 3. Localised dose optimisation is used to interactively tweak the dose distribution as detailed in Listing 4.

Algorithm 4 Stepwise optimisation of the spatial dose distribution. $pop(O)$ removes the first entry in O and returns **null** of O is empty; **while** stops on **null** argument. $\mathbf{convertToObjective}(\mathbf{R})$ converts saved relaxation R to an optimisation objective, to recover this relaxation.

$O = \{\emptyset\}$ ▷ Objective set is initially empty
$C = \{\mathbf{c}_1, .., \mathbf{c}_n\}$ ▷ Define initial constraint set
$R = \{\emptyset\}$ ▷ Relaxation set is initially empty
while $\exists$ opportunity for local dose change **do**
    Manually define f, local dose objective ▷ SCO, SMI, SSP, or SMA
    while $optimise(f)$ unsatisfactory **do**
        Select $\mathbf{c} \in C$ localised binding constraints
        $R = R \cup \{\mathbf{c}\}$ ▷ Save relaxation
        $relax(\mathbf{c})$ ▷ Perform local relaxation
    end while
    $fix(f)$ ▷ Fix current objective
    $O = O \cup \{f\}$ ▷ Add to objective set
end while
$O_r = convertToObjective(R)$ ▷ Generate objectives to recover relaxations
$O = O \cup O_r$
while $f = pop(O)$) **do** ▷ Establish Pareto efficiency
    $optimise(f)$
    $fix(f)$
end while

3.3 User Interfaces to the Optimisation Model

The planning framework is implemented in *Java* and uses the *IBM ILOG CPLEX* library to model and solve the optimisation problem. Two planning graphical user interfaces, Workflow based (WF) and Interactive DVH (IDVH), are used to carry out lexicographic ordering and stepwise planning respectively. Both are built with the Java SWING widget toolbox and use the Java2D and Java3D libraries to draw DVH curves and three-dimensional representations of the VOIs, beams, and patient anatomy.

As a result of the work presented here, a third user interface—*Interactive3D* (I3D)—implementing spatial dose planning, was created using Java SWING for the controls (buttons, sliders, lists, etc.) and Java3D and The Visualization Toolkit (VTK) for visualisation and interaction with the virtual scene. Ending this chapter, an overview is given of the two existing planning interfaces, while the methods employed for I3D planning are discussed in detail in the chapters that follow.

3.3.1 Workflow Based Planning (WF)

Computational units that perform anything from basic functions, such as file reading and writing, to more complex functions, such as dose computation or solving a linear program, are represented as *tasks*. Tasks are configurable through input and output parameters. Multiple tasks can be chained together into *workflows* to provide more complex functionality. Workflows are recursive: they can contain other workflows as tasks and, thus, make use if the more complex functionality.

A GUI allows the user to create workflows by dragging and dropping tasks into the work field. The order in which the tasks are chained together determines their run order. Figure 3.4 shows a workflow used to setup and solve an OCO optimisation problem.

Workflows can be seen as scripts, running all the task in a continuous, non-interactive manner. From a user's point of view, this planning abstraction is suited for lexicographic ordering planning. One needs only to setup a workflow that reads the plan data and runs the various optimisation steps in order. The final solution, as discussed earlier, is Pareto-efficient.

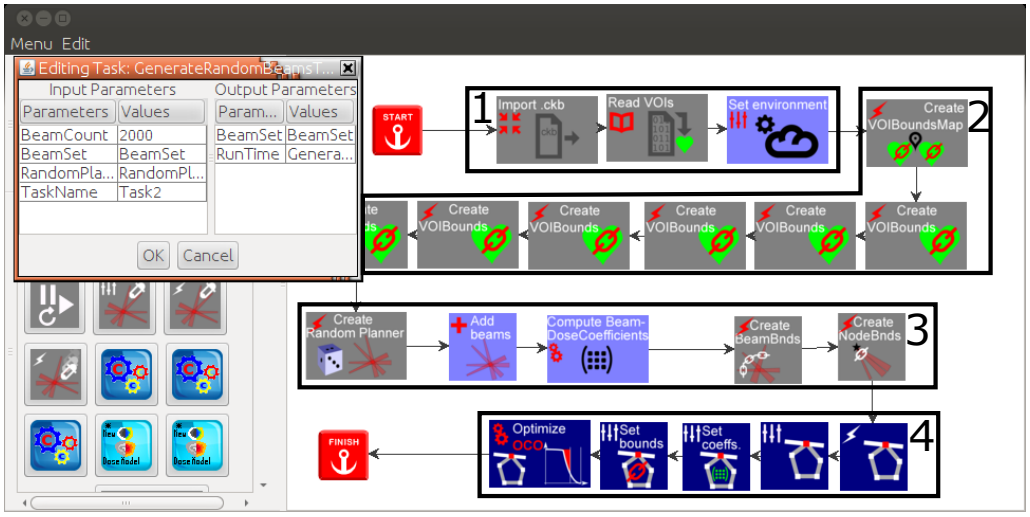


Figure 3.4: Example of a workflow to run an OCO optimisation step. Tasks are represented by icons and are executed in the order shown by the arrows. Each task has an input and an output parameter set, respectively (upper left window). The tasks group 1 loads the patient and therapy system data. Group 2 sets up the bounds for all VOIs. Group 3 generates the search space (optimisation beamset), computes the dose deposition matrix A and sets up miscellaneous beam data. Group 4 sets up the optimisation using the beam and bounds data and solves it.

3.3.2 Interactive DVH Planning (IDVH)

The Interactive DVH Planning GUI is used to implement the stepwise optimisation method. The user is presented with the DVH curves corresponding to the current solution. Goal selection and relaxation are performed by direct interaction with the DVH curves. For example, to set up a new goal, the user drags the vertical target bar to the desired dose position for the target VOI. Basic constraint relaxation is performed by dragging the vertical bounding bars in the direction of relaxed values. Underneath, all lower or upper voxel bounds belonging to that particular VOI V are updated. If the desired relaxed bound indicated by the user is b, then all voxel bounds are set to this value: $b'_{l_i} = b, \forall i \in V$ if lower bound relaxation is performed, and $b'_{u_i} = b, \forall i \in V$ if upper bound relaxation is performed, with $b < b_{l_i}$ or $b > b_{u_i}$ $\forall i \in V$, respectively. Obviously, relaxing all voxels of V is potentially destructive for any previous criteria involving the VOI.

For VOIs that have been targets of optimisation steps, the bound will be represented by a bounding curve that conforms to the DVH for all dose values worse than the target dose (see Figure 3.5). The user can then manually define a relaxed bounding curve starting from the original curve. One method is to shift the entire curve in the direction

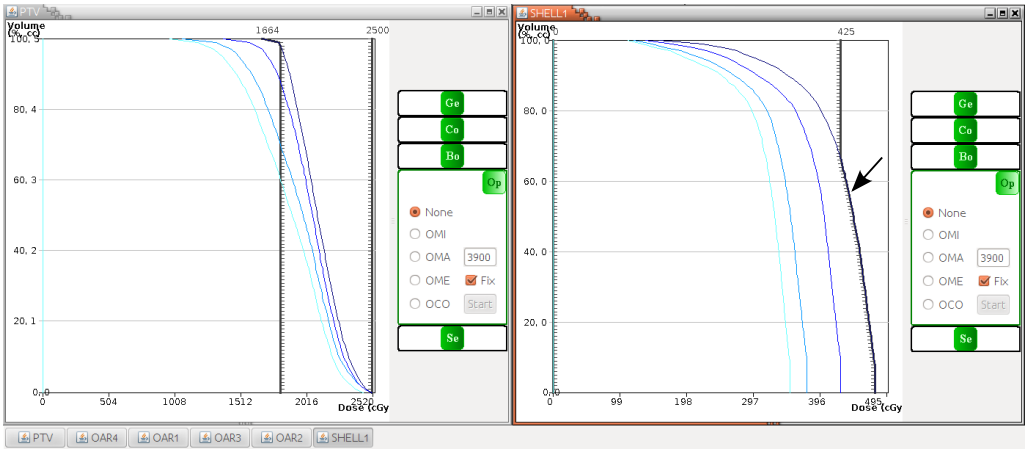


Figure 3.5: The IDVH GUI for planning a simplified meningioma case. The shell (right) upper bound was successively relaxed from $3.5\,\mathrm{Gy}$ (light blue) to $4.75\,\mathrm{Gy}$ (dark blue) leading to a substantial improvement in PTV coverage (left). The bounding curves partially mould to the DVH curves as the optimisation objectives are turned into constraints (arrow).

of more relaxed values. This is equivalent to relaxing all voxels of V by the amount ρ the curve was shifted. The second method is to shape the bounding curve to allow more relaxed values. To relax using the lower bounding curve, the user starts at a bound value b_1 on the curve and drags left to $b_2 < b_1$. The lower bound is set to $b_{l_i} = b_2, \forall i \in R_l$, where $R_l = \{i \in V \mid b_{l_i} \leq b_1 \text{ and } b_{l_i} > b_2\}$. To relax using the upper bounding curve, the user drags right from an initial bounding value b_1 to $b_2 > b_1$, which results in $b_{u_i} = b_2, \forall i \in R_u$, where $R_u = \{i \in V \mid b_{u_i} \geq b_1 \text{ and } b_{u_i} < b_2\}$.

3.3.3 Requirements for Interactive 3D Planning (I3D)

The main requirements of a planning tool implementing stepwise spatial optimisation are the subject of the next three chapters:

1. **Visualisation:** provides the planner with a view of the patient anatomy, planning structures (PTV, OARs, shell, etc.), and optimisation problem (current dose distribution, optimisation target, and limiting structures).

2. **Interaction:** translates manipulation of the visual elements into optimisation operations: setting targets, or performing relaxations.

3. **Computation runtime:** ensures user feedback in as short as possible time, for a good quality of interaction.

3.3.4 Synergic Planning

Neither of the three planning interfaces is the be-all end-all solution to the complex problem of treatment planning. Indeed, one of the aims of this work is to formulate a planning approach, in which all three tools are used conjunctively to achieve plans of higher quality than it would otherwise be possible, if one were to plan using only one of them. Conceptually, it is easy to grasp the advantages and disadvantages of each of the interfaces, as listed in Table 3.1. WF is a script-like planning approach: once the initial parameters are set, no user interference is needed. With IDVH, the user can tweak the optimisation problem after each step, allowing for greater flexibility. Finally, with I3D the user shall navigate a three-dimensional scene and shall modify the optimisation in much more granular steps than with both WF and IDVH.

Table 3.1: Overview of the main advantages and disadvantages of the three planning tools: WF, IDVH, and I3D.

	WF	IDVH	I3D
Planning effort	low	medium	high
Automatic	yes	no	no
Localised in 3D	no	no	yes
Flexibility	low	medium	high

Considering these characteristics, we propose a synergic approach to planning which exploits the advantages of each of the tools (see Figure 3.6). A first step establishes an initial solution in a completely automatic manner (WF). Next, the planner uses IDVH to adjust the plan in such a way that gains are achieved with reduced effort. Finally, the user analyses the three-dimensional dose distribution and operates local changes using the I3D tool, to further improve the plan.

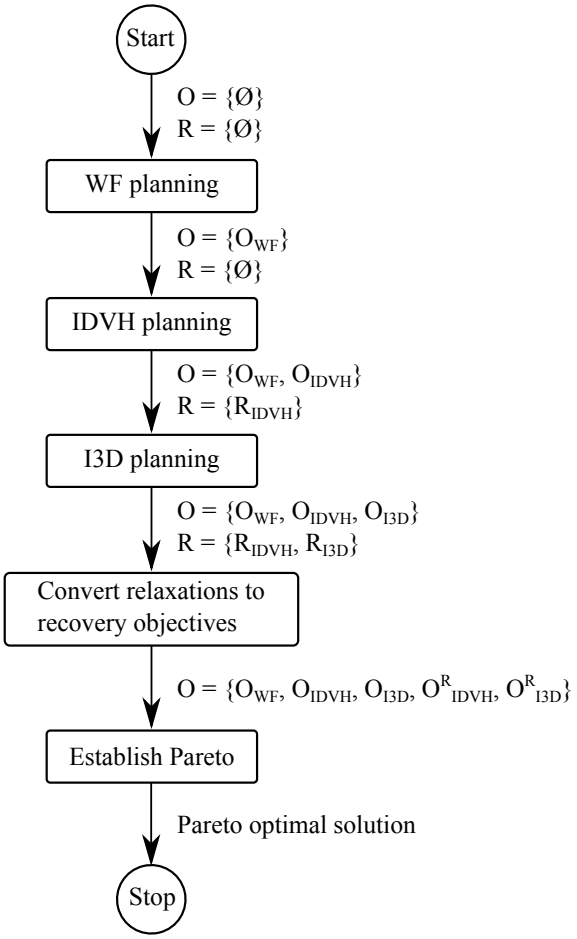


Figure 3.6: Synergic use of the three planning interfaces, WF, IDVH, and I3D to establish a Pareto optimal solution. Initially, the objective set, O, and the relaxation set, R, are empty. After each planning module, the goals and relaxations that were performed are added to these sets. Finally, the relaxation sets R_{IDVH} and R_{I3D} are converted to optimisation goals, O^R_{IDVH} and O^R_{I3D}, to attempt to recover any available slack. The last step is automatic, and establishes Pareto efficiency of the solution.

4 Visualising the Optimisation Problem

This chapter presents the methods used to visualise the optimisation model presented earlier. For each voxel in the planning scene, a number of scalar data are associated to it: current dose, upper and lower bounds, target dose (if the voxel is part of the goal), and CT Houndsfield value. To present these data to the user, the I3D tool uses isosurface extraction (see Section 2.5). The choice to use isosurfaces, instead of direct volume visualisation, is due to a number of factors. First, planners are used to visualise isodose surfaces when assessing the quality of a plan [116, 117]. Second, virtual surface manipulation is intuitive for most users, as there are a lot of real life analogies: for example deforming a balloon, by expanding or contracting it. Third, the rendering of surfaces as solid objects results in decreased visual complexity of the virtual scene.

Isosurfaces are extracted using the VTK implementation of the Marching Cubes algorithm, *Marching Tetrahedra* [102]. For a given scalar field and a threshold τ, the τ-valued isosurface partitions this field into areas with values larger than τ and areas with values lower than τ.

Hypothesis 2. *The optimisation problem can be completely visually represented through combined use of isodose surfaces, target surfaces and bounding surfaces which correspond to the current solution, the optimisation objective, and the voxel constraints respectively.*

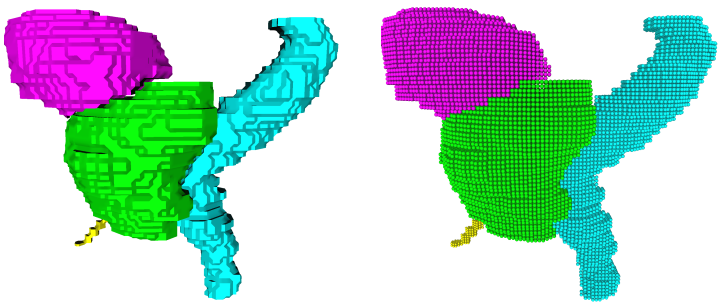
The following sections present the visualisation of the different graphical elements and make use of the common notations: S is the set of all voxels discretizing the box-shaped region of interest (ROI), which contains all p VOIs; each VOI V_k, discretized by m_k voxels, is contained within it's own bounding box. Both ROI and each VOI define their own local coordinate systems: $\mathbf{cs_S}$ and $\mathbf{cs_{V_k}}$.

4.1 Volumes of Interest

To extract the outer surface of a VOI V represented in the voxel discretized space S we use the following function to create a binary scalar field:

$$v(k) = \begin{cases} 1 & \text{if } k \in V \\ 0 & \text{if } k \notin V \end{cases} \tag{4.1}$$

The isosurface extracted for $\tau \in (0, 1]$ corresponds to the surface of the VOI. An alternative representation, is to render each voxel of the VOI as a simple geometric shape, such as a sphere [110]. This latter approach allows for precise visual inspection down to voxel level. Note that sparsely discretized VOIs, such as shells, are difficult to render accurately as surfaces using the method described above. For these VOIs, the voxel-wise representation is especially useful.



(a) Surface representation (b) Voxel representation

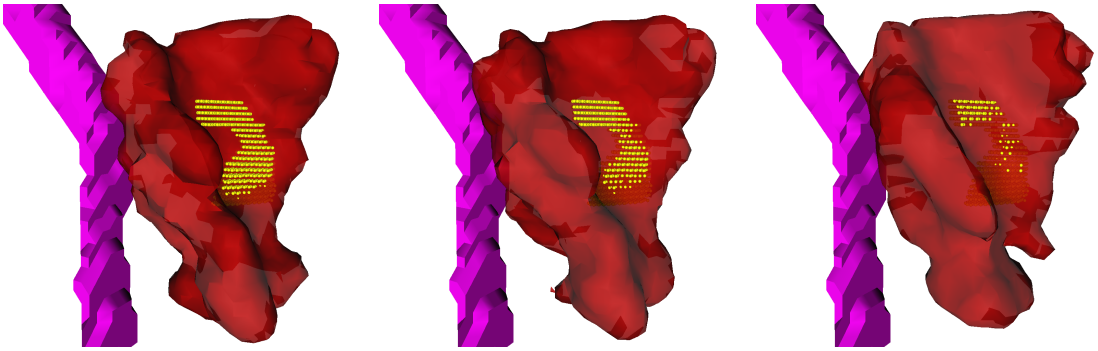
Figure 4.1: VOIs rendered as surfaces (a) and as voxels (b), for a prostate cancer case: prostate (PTV, green), rectum (cyan), bladder (magenta), and urethra (yellow). Rendering VOIs as individual voxels is more precise, but also more complex.

Both surface and voxel VOI representations can be seen in Figure 4.1, which shows screenshots taken during the planning of a prostate cancer case. The size of the individual voxels can be varied, to allow the user to see deep inside the VOIs. If the VOIs are very large, voxel wise rendering may be impractical due to the complexity of the scene.

4.2 Isodose Surfaces

The three-dimensional dose distribution is visualised as isosurfaces corresponding to the dose values of interest d. These *isodose* surfaces are simply extracted from the dose scalar field. By overlaying the isodose and VOI surfaces one can visually inspect the quality of

the current solution, such as coverage of the target, or sparing of the sensitive structures. For example, Figures 4.2 and 4.3 show isodose surfaces expanding during planning to fulfil the optimisation goal.



(a) Initial (b) During planning (c) Final result

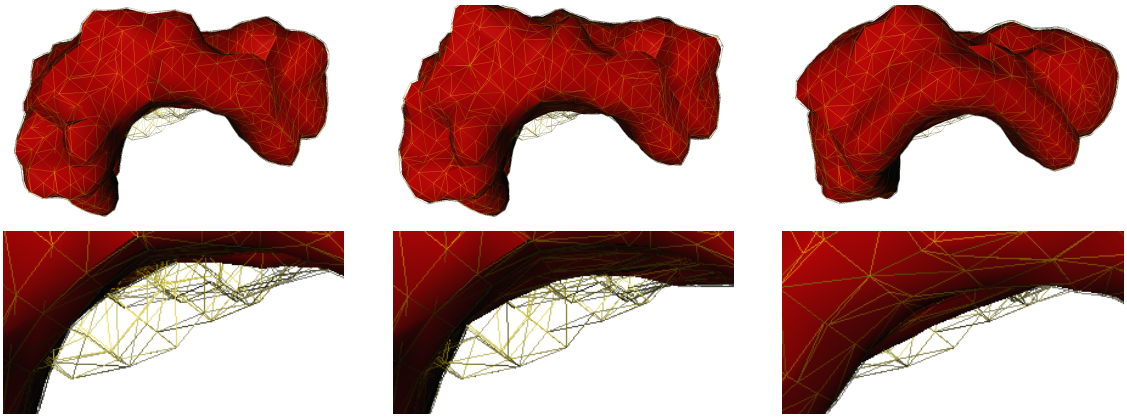
Figure 4.2: The evolution of the isodose surface (red) as it expands during planning to cover a target area containing the boost volume (yellow spheres); the esophagus is rendered as a magenta surface.

4.3 Target Surfaces

Local dose targets correspond to one of two scenarios, of either covering or uncovering a specific area with dose. For a target dose d_T and a set of target voxels T, the following indicator function $t(k)$ assigns a value for each voxel:

$$t(k) = \begin{cases} d_T & \text{if } k \in T \\ d_k & \text{if } k \notin T \end{cases} \tag{4.2}$$

where d_k is the dose received by voxel k. The target surface is the d_T-thresholded isosurface, which is extracted from the scalar space created by t. It is this surface, which the user shapes to define the optimisation goal. The spatial relationship between the d_T-thresholded isodose and the target surface gives a visual feedback on the progress of planning: the more the two surfaces coincide, the better the current solution. Figure 4.3 shows an isodose surface gradually conforming to the target surface, as a result of planning progress.



(a) Initial (b) During planning (c) Final solution

Figure 4.3: The isodose surface (red) as it evolves to conform to the shape of the target surface (wireframe, yellow) during three stages of planning. Second row provides an enlarged view of the area of interest.

4.4 Limiting Surfaces

Consider a d thresholded isodose surface and the corresponding target surface. We are interested in visualising the spatial constraints that prevent the two surfaces from coinciding.

Definition 4.1: Limiting voxels

A voxel is said to be lower limiting relative to a dose value d if and only if its lower bound b_l is greater than the reference dose: $b_l > d$. Analogously, a voxel is upper limiting relative to d if and only if its upper bound b_u is less than the reference dose: $b_u < d$.

By applying the indicator function $b_d(k)$ defined below, a binary scalar field is obtained, for which values of 1 signify limiting voxels. To visualise the limiting structures formed by the bounds of these voxels, one only has to extract the τ thresholded isosurface, for $\tau = 1$.

$$b_d(k) = \begin{cases} 1 & \text{if } k \text{ is limiting relative to } d \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

The proximity of the d-thresholded isodose surface to the d-valued lower or upper bounding surface gives a measure of how limiting certain constraints are. The closer the isodose surface is to the bounding surface, the more limiting is that bounding surface.

4.5 Colourmapping

Due to low depth perception, it is difficult to assess the proximity of three-dimensional objects on a two-dimensional display. While traditional cues, such as perspective projection or shading, do offer a sense of depth to the images, the absence of steropsis is an additional limiting factor [118]. When visualising dose distributions, the problem is compounded by the fact that in areas of high dose gradient, large dose variations happen within very small distances. For example, the d-valued isodose surface and one of its limiting surfaces may seem close to each other but, very often, they almost have to touch for the limiting surface to actually obstruct the evolution of the isodose surface. This problem is illustrated in Figure 4.5a, where it is difficult to discern which areas of the lower bounding surface are limiting (from their proximity to the dose surface). To tackle this problem *colourmapping* is used to highlight these areas (see Figure 4.5b, where limiting areas are in red). Colourmapping, is the process of assigning a colour to a point in the virtual space, based on a property measured at that point (CT value, dose, distance to bound, etc). Additional information, such as PET uptake values, can be easily integrated into the virtual planning environment by defining new colourmaps.

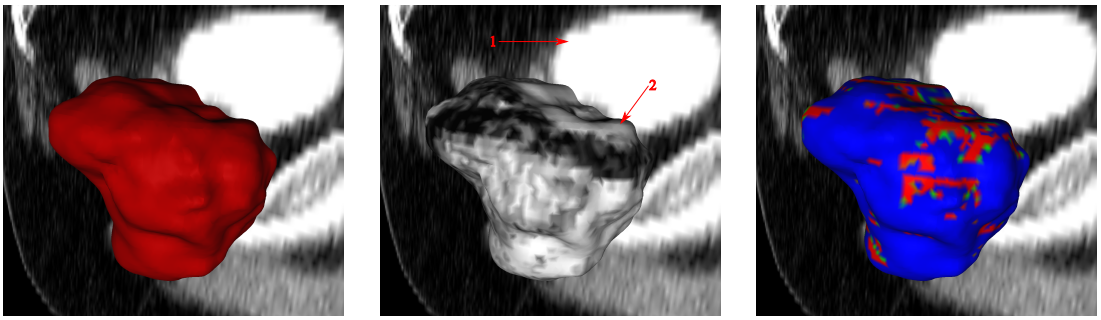
Interactive3D uses two types of colourmaps: *position colourmaps* and *value colourmaps*. A position colourmap assigns a colour to a point P in space. A value colourmap assigns a colour to a scalar value based on a mathematical function. For example, such a colourmap would be a linear mapping of the values in the $[0, 100]$ interval to a gray scale, with values lower than or equal to 0 being mapped to black and values greater than or equal to 100 being mapped to white. Position colourmaps use value colourmaps to assign a colour value to a point in space. The I3D tool allows the use of three value colourmaps: uniform (one colour), grayscale (black to white), and jet (blue to red). Figure 4.4 shows how colourmaps give different insight into a scene, although the objects remain the same.

4.5.1 CT Colouring

The CT colourmap looks-up the CT density value measured at a position P and converts it to a grayscale colour value. Additional conversion maps can easily be implemented.

4.5.2 Dose Colouring

The dose colourmap converts the dose value of the voxel corresponding to position P and converts it to a jet colour value. The highest dose within the scene is mapped to red while the lowest dose is mapped to blue.



(a) Uniform (b) Grayscale—CT (c) Jet—lower limiting

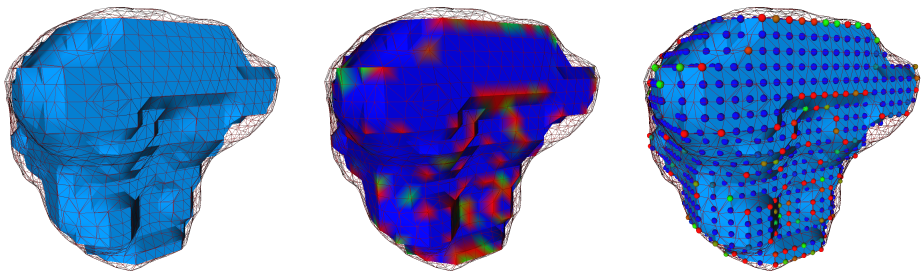
Figure 4.4: Isodose surface rendered with three colourmaps: uniform (a), grayscale (b), and jet (c). In the background, a sagittal CT slice is also rendered to give context to the scene. In (b), the CT values are mapped to shades of gray. Arrow (1) shows the location of the bladder on the CT slice; arrow (2) shows the bladder CT values mapped onto the dose surface. The planner can have an idea of the structures affected by that particular dose. In (c), red indicates lower limiting areas, in which the dose is close to the PTV lower bound. Notice that one may conclude, that some of these regions belong to the bladder, as they are depicted as such in (a). However, due to the fact that the PTV and the bladder overlap, these areas belong exclusively to the PTV during planning.

4.5.3 Constraint Colouring

When considering potential relaxations one only regards those areas that are limiting. For a point P in space, a colour in the inverted jet colourspace is assigned based on the dose distance to bound (i.e. the absolute difference between the dose and either the lower, or the upper bound at P). For a distance of zero (i.e. the dose is equal to the bound value) the voxel is considered to be maximally limiting, and it is assigned the colour red. The maximum difference is set by the user to correspond to "not limiting" and maps to the colour blue. Figure 4.5 shows a rendering of the $36.25\,\mathrm{Gy}$ isodose surface encasing the corresponding lower bounding surface. Without constraint colouring, the limiting areas are difficult to discern in (a), whereas in (b) and (c) they become evident.

Note that the distance to bound is not a linear description of *how limiting the voxel is*, although the jet colourmap may give this impression to some users[1]. Indeed, it is impossible to predict how much more the objective value will improve as a result of re-

[1]The jet colourmap itself suffers from a number of shortcomings: it does not have a natural perceptual ordering of the colours, it obscures details (e.g. green and cyan are difficult to distinguish) while overemphasising other features, etc. [119]. For the planning use-case however, the jet colourmap easily serves its purpose.



(a) (b) (c)

Figure 4.5: The lower bounding surface (blue) corresponding to the shape of the PTV and restricting the isodose surface (wireframe, transparent red) from shrinking into it. The limiting areas, where the isodose surface comes in the proximity of the bounding surface, are highlighted in (b) as red, but very hard to discern in (a). (c) Alternatively, one can view the limiting PTV voxels by colour-coding them directly.

laxing a "more limiting" voxel than another "less limiting" voxel. Additionally, relaxing a maximally limiting voxel may result in no improvement in the objective value whatsoever. This is due to the complex coupling of constraints, where other voxels may be limiting the optimisation. It is therefore futile to define a degree of limitation for a voxel. The colouring simply highlights potential limiting areas and offers cues on the remaining dose slack.

4.6 Additional Visualisation Elements

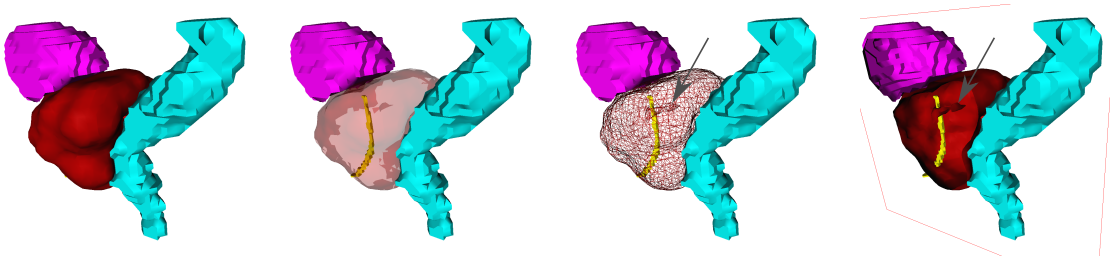
This section presents to two additional visualisation options that are available to the user during planning: visualisation of the patient anatomy, through slicing, and simplification of the virtual scene when complexity is an issue.

4.6.1 Slicing

The volumetric CT data is rendered into the scene through the cutting plane method as three orthogonal planes, corresponding to the coronal, sagittal and axial slices, and a fourth arbitrarily oriented plane. The decision to use slicing rather than direct volume rendering was mainly for convenience. The implementation is straight forward and details are visible on the full resolution images. Furthermore, planners are used to evaluating CT data as slices rather than volumes. Lastly, slices do not overly complicate the scene, as would a complete volume. The obvious trade-off in this case is the loss of three-

dimensional information, as the slices have to be repositioned in order to make areas of interest visible. We consider this problem to be mitigated by the other visualisation options, such as projection of the CT information on scene objects though colourmapping.

4.6.2 Reducing Visual Complexity

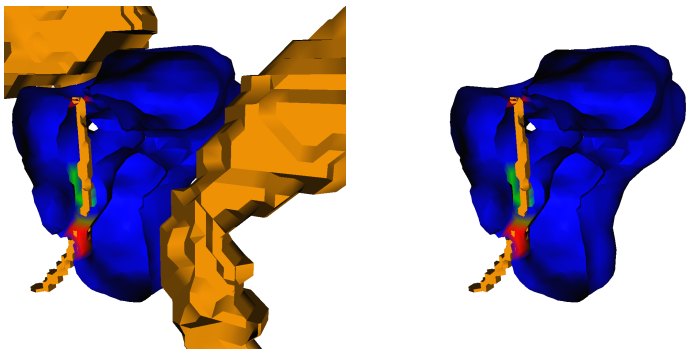


(a) Normal rendering (b) Transparent isodose surface (c) Wireframe isodose surface (d) Cutting plane rendering

Figure 4.6: Visualising the treatment isodose surface for a prostate cancer case in the context of the OARs: rectum (cyan), bladder (magenta), and urethra (yellow). Depending on the visualisation options, the coldspot indicated by the arrow may not be visible. (a) The isodose is depicted as a solid surface. (b) The isodose is drawn transparent, but because the composing triangles are not depth sorted, the rendering is incorrect. The cold area is practically invisible. (c) With a wireframe representation of the surface, the cold area (arrow) is visible, yet the rendering is cluttered. (d) An x-axis aligned cutting plane (red square) simplifies the scene and makes the interior of the surface as well as the cold area (arrow) visible.

As seen in Figure 4.6b, Interactive3D supports transparency to display multiple overlapping objects. This gives good results for a limited number of simple objects, yet it is not feasible for many or complex objects (e.g. self-intersecting surfaces). Special techniques, such as illustrative rendering, can be used to emphasise shape through shading [120], and thus create simplified yet highly suggestive depictions. Implementation of both illustrative rendering and correct transparency requires low level access to the rendering pipeline, where one has to define a custom shader program. For example, in the case of transparency, primitives must be displayed in *back-to-front* order from the camera when using the standard method of alpha-blending. Sorting the primitives in real-time can only be achieved if implemented on the GPU, as it is computationally expensive. An alternative to alpha-blending, *depth peeling*, also requires low level access to the pipeline [121, 122], which is unusual for a high level library such as Java3D. Due to the fact that shader programming was never fully supported and documented in Java3D, these techniques were not implemented in the planning tool.

One alternative to transparency, is to display surfaces as wireframes, which makes their interior visible (see Figure 4.6c). However, the number of overlapping objects has to remain small, to keep the scene understandable. If the scene does grow complicated, the user may limit the rendering to a smaller area of interest. This is achieved through the use of clip planes—the scene is only rendered on one side of the plane, while the other side is completely discarded. Graphical elements can thus be cut through, to reveal their insides. This is the case in Figure 4.6d, where a clip plane aligned with the x-axis has been translated, along this axis, to cut through the VOIs.



(a) $\tau = \infty$ (b) $\tau \in (0, 1)$ Gy

Figure 4.7: Visualising the $38\,\mathrm{Gy}$ upper limiting surface (orange) for a prostate cancer case. (a) Without thresholding, the surfaces of the rectum ($37\,\mathrm{Gy}$ upper bound) and bladder ($33\,\mathrm{Gy}$ upper bound) are also visible show. (b) Setting the selector threshold $\tau \in (0, 1)$ Gy, only the limiting surface corresponding to the urethra ($38\,\mathrm{Gy}$ upper bound) is visible. The $38\,\mathrm{Gy}$ isodose is also shown, colour coded to emphasise the areas where the dose is close to the bound.

To simplify the scene further, a selector threshold τ is used when extracting the limiting surfaces. The indicator function takes the form from Equation 4.4. Only those voxels, for which the dose distance to bound is less or equal to τ, are displayed as limiting. This is because typically, only bounds that are relatively close to the reference dose are relevant for relaxation.

$$b_{d,\tau}(k) = \begin{cases} 1 & \text{if } k \text{ is limiting relative to } d \text{ and } |b_k - d_k| \leq \tau \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

Take the example in Figure 4.7, where the urethra, rectum, and bladder have an upper bound of $38\,\mathrm{Gy}$, $37\,\mathrm{Gy}$, and $33\,\mathrm{Gy}$ respectively. If one is interested in analysing which areas are upper limiting of dose greater than $38\,\mathrm{Gy}$, the $37\,\mathrm{Gy}$ and $33\,\mathrm{Gy}$ bound voxels will be included, cluttering the scene (Figure 4.7a). Very often, the planner will only be

interested in those bounds that are close to the value of the isodose (i.e. limiting structures to which the isodose surface is allowed to get near to). In this example, she may simply set τ to a value in $(0, 1)$ Gy, thus ignoring the bounding surfaces formed by the rectum and bladder (see Figure 4.7b).

5 Direct Optimisation Interactions

Previous chapter detailed the graphical representation of the optimisation problem. This chapter presents the user interaction metaphors, mapping interactions with the virtual scene to common optimisation operations: setting the optimisation target, constraint relaxation, and goal fixation.

Hypothesis 3. *Localised goals and trade-offs can be expressed as interactions that remodel the various graphical elements represented by isosurfaces.*

Clearly, interacting in a virtual environment only makes sense in the context of localised goals and relaxations. However, non-localised operations, such as setting VOI-wide goals, can be mixed with localised operations. An example would be a scenario in which the planner desires to increase PTV coverage, while considering localised relaxations. Although it is possible to manually select the goal of covering the entire PTV, this would be an exercise in futility, as such actions are standard practice and do not require interaction with a virtual scene.

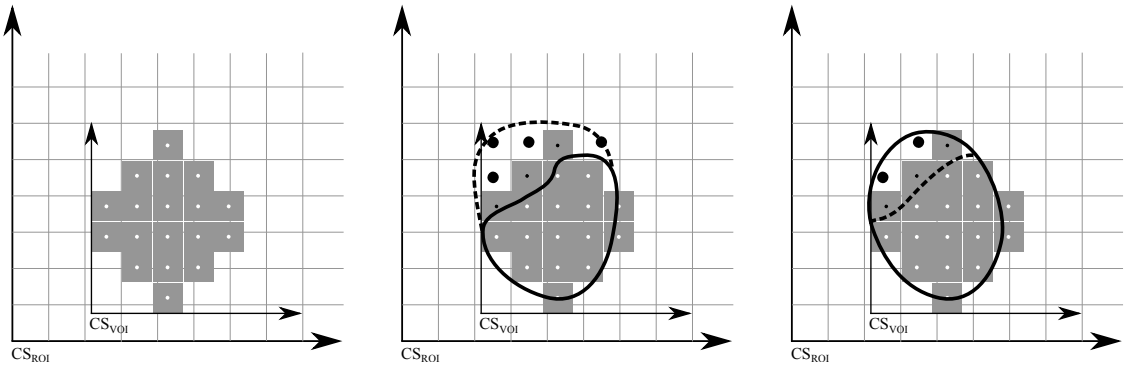
5.1 Interactions as Optimisation Operations

Interactive3D is built around classical mouse and keyboard input, in which the user controls a virtual sphere-shaped bumper that clings to the objects in the scene. The sphere indicates the area of influence of the tool and its radius can be varied as required. Operation is triggered by mouse clicks and, as a result, the subject of interaction is expanded or clipped to the shape of the bumper, or its colour changes accordingly. Wide areas may be covered by allowing the bumper to paint a trail as it is dragged across the surfaces. All the while, the camera may be freely rotated or translated.

5.1.1 The Interaction Grid

Each VOI is discretized on its own grid, with the origin coinciding with the origin of the VOI bounding box—the smallest cuboid containing the VOI. Each VOI voxel enters the optimisation, but to be able to control the dose shape outside the VOIs one must be able

to introduce new voxels. Therefore, a common interaction grid is used (see Figure 5.1a), which coincides with the visualisation grid and, which covers the entire region of interest (ROI). Corresponding to the grid, an artificial grid VOI is introduced, which can be resized by adding and removing voxels during planning; initially, the grid VOI is empty.



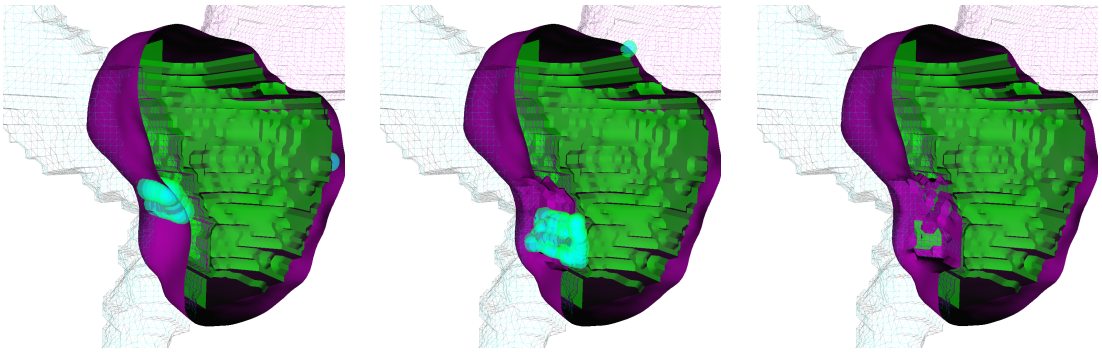
(a) Interaction and VOI grids (b) Target: dose expansion (c) Target: dose contraction

Figure 5.1: (a) The interaction grid defined at the origin of the ROI and a single VOI (discretized on its own grid). (b) Locally extending the coverage of the dose surface (continuous curve) by expanding the target surface (dashed curve). To approximate the target surface, new voxels are added (large black dots) to the grid VOI and enter the optimisation. Together with the small black dots, they represent the optimisation targets. (c) Sparing an area of the VOI by pushing the target curve (dashed curve) inwards. Two new voxels (large black dots) are added to the grid VOI and enter the optimisation.

5.1.2 Optimisation Target Selection

With a high degree or granularity, most optimisation goals can be represented through one of two complementary operations: increasing or decreasing dose coverage in a given area. The direct interaction metaphor of the former is to expand the respective target surface to cover the target area. Likewise, the interaction metaphor of the latter is to clip (contract) the respective target surface, to uncover the target area. Consider the situation shown in Figure 5.1b: a d_0 isodose surface insufficiently covers a critical area of the PTV. One can pull the d_0 target surface, and the newly covered voxels become targets of either an SCO, or of an SMI optimisation step.

Conversely, if it is advantageous to reduce the dosage in a specific area, the user will push the target surface to avoid that area, as sketched out in Figure 5.1c. The uncovered voxels then become targets of either an SME or of an SMA optimisation step. An actual



(a) (b) (c)

Figure 5.2: Steps in subtractive sculpting of the target surface (magenta) in order to spare a sensitive area of the rectum (partially transparent, left of image) for a prostate cancer case. Also visible are the prostate (green) and the bladder (partially transparent, right of image). The interaction bumper in (a) and (b) is used to select the target area. As a result, the target surface contracts (in (b) and (c)), to partially show the surface of the PTV in (c).

example is given in Figure 5.2: the planner uses the bumper to carve out the target surface, which has the initial shape of the isodose surface corresponding to the target dose.

If the user interacts beyond the limit of the VOIs, new voxels need to be introduced into the optimisation (as part of the grid VOI) to accurately represent the shape of the target surface. To ensure feasibility, the voxel bounds (lower or upper bounds, depending on the optimisation goal) are set equal to the target dose and the bound slack is set equal to the difference between the target dose and the actual dose. For example, consider a new voxel P currently receiving a dose d_P less than a target dose d_T. P enters the optimisation by setting $b_{u_P} = \infty$, $b_{l_P} = d_T$ and the associated lower bound slack $b_{\check{s}_p} = d_t - d_P$. Obviously, the new voxels increase the optimisation problem size, which in turn will increase the computational cost. Furthermore, the optimisation model has to be rebuilt every time new voxels are added, which is itself a costly operation.

5.1.3 Constraint Relaxation

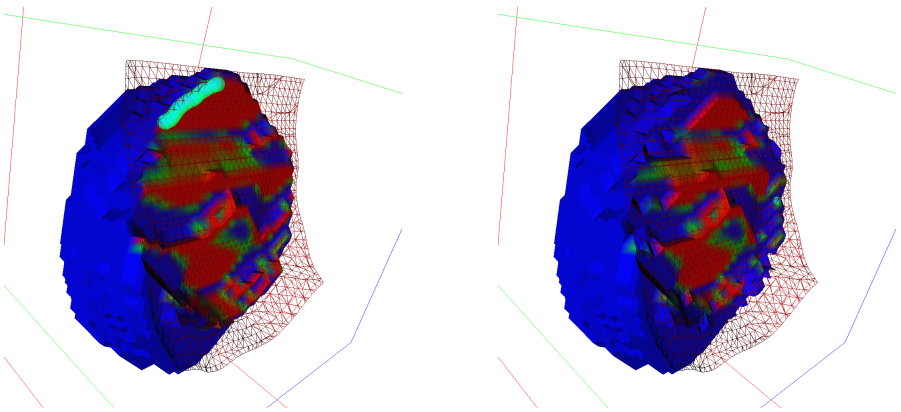
Voxel constraint relaxation can be understood as allowing dose to areas where either the lower or the upper bounds previously prohibited this. Consider a reference dose d and the existence of voxels with at least two bound values, b_l and b_u, such that $b_l > d$ and $b_u < d$. These bounds limit the d isodose surface from shrinking or extending beyond them and, as discussed in the previous chapter, the structures they form can be visualised as lower and upper bounding surfaces, respectively: the b_l lower bounding surface and

the b_u upper bounding surface. Let L_l and L_u be the set of all voxels covered by the lower and upper bounding surfaces respectively.

5.1.3.1 Sculpting the Bound Surfaces

The interaction metaphor of relaxing the lower bound is that of either pushing the isodose surface towards the lower limiting surface, or of directly sculpting the lower limiting surface itself. Let $R_l \subseteq L_l$ be the set of all voxels uncovered by either pushing or sculpting the respective surfaces. Then, for all voxels $i \in R_l$ we relax by the amount $\rho_i = b_{l_i} - d$ to set the lower bound $b'_{l_i} = d$.

Similarly, we relax the upper bounds by either pulling the isodose surface or by sculpting the upper bounding surface. Let $R_u \subseteq L_u$ be the set of newly covered/uncovered voxels by the modified isodose/upper bounding surface respectively. Then, for all $i \in R_u$ we relax the upper bound by $\rho_i = d - b_{u_i}$ and the new bound is $b'_{u_i} = d$. Figure 5.3 shows an example of upper bound relaxation, where the planner sculpts out areas of the limiting surface formed by the voxels of an OAR.

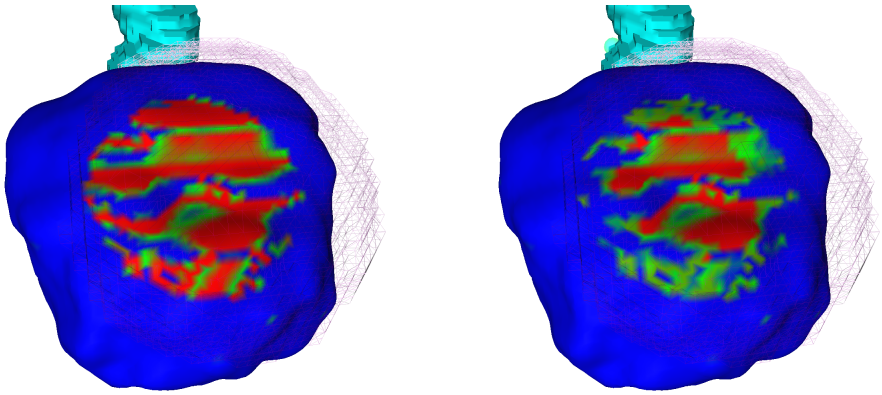


(a) Relaxation selection (b) Relaxation performed

Figure 5.3: Relaxation of the bladder upper bound by sculpting the upper bounding surface (filled; limiting areas are in red). In the limiting areas, the isodose surface (red, wire-frame) comes in close proximity to the upper bounding surface. (b) Following relaxation, the upper bounding surface is sculpted, thus allowing the isodose surface to extend into the newly freed space.

5.1.3.2 Direct Relaxation Selection

Interactive3D offers another possibility of performing trade-offs, by selecting the relaxation targets directly on the isodose surface itself. Consider the $b_l + \delta$ valued isodose surface, where δ is a small threshold. Voxels with lower bounds less than b_l and receiving a dose d such that $d - b_l \leq \delta$ are considered to be limiting. We colour the isodose surface to highlight the lower limiting areas and we directly select the ones we consider for relaxation using the virtual bumper. The relaxation amount can be fixed to a certain value ($\rho = \rho_0$) or can be computed to relax all selected voxels to a certain bound b_l': $\rho = b_{l_i} - b_l$, where b_{l_i} is the current lower bound of each selected voxel i. The colouring of the isodose surface changes accordingly. A similar approach is required to relax the upper bound of limiting voxels. Direct relaxation is shown in Figure 5.4: the limiting areas, shown in red, correspond to voxels belonging to the OAR; after relaxation, the surface colouring changes to indicate that the area of relaxation is now less limiting.



(a) Before relaxation (b) After relaxation

Figure 5.4: Relaxation of the bladder (transparent, wireframe) upper bound by direct selection on the isodose surface which is coloured so as to highlight the limiting areas as red. (b) After relaxation the selected areas show up as less limiting (green).

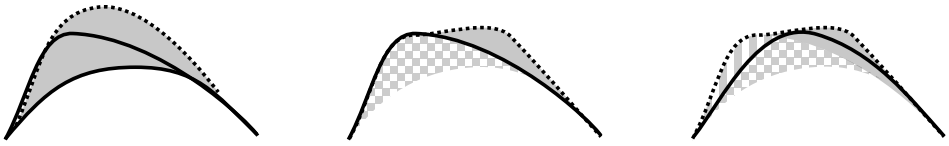
5.1.4 Objective Fixation

The shape of the isodose surface represents a criterion in the proposed optimisation approach. To fix the current goal, after it has been partly or completely satisfied, the bounds of the target voxels are tightened similarly to the method presented in Section 3.1.4.1. Let T be the set of all voxels for which a dose d_T was targeted. Depending on the optimisation step, we set the dose bound for each target voxel i as follows:

$$\begin{array}{ll} \text{SCO} & b_{l_i} = \min(d_T, d_i) \\ \text{SSP} & b_{u_i} = \max(d_T, d_i) \\ \text{SMI} & b_{l_i} = d_T - \max\limits_{i \in T}(d_T - d_i, 0) \\ \text{SMA} & b_{u_i} = d_T + \max\limits_{i \in T}(d_i - d_T, 0) \end{array} \tag{5.1}$$

5.2 Multicriteria Optimisation with Localised Goals

Each voxel in the target set T is in itself a simple criterion to be optimised. Within a purely LO framework, one would need to prioritise the voxels and optimise for each of them individually. Obviously, this is not practical, and clustering all voxels together in T implies no explicit prioritisation. Satisfying all voxel criteria is equivalent to achieving the criterion of desired local dose distribution shape. Very often though, a subset of voxels of T will not reach their target doses. Indeed the planner has no control over which voxels of T are supposed to achieve their goal first. But if such a control is desired, the planner can simply fix the current solution and target a subarea of the original target in a subsequent optimisation step.



(a) First achieved goal

(b) Goal fixation and selection of new goal

(c) Second achieved goal

Figure 5.5: Example of increasing the interaction granularity for finer local dose control. (a) First, the user pulls the target surface (dotted) away from the current isodose surface (lower, solid), thus defining the goal area (gray). The resulting isodose surface (extended, solid) partially conforms to the target surface. (b) The result is fixed (chequered area) and the user defines a new goal (gray area). (c) To partially satisfy the new goal, the user sacrifices fractions of the original result (striped area). The resulting isodose curve (solid black) shifts to partial cover the new target area.

Figure 5.5 sketches out such an example, in which an initial target (dotted line) is only partially satisfied (a) by the resulting expanded isodose curve. In (b), the current result is fixed (chequered area) and a new goal is defined as a subarea of the original target (dotted line). This new goal is partially satisfied in (c), by trading off fractions of the original result (striped area).

6 Sampling Issues

In the optimisation model from Section 3.1.3, each voxel contributes two rows and two columns to the linear program matrix (corresponding to the two bounds and two slack variables respectively), while each beam adds one new row and one new column (corresponding to the beam variable and beam weight bound respectively). Keeping in mind that the complexity of the simplex algorithm ranges from polynomial to $\mathcal{O}(n^2)$ for the worst-case [123], this chapter presents the sampling approaches implemented in Interactive3D, for keeping the optimisation problem size as small as practically possible. There is an inherent trade-off in decreasing either the anatomical or the beam space sample numbers, as coarsely sampled problems lead to inferior solutions. As such, sampling approaches have to balance optimisation runtime against quality of the solution.

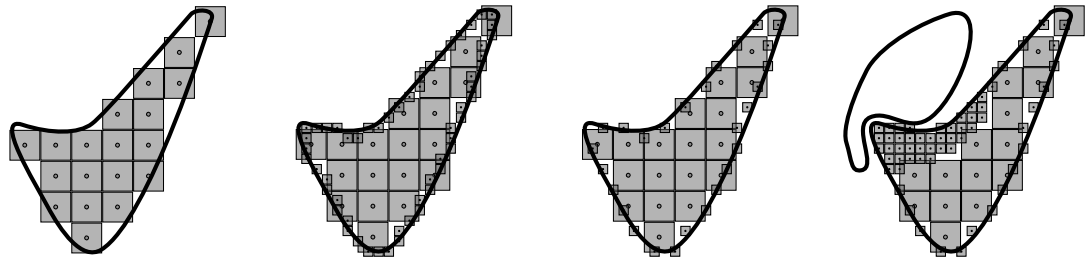
6.1 Anatomical Space Sampling

Note that the aim of the I3D tool is to enable exploration of the spatial trade-offs, by continuously modifying the optimisation problem. Approaches which adapt the sampling strategy to the current optimisation goal, would change the formulation of the trade-offs for each new goal. They are, as such, of no practical use for I3D. Accordingly, we shall look into adaptive sampling approaches that are invariable to the optimisation problem.

Hypothesis 4. *Sparse sampling of the patient anatomy decreases the optimisation runtime, while aliasing errors can be kept within limits by increasing the sampling rate in areas of high dose gradients.*

6.1.1 Two Resolution Approach

Following Hypothesis 4, we use a two-level sampling strategy for representing VOIs: areas that are likely to be exposed to high dose gradients are more finely discretized than the other areas. For OARs, one expects large dose gradients mainly at the interface with VOIs, for which the dose bounds differ by large amounts. It is then sufficient to accurately represent the surface of the organs only in these areas, while using coarser sampling for the rest.



(a) Uniform sampling (b) Fine surface sampling (c) Sparse surface sampling (d) Adaptive sampling

Figure 6.1: Different sampling strategies for a PTV. (a) Coarse sampling leaves areas uncovered, most evident on the edges. (b) Coarse sampling and fine sampling of the PTV surface. The number of PTV surface samples is comparable to the total number of coarse samples. (c) Coarse sampling and sampling of the PTV surface with a sparsity of 1 - every neighbouring sample is discarded. The total number of samples decreases. (d) Adaptive sampling in the presence of another VOI: fine sampling for the area within a certain distance to the VOI, sparse sampling for the surface and coarse sampling for the interior of the PTV.

For PTVs, besides the interfaces to OARs, one can expect high dose gradients (hotspots and coldspots) to appear deep within the volume as well. This is due to the large number of incidental beams which overlap mainly in the PTV. Also note that the entire surface of the PTV needs to be accurately represented in order to ensure proper target coverage. The various discretization options for target volumes are detailed in Figure 6.1: fine sampling of the entire VOI surface, preferential sparse sampling of the VOI surface, and preferential sparse sampling of the VOI surface with fine sampling of the interface areas. Note that the surface of a VOI requires a large number of sample points due to its sheer size. Indeed, this is a well known effect to treatment planners, as extending the margins of an CTV even by a small amount, to create the PTV, can greatly increase the total volume of the VOI. However, the surface may be sufficiently represented through only a sparse set of points lying on it, as shown in Figure 6.1c. For this approach, a *sparsity* factor is defined: the larger the sparsity the coarser the sampling. For example, a sparsity of 2 implies that every first two neighbours of a pixel along each coordinate are discarded.

VOIs represented as unions of surface and interior sub-VOIs are a good match for the step-wise framework, in which VOIs are the planning unit. Comparatively, I3D interactions, for which the voxel is the planning unit, cover decidedly smaller areas. These areas require a high spatial resolution to allow for precise dose shape manipulation. Consider a VOI discretized as surface and interior sub-VOIs. Typically, limiting voxels are found near the surface and, if one performs relaxations in this area, the extent of dose change may be a lot greater than desired, as the interior is coarsely represented. It is for this

reason that we prefer fine homogeneous sampling in areas that are likely to be subject of interactions. Consider a VOI V, discretized on both a coarse (V^l) and on a fine (V^h) uniform grid respectively; its rind (surface), S_V, is sparsely discretized. If V is an OAR, it is represented as the union of the following voxel sets:

$$\begin{aligned} OAR^l &= \{i \in V^l \mid i \notin S_V\} \\ OAR^h &= S_V \cup \{i \in V^h \mid d(i, PTV) \leq l_T\} \end{aligned}$$

The value l_T it the distance threshold within which the OAR is finely discretized. Basically, high resolution discretization is used in areas in which the OAR is close to the PTV.

If V is itself a PTV, a fine representation is used for all areas that are close to the OARs (V_j) (see Figure 6.1d):

$$\begin{aligned} PTV^l &= \{i \in V^l \mid i \notin S_V\} \\ PTV^h &= S_V \cup \{i \in V^h \mid d(i, V_j) \leq l_t, \forall \text{ s.t. } V_j \neq V\} \end{aligned}$$

6.1.2 Hotspot Error Compensation

Large dose variations can lead to hotspots appearing between sampling points. Such is the case deep inside the PTV, where the beams converge. This effect is largely attributed to the lateral beam profile, which exhibits a steep fall-off due to collimation, as explained in Figure 3.1a. Figure 6.2 shows how two beams can overlap to form a hotspot just between two sampling points.

For some cancer cases, homogeneity is a clinical criterion and, thus, hotspots within the PTV cannot be tolerated. An immediate solution would be to use a finer resolution when sampling the PTV interior (which we already know to be computationally taxing). Alternatively, we can estimate the probability of overdosage as a function of voxel upper bound, and then adapt the PTV constraints accordingly. The method of error compensation by estimating the overdosage *probability mass function* (PMF) was published in [124].

6.1.2.1 Estimating the Probability of Overdosage

The aim to estimate the reduction of the overdosage error over the entire PTV as a function of upper bound variation. First, we estimate the probability of a hotspot appearing between eight sampling points, arranged on a regular grid. We create a supersampled cubic VOI with the side length l equal to the PTV voxel length (Figure 6.3a). For a given beamset, let D_c be the maximum dose over the eight corners, and D_i the maximum dose

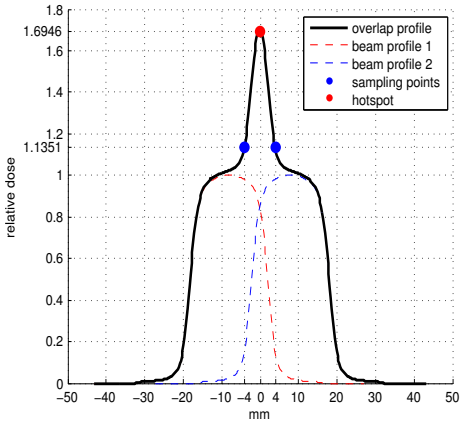
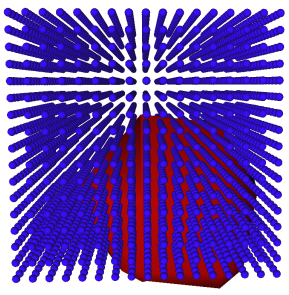


Figure 6.2: Overlap of two normalised lateral beam profiles showing the formation of a hotspot. Consistent with [79], the OCR is responsible for the largest dose variation per unit of distance, in comparison with the other beam effects, such as Tissue-Phantom Ration (TPR), which slowly varies with depth.

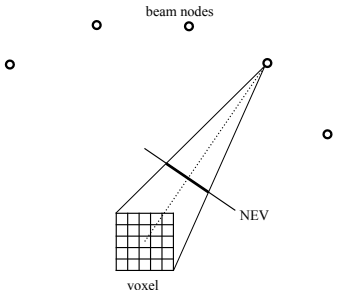
measured inside the cube, including the edges but excluding the corners. The hotspot error is defined as:

$$e_H = \begin{cases} D_i - D_c & \text{, if } D_i \geq D_c \\ 0 & \text{, otherwise} \end{cases} \tag{6.1}$$

For a given cancer case, the cubic VOI is placed at the centre of the original PTV and it is exposed to randomly generated sets of beams, as shown in Figure 6.3a. For each of these beamsets, the e_H is measured and an estimation of its PMF is built.



(a) Hotspot within cubic VOI



(b) Beam generation

Figure 6.3: (a) An $8\,\mathrm{mm}$ voxel modelled as a finely sampled cubic VOI. The reference doses are taken in each of the eight corners. The isodose surface (red) depicts the hotspot that formed between the sampling points. (b) Random beam generation: beams originate in any of the nodes around the patient and are directed based on a node eye view of the target (perspective projection of the target on a plane at a fixed distance, see [10]).

When generating the beams, there are a number of factors to take into consideration which impact the hotspot error. First, the cardinality of each beamset must be equal to the typical number of weighted beam, and the beams must be formed by collimators with radii that are typical for the given case. Second, the beams originate from any of the nodes used by the machine (distributed uniformly) and are targeted towards the cube, such that each beam delivers at least some measurable dose to the target. Specifically, the beam ends are selected uniformly from the projection (uniform or perspective) of the target onto a plane at a fixed distance from the beam node, as shown in Figure 6.3b.

As the cubic VOI is relatively small, we can compute the hotspot error for a large number of beam configurations. This allows us to build a good estimation of the PMF $f_H(x) = P(e_H = x)$. The cumulative distribution function (CDF) of the hotspot errors is then the probability that the hotspot error is less than a value x:

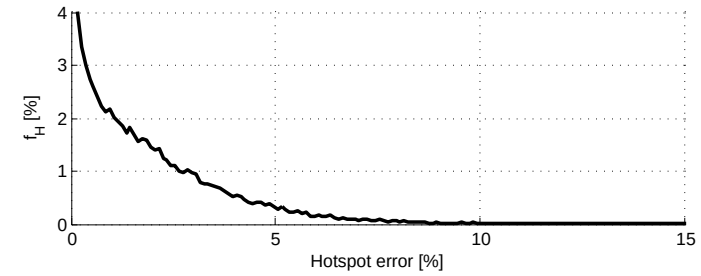
$$F_H(x) = P(e_H \leq x) = \sum_{x_i \leq x} f_H(x_i) \tag{6.2}$$

The complementary cumulative distribution function (CCDF) is the probability of the hotspot error being larger than a value x. It is this probability, given by $\overline{F}_H(x)$, that we desire to reduce:

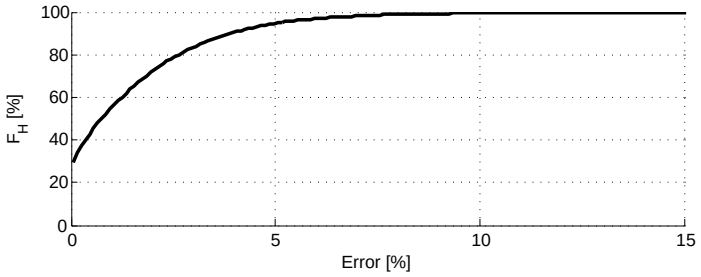
$$\overline{F}_H(x) = P(e_H > x) = 1 - F_H(x) \tag{6.3}$$

As an example, Figure 6.4 shows the resulting PMF, CDF and CCDF for an $8\,\text{mm}$ voxel. $\overline{F}_H(x)$ is expressed for any value of the voxel upper bound. If we were to reduce the reference bound, the entire CCDF would be shifted to the left and the probability of the hotspot error relative to the original bound would be reduced accordingly. More specifically, in the previous example, if the upper bound is reduced by $5\,\%$, the origin of the new CCSF within the old axes would be $-5\,\%$. Hence, the probability that hotspots of at least $5\,\%$ of the initial bound appear would be reduced by about $90\,\%$.

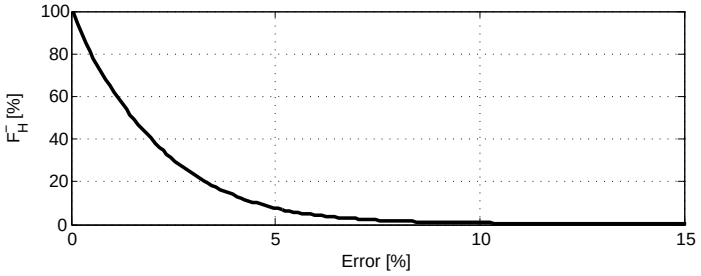
Assuming the hotspot error probability is the same across all voxels of a VOI, it is reasonable to expect that the incidence of overdosage drops accordingly, when tightening the VOI upper bound. This overdosage is expressed as the percentage of the VOI receiving a dose larger than the optimisation bounds.



(a) PMF



(b) CDF



(c) CCDF

Figure 6.4: The measured hotspot error PMF (a), CDF (b), and CCDF (c) for a voxel of size of $8\,\mathrm{mm}$. The CCDF can be used to estimate the reduction in overdosage as the upper bound is tightened.

6.2 Concurrent Beam Space Resampling

To further reduce the computation time, beamset cardinality must be kept as small as possible. As this impacts the solution, it is of particular interest to try to recover the solution quality. Consider first the serial algorithm by Schweikard et al. [39] shown in Listing 1. At each iteration, the objective has a chance to improve, as N_r randomly selected non-weighted beams are replaced by N_r randomly generated beams. Because the optimisation starts from a good solution, a new solution is found in seconds rather than minutes. If the old beams can be replaced with better than randomly generated beams, the objective improvement may be accelerated. In using a heuristic to rank the beams from most to least useful, the IBR algorithm (see Section 2.4.2) does exactly that.

Our approach is to run multiple threads in parallel, each executing the same modified version of the RBR algorithm (also in Section 2.4.2) and thus taking advantage of the multicore architecture of current CPUs. This method has been published in [125]. On each iteration, a thread finds a solution (*internal solution*) to the optimisation problem and it presents it to the other threads as the tuple (B^{sol}, o^{sol}): weighted beams and associated objective value. A thread accepts a solution as its *external solution* if and only if it is better than previous external solution. On the next iteration, the thread compares its internal solution with the current external solution. If the external solution is better, the beams defined by the external solution are inserted into the optimisation. Additional beams B^n are generated such that the optimisation beamset remains N_r: $|B^{ext}| + |B^n| = N_r$.

Hypothesis 5. *Solution beams, found by parallel running optimisations, lead to a better improvement in the objective value than randomly generated beams. Thus, concurrent beam resampling leverages the architecture of multiprocessor systems to reduce computation time.*

Obviously, the solution can only improve from one iteration to another, for the same reasons for which RBR works. To see this, consider the following linear program in standard form for which, without any loss in generality, the total beam weight is minimised:

$$\begin{aligned} min \quad & \mathbf{1^T x} \\ s.t. \quad & \mathbf{A}^i\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \quad \geq \mathbf{0} \end{aligned} \tag{6.4}$$

Where $\mathbf{A}^i = [\mathbf{a}_1^i..\mathbf{a}_n^i]$ is the coefficients matrix of the i-th LP written out in column form, $\mathbf{A}_i \in \mathbb{R}_+^{m\times n}$, $\mathbf{x} \in \mathbb{R}_+^n$ is the vector of variables and $\mathbf{b} \in \mathbb{R}^m$ is the constraint vector. Mathematically, beam resampling is done by simply replacing columns of $\mathbf{A}^i$. Suppose that $\mathbf{x}^I$ and $\mathbf{x}^E$ are the interior and the exterior solutions, respectively, to two LPs corresponding to the coefficient matrices $\mathbf{A}^I$ and $\mathbf{A}^E$, and that $\mathbf{1}^T\mathbf{x}^E < \mathbf{1}^T\mathbf{x}^I$ ($\mathbf{x}^E$ is a superior

solution to $\mathbf{x}^I$). Without loss of generality, let the non-zero entries of both $\mathbf{x}^I$ and $\mathbf{x}^E$ be the first p and q entries respectively:

$$\mathbf{x}^I = [x_1^I \,..\, x_p^I \, 0 \,..\, 0]$$
$$\mathbf{x}^E = [x_1^E \,..\, x_q^E \, 0 \,..\, 0]$$

According to the modified resampling algorithm, we build the hybrid LP containing the columns corresponding to the non-zero entries of the two solutions:

$$\mathbf{A}^c = [\mathbf{a}_1^I \,..\, \mathbf{a}_p^I \, \mathbf{a}_1^E \,..\, \mathbf{a}_q^E \, \mathbf{a}'_1 \,..\, \mathbf{a}'_r]$$

The entries $\mathbf{a}'_i$, with $i \in \{1..r\}$, belong to new beams being added to the LP, such that $p+q+r = n$. Preserving the constraints $\mathbf{b}$, the linear program remains feasible as both x^I and x^E are solutions. Yet, the actual optimum may very well be better than both of them, and at least as good as the best, x^E.

The complete algorithm for a single thread is shown in Listing 5. Because each thread executes almost independently of other threads, synchronisation is only needed when starting and stopping the resampling. Only the operations of writing and reading the external solutions need to be synchronised to ensure the correctness of the algorithm. Concurrent beam resampling is thus highly parallelisable, as all other resources are not shared between the threads.

6.2.1 Integration of Beam Resampling with I3D Planning

All three planning tools that have been presented explore a fixed search space to identify the optimal beamset. The size of the search space is limited by the interactivity requirements of each of the tools. Obviously, WF and IDVH will have much laxer runtime constraints than I3D, as the number of planning steps required to arrive at a final solution is disproportionately higher for the latter. Consequently, smaller beamsets are used during I3D planning, which, in turn, negatively impacts the final solution quality. To overcome this, concurrent beam resampling is integrated within the interactive planning workflow.

During beam resampling, the three dimensional dose distribution changes with every new solution. Areas that are limiting for the current goal vary continuously. It is then obvious that there is little sense in performing spatial trade-off exploration based on the distance to bound heuristic, which identifies limiting voxels. Consequently, we are compelled to perform these two operations serially: first we explore the spatial trade-offs with a *suitably sized* and fixed beamset; second, we perform beam resampling as a final step.

Algorithm 5 The main procedure for a thread that randomly resamples a beamset. The exterior solution—beamset and objective value (B_+^{ext}, o^{ext})—is atomically read and written. If the exterior solution is better than the last solution, the exterior beams are inserted into the beamset. When a new solution is found, the result is posted (line 8), which calls the **resultPosted**() function for each of the peer threads. The other functions and variables are explained in Listing 1.

```
1:  function EXECUTE(N_t, N_r)
2:      B ← generateRandom(N_t)
3:      (B_+^ext, o^ext) ← (∅, ∞)                    ▷ Initialise exterior solution
4:      Δ_obj ← ∞                                    ▷ Initialise objective improvement
5:      while Δ_obj > T_obj do
6:          (B_+^int, B_0^int, o^int) ← solve(B)     ▷ Optimise with current beamset
7:          Δ_obj ← o^prev − o^int                   ▷ Compute objective improvement
8:          postResult(B_+^int, o^int)               ▷ Post result to other threads
9:          if o^ext − o^int < T_improv then         ▷ Accept external if good enough
10:             B_n ← generateRandom(N_r − |B_+^ext|)   ▷ Randomly generate new beams
11:             B_r ← selectRandom(B_0^int, N_r)        ▷ Randomly select beams to remove
12:             B = B_+^int ∪ B_n ∪ (B_0^int \ B_r)     ▷ Update current beamset
13:         else
14:             B_n ← generateRandom(N_r)
15:             B_r ← selectRandom(B_0^int, N_r)
16:             B = B_+^int ∪ B_n ∪ (B_0^int \ B_r)
17:         end if
18:         o^prev = o^int
19:     end while
20:     return (B_+^int, o^int)
21: end function

1:  function RESULTPOSTED(B_+, o)
2:      if o < o^ext then                            ▷ Accept only the best external solution
3:          (B_+^ext, o^ext) ← (B_+, o)
4:      end if
5:  end function
```

We say that a beamset is of suitable size, if it sufficiently describes the planning problem: (1) the objective value after I3D planning is close to the objective value after resampling and (2), relaxations made during I3D planning are advantageous to the optimisation goals during resampling. Note that any unnecessary relaxations will be recovered during beam resampling,

Similar to stepwise planning, optimisation goals are resampled in order of their priority. After each resampling step, the solution is fixed as discussed in Chapter 3 and Chapter 5. An attempt is also made to recover any relaxations performed during planning, by resampling with the dual goal of minimising the sum of differences to the original bounds. Assuming the relaxations were executed in increasing order of their importance (less critical relaxations first), beam resampling of recovering the slack will be applied in the inverse order. Such relaxation recovery objectives can be interleaved with the objectives for which they were originally performed, or they can be considered as the last steps of resampling. The process is explained in Figure 6.5.

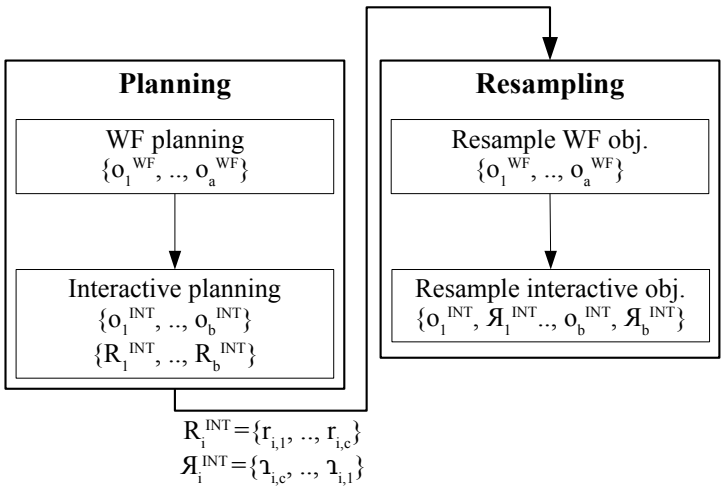


Figure 6.5: Integration of beam resampling in the interactive planning workflow, as a final step. Objectives are resampled in descending order of their priority. Each interactive planning objective is followed by a series of relaxation steps, R_i^{INT}, consisting of one or more trade-offs, $\{r_{i,1}, .., r_{i,c}\}$. Relaxations are recovered by resampling with the respective dual objectives in the inverse ordered the relaxations were performed $Я_i^{INT}$. Note that the resampling order is flexible: one could conceivably relax the WF objectives, once, before starting I3D planning and, again, after the resampling of the I3D objectives.

7 Experimental Results

This chapter presents the experimental evaluation of the methods presented in previous chapters and which have been implemented in the interactive planning tool. A special emphasis is given to the use of Interactive3D for spatial trade-off navigation. Section 7.1 demonstrates the combined use of the interactive planning tools—WF-based, IDVH, and I3D—for finding a Pareto optimal solution with respect to both localised and non-localised goals. Next, Sections 7.2 and 7.4 give examples of practical uses of localised dose shaping in a series of planning scenarios for three different cancer cases. In each scenario, educated assumptions are made about the benefit of local manipulation; the steps involved during planning are shown in detail. Sections 7.5 and 7.6 evaluate the anatomy and beam space sampling approaches from Chapter 6. Finally, in Section 7.7, the influence of the different optimisation parameters on the serial performance of the CPLEX optimiser is analysed.

In our experiments, we used plan data of real patients who underwent radiation treatment at CyberKnife centres in Güstrow (Germany), München (Germany), and Milan (Italy). All personal data have been anonymized.

7.1 Synergic Planning for Multicriteria Optimisation

We demonstrate the synergic use of the planning interfaces on a simplified optic meningioma case containing two VOIs: the PTV and the shell, located at a distance of $12\,\text{mm}$ of the PTV. In this scenario, we analyse the trade-off between target coverage at $18\,\text{Gy}$ and shell dosage to under $3.5\,\text{Gy}$ (target is to minimise the percentage of the shell receiving $3.5\,\text{Gy}$ or more). We compare the spatial dose optimisation method to two conventional approaches: unconstrained multicriteria optimisation by scalarization—UMO (see Section 2.3.2.1)—and constrained multicriteria optimisation by lexicographic ordering—CLO (see Section 2.3.2.2). The beamset is made up of $7.5\,\text{mm}$ and $10\,\text{mm}$ beams, generated as to avoid both eyes; the maximum beam bound is $180\,\text{MU}$.

7.1.1 Unconstrained and Constrained Optimisation

For the UMO, we setup the following optimisation problem, in which the total dose deviation from the prescription dose for all PTV voxels ($\mathbf{b_{l_{PTV}}} = \mathbf{18}$ Gy) and the total dose deviation from the reference dose for all shell voxels ($\mathbf{b_{u_{SHELL}}} = \mathbf{3.5}$ Gy) are simultaneously optimised:

$$
\begin{array}{lllll}
min & \mathbf{c}^T_{\hat{s}_{SHELL}}\hat{s}_{SHELL} + \mathbf{c}^T_{\check{s}_{PTV}}\check{s}_{PTV} & & & \\
s.t. & \mathbf{Ax} - \hat{\mathbf{s}}_{SHELL} & & & \leq \mathbf{3.5} \\
 & \mathbf{Ax} & & + \check{\mathbf{s}}_{PTV} & \geq \mathbf{18} \\
 & \mathbf{Ax} & & & \leq \mathbf{b}_{u_{PTV}} \\
 & \mathbf{x} & & & \leq \mathbf{b}_{w} \\
 & \mathbf{x}, \hat{\mathbf{s}}, \check{\mathbf{s}}, \hat{\mathbf{t}}, \check{\mathbf{t}} & & & \geq \mathbf{0}
\end{array}
$$

Note that this formulation is equivalent to simultaneously optimising an OCO and an OME step. To obtain the Pareto front, we fix $\mathbf{c^T}_{\check{s}_{PTV}}$ and uniformly sample $\mathbf{c^T}_{\hat{s}_{SHELL}}$ in a given interval (empirically deduced to allow for the balancing of both goals). The Pareto front of the UMO formulation is depicted as the blue curve in Figure 7.1a.

On the other hand, the CLO problem is solved in two steps, each optimising a simple objective: an OCO step, establishing PTV coverage, and an OME step, minimising shell dosage. To obtain the Pareto front for the CLO formulation, we increase the shell upper bound in 0.5 Gy increments, starting from an original value of 3.5 Gy. The optimisation results are plotted in the objective space, resulting in the red Pareto curve from Figure 7.1a.

7.1.2 Interactive Planning

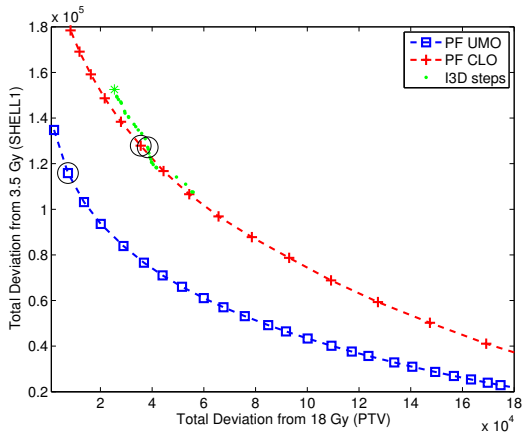
We start interactive planning with the IDVH tool, by setting the shell upper bound set 3.5 Gy and optimising for PTV coverage. This results in an initial PTV coverage of 60.8 %. Next, we successively relax the shell upper bounds to 3.75 Gy, 4.25 Gy, and 4.75 Gy, thus pushing PTV coverage to 70.4 Gy, 88.5 Gy, and 98.7 Gy, respectively. The individual steps plotted in the objective space are shown in Figure 7.2a and key statistics are given in Table 7.1. We deem step (3), shell relaxation to 4.75 Gy, to be unacceptable and we return to a solution close to that of step (2), as explained in Figure 7.2a. Note that to re-establish (2), one could simply have reset the shell upper bound to 4.25 Gy, followed by an OCO step. This example merely shows the exploration of the solution space through successive optimisation and relaxation steps.

Next, we continue planning with the Interactive3D tool, for studying the spatial trade-offs. First, we import the solution from step (6) into I3D. We consider it acceptable to allow a flatter dose gradient through the nasal and the ear cavities and, as such, we relax the shell voxels in these areas to 5.75 Gy and 4.75 Gy respectively. This allows the PTV coverage to increase to 93 % from 88.2 %. Figure 7.3 shows the evolution of the DVH curves and of the 4.25 Gy isodose surface during interactive planning. Dose fingers appear only in the relaxation areas of the nose cavity; this dose increase is also noticeable, yet not localised, on the DVH curves.

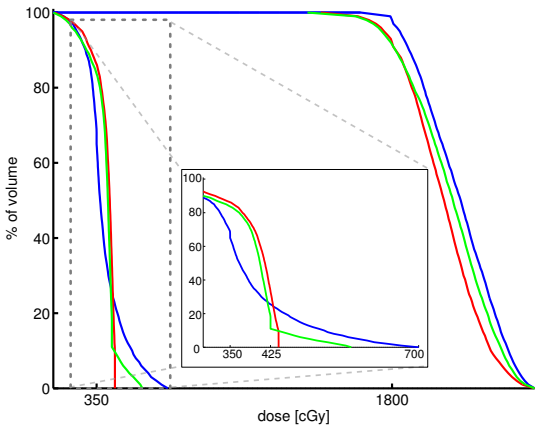
From looking at Figure 7.1a, it is obvious that the UCO Pareto front dominates the CLO front, which in turn dominates the solution obtained through combined use of IDVH and I3D. Despite this, UCO solutions are not necessarily better than the others. For approximately the same shell objective value, the UCO solution delivers a much higher maximum dose to the shell (compare DVH curves in Figure 7.1b). While CLO can limit the amount of overdosage, it cannot control the areas where it occurs: consequently, the 4.25 Gy isodose surface is able to expand beyond the shell in all directions. By comparison, I3D planning allows the user to both control the trade-off amount as well as its location (compare Figures 7.1 c–d). This is a practical example of the interactive planning workflow we explained Chapter 3, Figure 3.6: first we use the IDVH tool to quickly navigate the CLO Pareto front in search of an acceptable solution; second, we tweak this solution with the I3D tool, by exploring the possible spatial trade-offs.

Table 7.1: Combined IDVH and I3D interaction steps for planning the synthetic meningioma case. Plan metrics are PTV coverage (CO), minimum (min) and maximum (max) doses, objective value (total dose deviation—TD), and $V_{4.25}$ for the shell.

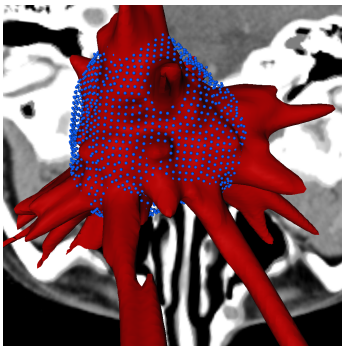
step	PTV			shell		
	CO (%)	min (Gy)	TD	$V_{4.25}$ (%)	max (Gy)	TD
0. initial	60.8	9.64	421 156	0.0	3.50	0
1. IDVH r. shell	70.4	10.56	247 033	0.0	3.75	18 830
2. IDVH r. shell	88.5	13.65	54 409	82.8	4.25	106 811
3. IDVH r. shell	98.7	16.63	1921	89.7	4.75	203 686
4. IDVH OME shell	98.7	16.63	1921	89.7	4.75	203 686
5. IDVH r. PTV	66.8	12.28	1921	62.3	4.25	64 231
6. IDVH OCO PTV	88.2	13.57	248 992	82.7	4.25	106 628
7. IDVH OME shell	88.2	13.57	54 599	82.7	4.25	106 628
8. I3D r. shell	93.0	13.98	25 419	19.2	5.75	152 722



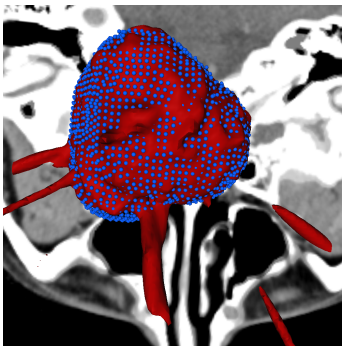
(a) Pareto fronts



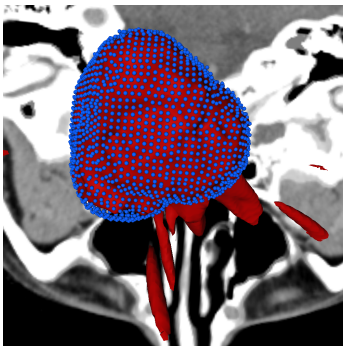
(b) DVH curves for PTV & shell



(c) UMO

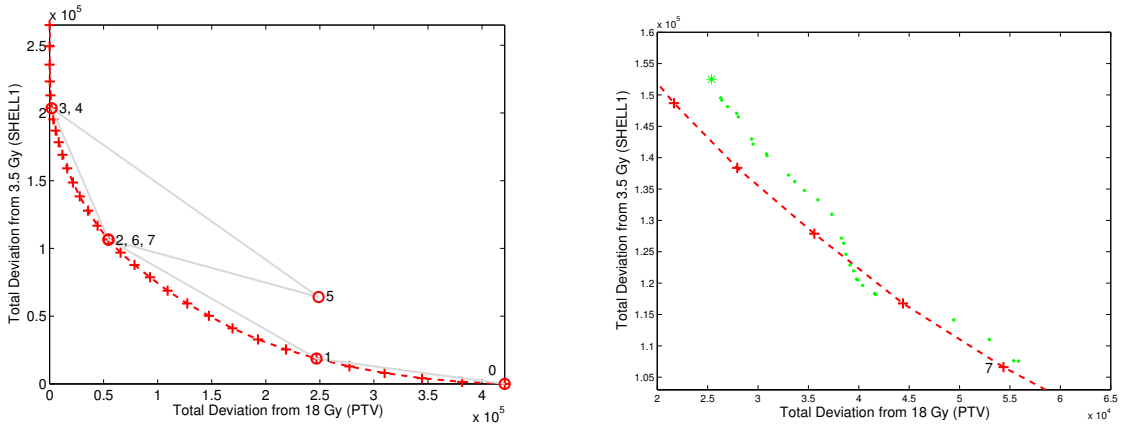


(d) CLO



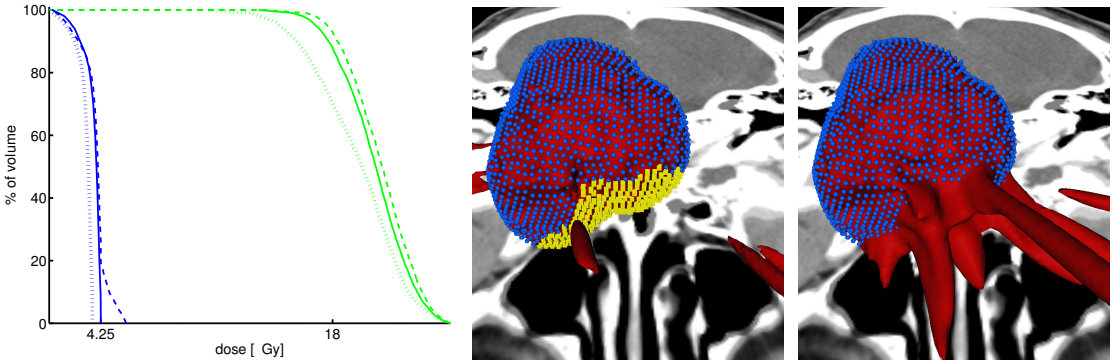
(e) Interactive

Figure 7.1: (a) The Pareto fronts obtained for UMO (blue) and CLO (red), and the individual I3D steps (green) projected in the objective space. (b) The PTV and shell DVH curves corresponding to the encircled points in (a). The inset shows the rescaled shell DVH curves from the gray dotted box: notice how with I3D bounds of $4.25\,\mathrm{Gy}$ are only violated for the specifically relaxed voxels; the comparable CLO solution uses a hard VOI bound larger than $4.25\,\mathrm{Gy}$. (c, d, and e) The $4.25\,\mathrm{Gy}$ isodose surface (red) and shell voxels (blue spheres) for the encircled UMO (c), CLO (d), and I3D (e) solutions respectively. The shell violation is clearly visible and, only in the case of I3D, is it contained to a specific area.



(a) (b)

Figure 7.2: (a) Navigation of the CLO Pareto front with the IDVH tool (circles show the individual IDVH solutions). Starting with a bound of $3.5\,\mathrm{Gy}$ (0), the shell is further relaxed to $3.75\,\mathrm{Gy}$ (1), $4.25\,\mathrm{Gy}$ (2), and $4.75\,\mathrm{Gy}$ (3) respectively. After each relaxation, the OCO optimisation is solved. Next, the solution is fixed and an OME step to reduce shell dose to under $4.25\,\mathrm{Gy}$ follows (4). To return to (2), the PTV lower bound is relaxed (5) and a new OCO step reestablishes PTV coverage (6). A final OME step to $3.5\,\mathrm{Gy}$ ensures the Pareto-efficiency of point (7). (b) The individual I3D planning steps (green) shown in the objective space in relation to the CLO Pareto front (red).



(a) DVHs for initial, IDVH and I3D (b) Before I3D (c) After I3D

Figure 7.3: (a) The DVH curves for the PTV (green) and shell (blue), corresponding to the initial solution (1) (dotted), after IDVH (6) (solid), and after I3D (8) (dashed). (b, c) The $4.25\,\mathrm{Gy}$ isodose surface before (b) and after (c) I3D relaxations. The relaxed voxels are depicted as yellow boxes in (b).

7.2 Prostate Cancer

Stereotactic body radiation therapy (SBRT) is one of the treatment options for patients suffering from prostate cancer due to its ability to deliver escalated dose in a low number of fractions [126, 127]. The rectum, the bladder, and, to a certain degree, the urethra are the main organs at risk as they are in close proximity to the prostate [8, 128–130].

The case we consider is that of an enlarged prostate of $124\,\mathrm{cm}^3$, for which a dose of $36.25\,\mathrm{Gy}$ was prescribed at the $87\,\%$ isodose (prostate maximum dose of $41.50\,\mathrm{Gy}$). Conformality is ensured by the use of two shell structures at $5\,\mathrm{mm}$ and $12\,\mathrm{mm}$ from the target. Beams with sizes ranging from $10\,\mathrm{mm}$ to $40\,\mathrm{mm}$ are used during treatment. The original bounds and planning settings, listed in Table 7.2, are compatible with published clinical practice [131].

Table 7.2: Initial planning settings for the prostate case.

Parameter	Value	Comment
PTV UB/TX	$41.50/36.25\,\mathrm{Gy}$	upper bound / TX dose ($87\,\%$ isodose)
rectum UB	$33.00\,\mathrm{Gy}$	—
bladder UB	$33.00\,\mathrm{Gy}$	—
urethra UB	$38.00\,\mathrm{Gy}$	—
SHELL1 UB	$35.00\,\mathrm{Gy}$	inner shell, $5\,\mathrm{mm}$ from PTV
SHELL2 UB	$20.00\,\mathrm{Gy}$	outer shell, $12\,\mathrm{mm}$ from PTV
beam UB	$500\,\mathrm{MU}$	1500 beams used for I3D
max. MU	$40\,000\,\mathrm{MU}$	total maximum beam weight
collimators	10, 15, 20, 30, and $40\,\mathrm{mm}$	diameters
beam node UB	$2000\,\mathrm{MU}$	max. tot. beam weight out of one node
optimisation grid	$1\,\mathrm{mm}$ and $3\,\mathrm{mm}$	urethra and all other VOIs, respectively
interaction grid	$3\,\mathrm{mm}$	grid VOI

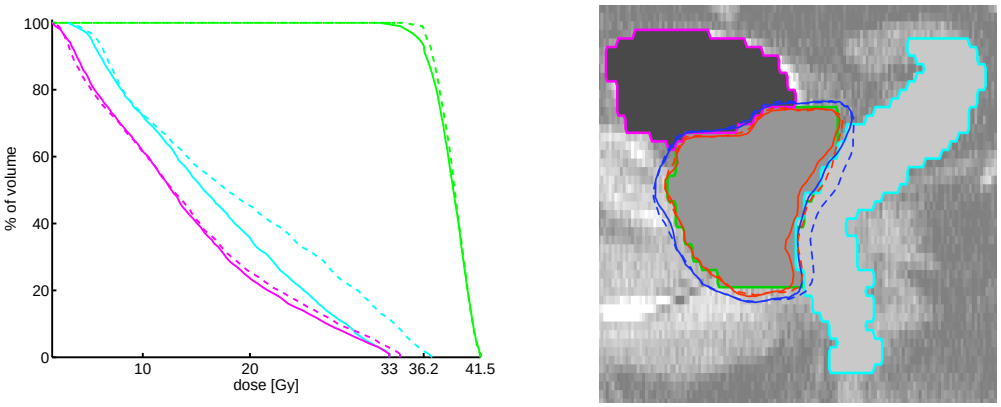
7.2.1 IDVH Preplanning

An initial OCO step establishes a PTV coverage of $92.9\,\%$ at $36.25\,\mathrm{Gy}$. Next, we successively relax the rectum upper bound to $35\,\mathrm{Gy}$ and $37\,\mathrm{Gy}$, thus increasing PTV coverage to $95.8\,\%$ and $97.5\,\%$, respectively. The last trade-off we attempt is to relax the bladder to $34\,\mathrm{Gy}$, resulting in a target coverage of $98.7\,\%$. Finally, we establish Pareto efficiency by two OME steps targeting $33\,\mathrm{Gy}$ for both rectum and bladder.

The IDVH planning steps are summarised in Table 7.3. Figure 7.4 shows the DVHs curves and the isodose curves for the initial and final solutions using IDVH. As coverage increases, the 36.25 Gy isodose curve expands to cover almost the entire prostate, while the 33 Gy curve cuts through the surface of the rectum, facing the PTV.

Table 7.3: Summary of the IDVH planning steps for the prostate case. Listed metrics are: PTV coverage (CO), minimum (min) and maximum (max) doses; V_{33} is the percentage of the respective VOI receiving a dose of at least 33 Gy.

	PTV			rectum		bladder	
step	CO (%)	min (Gy)	max (Gy)	max (Gy)	V_{33} (%)	max (Gy)	V_{33} (%)
init. OCO	92.9	32.11	41.50	33.00	0.0	33.00	0.0
r. rect. 35 Gy	95.8	33.35	41.50	35.00	6.3	33.00	0.0
r. rect. 37 Gy	97.5	33.42	41.50	37.00	11.2	33.00	0.0
r. blad. 34 Gy	98.7	34.33	41.50	37.00	11.4	34.00	3.6
est. Pareto.	98.7	34.33	41.50	37.00	11.4	34.00	2.7

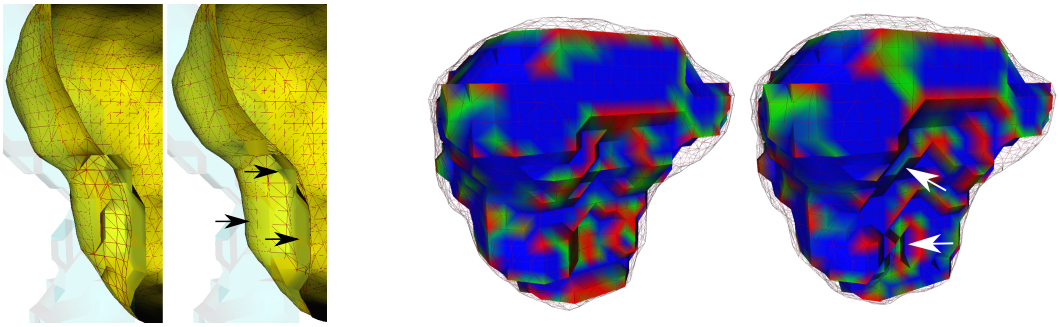


(a) (b)

Figure 7.4: (a) DVH and (b) isodose curves for the PTV (green), rectum (cyan) and bladder (magenta) showing the initial (solid) and final (dashed) IDVH solutions. (b) Sagittal slice showing the 33 Gy (blue) and the 36.25 Gy (red) isodose curves. As the rectum and bladder are relaxed, the 33 Gy isodose expands, increasing exposure to the OARs; as a result, PTV coverage is improved.

7.2.2 Scenario 1: Rectal Wall Sparing

The PTV coverage of $98.7\,\%$ comes with the cost of a maximum dose to the rectum of $37\,\mathrm{Gy}$. Rectal toxicity has been extensively studied and a strong correlation with dose exposure has been demonstrated [8, 128, 129, 132, 133]. Furthermore, Acosta et al. [134, 135] showed that most rectal side-effects are located within $1\,\mathrm{cm}$ of the anterior rectal wall. Motivated by this, we push the $33\,\mathrm{Gy}$ target surface to uncover the rectal wall facing the prostate (see Figure 7.5a). The resulting SME step, targeting a minimisation of the rectum V_{33}, only slightly improves the goal, as V_{33} only drops by $0.1\,\%$. Thus, we are compelled to consider possible trade-offs.



(a) Isodose surface evolution

(b) Lower bounding surface (PTV) relaxation

Figure 7.5: (a) Clipping plane view of the target (yellow) and of the $33\,\mathrm{Gy}$ isodose surfaces (red, wireframe), before (left) and after (right) I3D planning to spare the rectum (partially transparent). Arrows highlight area of change, as the isodose surface moulds to the target surface. (b) The $34.5\,\mathrm{Gy}$ lower bounding surface before (left) and after (right) prostate lower bound relaxation. Limiting areas are highlighted in red. Arrows point to areas where the bounding surface was sculpted to relax the prostate.

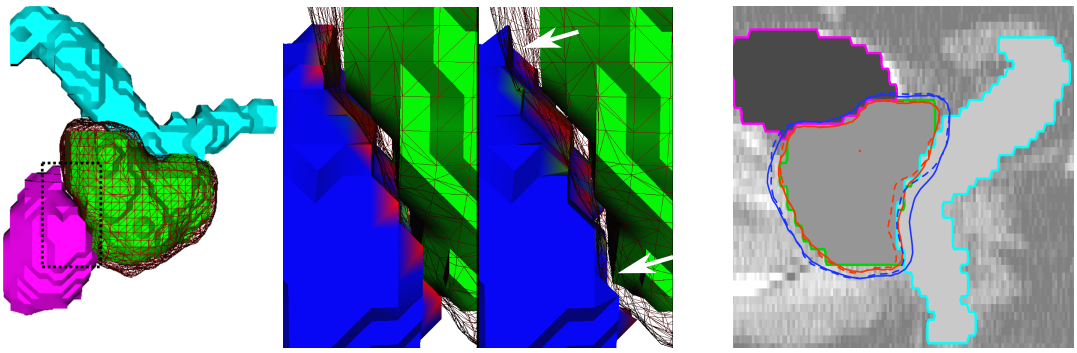
Due to its close proximity to the rectum, the prostate shows up as highly limiting while inspecting the respective lower bounding surface. Consequently, we relax parts of the PTV surface facing the target to $34.5\,\mathrm{Gy}$ by sculpting out the lower bounding surface, as shown in Figure 7.5b. Following optimisation, target area sparing improves, as target V_{33} (percentage of the target receiving at least $33\,\mathrm{Gy}$; initially target $V_{33} = 100\%$) drops to $21.6\,\%$ at a cost of $2.2\,\%$ drop in PTV coverage. This relaxation step is summarised in the first two entries of Table 7.4.

The next trade-off we consider is to allow slightly more dose to the bladder. We locally relax limiting margins of the bladder to $35\,\mathrm{Gy}$ by sculpting out the $33\,\mathrm{Gy}$ upper bounding surface as shown in Figure 7.6a. We limit our interactions to the edges of the OAR, where due to motion and bladder fill there is uncertainty about the actual extent

Table 7.4: Summary of the prostate planning steps to selectively spare the rectum. Metrics include PTV coverage (CO), minimum (min) and maximum (max) doses and V_{33}, as the percentage of the a VOI (target, rectum, and bladder) receiving a dose of at least $33\,\mathrm{Gy}$. Notice that during the last step, the PTV coverage drops further. This is due to recovering bladder relaxations, by using the available PTV slack from before.

	PTV		target			rectum	bladder	
step	CO (%)	min (Gy)	mean (Gy)	max (Gy)	V_{33} (%)	V_{33} (%)	max (Gy)	V_{34} (%)
OME tgs.	98.7	34.33	34.74	37.00	100.0	11.1	34.00	0.0
r. PTV	96.5	32.80	31.62	35.12	21.7	6.1	34.00	0.0
r. blad.	96.7	31.86	30.90	34.38	8.2	4.9	35.00	1.9
est. Pareto	95.9	32.67	31.32	34.60	8.2	4.9	35.00	1.2

of the organ. The objective improves once again, as the dose shifts towards the bladder, thus decreasing target V_{33} to $8.2\,\%$.



(a) Upper bounding surface (bladder) relaxation

(b) Isodose curves on sagittal slice

Figure 7.6: (a) Relaxation of the bladder upper bound. The overview on the left is shown to help orientate the reader (prostate—green, rectum—cyan, bladder—magenta, and isodose surface—red, wireframe) as it outlines the zoomed in areas to its right. The two images to the right show the $35\,\mathrm{Gy}$ upper bounding surface (corresponding to the bladder), coloured to highlight limiting areas in red, before (left) and after (right) bladder relaxation. Relaxation consists of sculpting out the upper bounding surface, as pointed by the arrows. (b) Sagittal slice showing the $33\,\mathrm{Gy}$ (blue) and the $36.25\,\mathrm{Gy}$ (red) isodose curves, before (solid) and after (dashed) I3D planning of Scenario 1. The $33\,\mathrm{Gy}$ isodose curve contracts to spare the target area of the rectum, while the there is a noticeable drop in PTV coverage.

Because we neither want to continue compromising PTV or bladder dosage, nor can we further improve the objective through other relaxations, we end I3D planning at this

step. Wrapping up, we recover any available slack and we establish Pareto optimality as already discussed (see summary in last entry of Table 7.4).

Comparing the initial and resulting dose distributions in Figures 7.5a and 7.6b, we see obvious changes due to I3D planning, as the 33 Gy isodose shrinks to uncover the target area of the rectum. The DVH curves in Figure 7.7 reveal that most dosimetric variation is limited to the target and relaxation subvolumes, with no other dose criteria worsening.

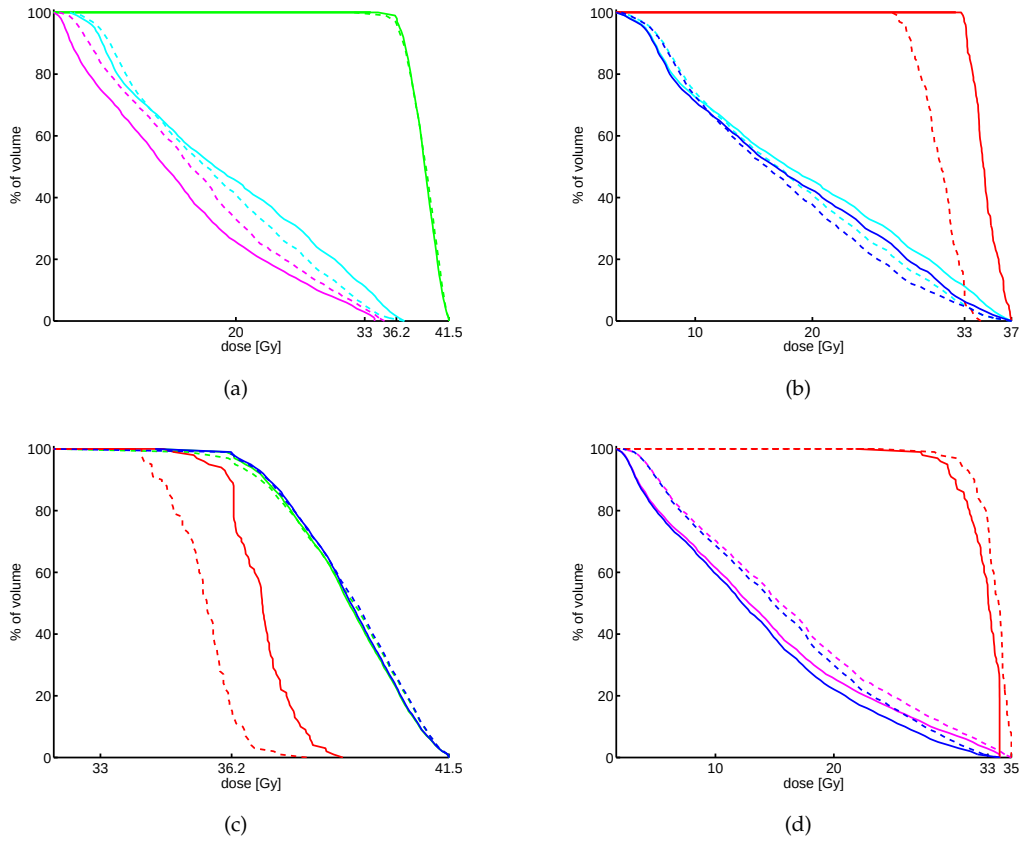


Figure 7.7: DVH curves summarising the evolution of the dose distribution for Scenario 1 of the prostate case, before (solid) and after I3D planning (dashed). (a) DVHs for the PTV (green), rectum (cyan) and bladder (magenta). Notice the rectum V_{33} decreasing at a cost to the PTV and bladder. (b) DVHs for the entire rectum (cyan) and for its target (red) and non-target (blue) subvolumes. Initially, the entire target receives a dose of at least $33\,\mathrm{Gy}$; after planning, target V_{33} drops close to $0\,\%$. The non-target subvolume sees almost no change. (c) DVHs for the entire PTV (green) and for its relaxed (red) and non-relaxed (blue) subvolumes. While dose to the relaxed subvolume drops abruptly, almost no change can be detected for the non-relaxed subvolume. (d) DVHs for the entire bladder (magenta) and for its relaxed (red) and non-relaxed (blue) subvolumes. As in (c), dose change is mostly limited to the relaxed subvolume.

7.2.3 Scenario 2: Boost Volume Coverage

Next we investigate the feasibility of escalating the dose to a subvolume of the PTV. This scenario is similar to *dose painting by contours*, where an increased uniform dose is prescribed to a subvolume of the original target in an effort to improve outcome predictors [47, 136]. SBRT has been recently shown to be able to produce such a dose distribution, which has actually been a characteristic of HDR brachytherapy [137, 138].

(a) Target and isodose surface evolution

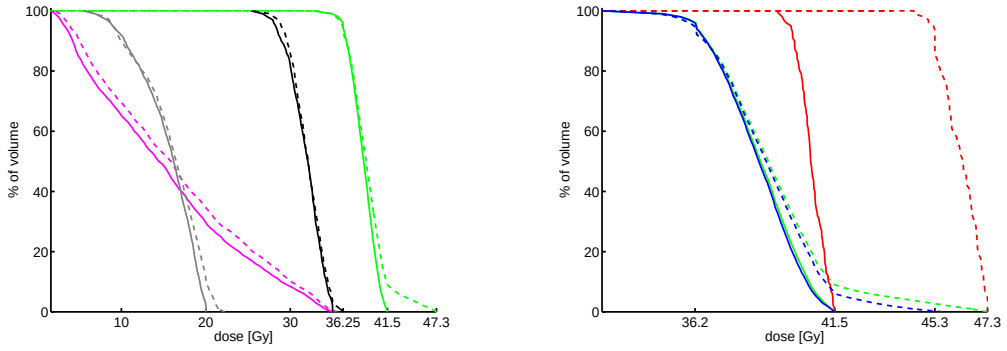
(b) Upper bounding surface (PTV) relaxation

(c) Isodose curves on sagittal slice

Figure 7.8: Rendering of the I3D Scenario 2 showing the PTV (green), the target boost volume (yellow), the rectum (cyan) and the bladder (magenta). (a) The target surface (yellow) and the evolution of the $45.30\,\mathrm{Gy}$ **isodose surface (red) during planning. The PTV is rendered as a partially transparent green wireframe surface. The isodose surface conforms to the target shape after the relaxation of the PTV (upper), shells (middle) and, bladder (lower) bounds. (b) The** $41.50\,\mathrm{Gy}$ **isodose surface (red) within the containing** $41.50\,\mathrm{Gy}$ **upper bounding surface (dark green) before (left) and after (right) PTV upper bound relaxation. The isodose surface expands within the hollowed out area of the bounding surface. (c) Sagittal slice showing the** $20\,\mathrm{Gy}$ **(black),** $33\,\mathrm{Gy}$ **(blue),** $41.5\,\mathrm{Gy}$ **(red), and** $45.3\,\mathrm{Gy}$ **(yellow) isodose curves, before (solid) and after (dashed) I3D planning. As a result of planning, the** $45.3\,\mathrm{Gy}$ **isodose appears, whereas it did not exist before.**

Starting from the Pareto efficient solution obtained in the previous scenario, we define the target as a boost volume inside one of the lobes of the PTV (Figure 7.8a) for which we target a dose escalation of $25\,\%$ (or, $45.3\,\mathrm{Gy}$), similarly to [47]. As expected, the subsequent SCO step fails to increase boost volume coverage, as the upper bounds of the adjacent PTV voxels enforce a very steep dose gradient. Consequently, we relax PTV voxels to between $43\,\mathrm{Gy}$ and $45\,\mathrm{Gy}$ around the boost volume, by sculpting out the upper bounding surface and taking care not to affect neither the surface of the PTV, nor the urethra (see Figure 7.8b). The ensuing optimisation succeeds to increase boost coverage

from 0 % to 56.2 %—compare first two entries in Table 7.5.



(a) (b)

Figure 7.9: DVH curves summarising the evolution of the dose distribution for the prostate Scenario 2, before (solid) and after (dashed) planning for the creation of the boost subvolume. (a) DVH curves for the PTV (green), bladder (magenta), inner (black) and outer (gray) shells. The slight increase in dosage to the shells and to the bladder (due to relaxations) are obvious. (b) DVH curves for the PTV (green), boost volume (red) and PTV non-target (blue) subvolumes. The PTV DVH shows only a small effect due to planning, yet the DVH of the target subvolume of the PTV, shows a large increase in dosage.

Additional inspection of the dose distribution shows both outer and inner shells as limiting. Correspondingly, we relax individual voxels belonging to the outer and inner shells to 23 Gy and 36 Gy, respectively. Again, target coverage increases as it reaches 63.5 %—third entry in Table 7.5.

For the third trade-off, we revisit the bladder, where we relax limiting areas on the bladder wall to 35.5 Gy. Boost volume coverage jumps to 70.8 %. Finally, we again relax the PTV and shell upper bounds, in areas that have now become limiting, and thus reaching an end coverage of 93.4 %. Tying everything up, we establish Pareto efficiency of the solution, which succeeds to increase overall PTV coverage and to reduce dosage to the OARs and to the shell (see last entry in Table 7.5).

The planning steps and results are summarised in Figure 7.9 and in Table 7.5: the creation of the boost volume is the result of careful and conservative relaxation of the PTV, bladder, and shell bounds. At the same time, the criterion pursued in the previous scenario does not worsen at all, as the rectum bounds remain untouched.

Table 7.5: Summary of the prostate planning steps to increase coverage to the boost volume inside the PTV. Metrics include boost volume coverage (CO), minimum (min) and maximum (max) doses per VOI and V_{34} (the percentage of a VOI receiving a dose greater than 34 Gy).

	boost		PTV			bladder		SHELL1	SHELL2
step	CO (%)	min (Gy)	CO	HI	CI	V_{34} (%)	max (Gy)	max(Gy)	max (Gy)
SCO boost	0.0	43.60	95.8	1.33	1.05	1.1	35.00	35.00	20.00
r. PTV	56.2	47.30	94.9	1.44	1.05	1.0	35.00	35.00	20.00
r. shells	63.5	47.30	95.1	1.44	1.07	1.1	35.00	36.00	22.84
r. blad.	70.8	47.30	94.2	1.44	1.06	2.4	35.50	36.00	21.86
r. misc	93.4	47.30	94.5	1.43	1.07	2.2	35.50	36.39	22.22
est. Pareto	93.4	47.30	95.7	1.43	1.06	2.0	35.50	36.00	22.12

7.2.4 Conclusions

Typical prostate cases present difficulties due to the proximity of the healthy organs to the tumour and side-effects are not uncommon. As such, the planning goals of special interest are those of improving organ sparing or of escalating dose to disease areas. In the scenarios we investigated, we were able to remodel the dose distribution, first to achieve better sparing in a critical area of the rectum and, second, to create a high dose subvolume within one of the prostate lobes. In reaching our goals we were compelled to accept some compromises, the extent of which we could precisely control down to voxel level. Relaxing the shells proved to be useless during the first scenario, and we promptly reverted the bounds to the original values. In Scenario 2, the same relaxations led to an important increase in boost volume coverage, from $56.2\,\%$ to $63.5\,\%$. This emphasises the fact that trade-offs are meaningful only relative to the current objective. Clearly, a systematic approach to interactive planning is critical to the outcome.

7.3 Optic Nerve Meningioma

Next case we consider is that of an optic nerve sheath meningioma (ONSM). Because the tumour is located very close to the optic nerve (see Figure 7.10), surgery is rarely an option [139] and radiotherapy, delivered as fractionated stereotactic radiotherapy or stereotactic radiosurgery, is preferred [140–142]. Part of the results in this section have been published in [143].

The original clinical goal was to cover the PTV with a therapeutic dose of $18\,\mathrm{Gy}$ while

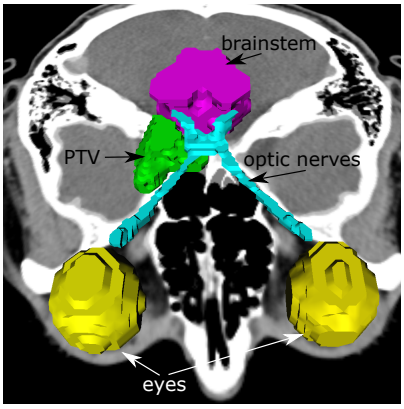


Figure 7.10: The VOIs involved in the ONSM case: PTV - green, optic nerves and chiasm - cyan, brainstem - magenta. and both eyes - yellow.

sparing the optic system (nerves and optic chiasm) and the brainstem as much as possible. We set the upper bounds for both OARs to 7 Gy, well within the dose limits recommended by the QUANTEC study [144] of 8 Gy and 12.5 Gy for the optic nerves and brainstem, respectively. Though the cochlea was not delineated, we take note of the QUANTEC recommendation to limit its exposure to a maximum of 12–14 Gy for SRS. Two shell structures, at a distance of 5 mm (SHELL1) and 12 mm (SHELL2) from the PTV, ensure dose conformity and limit the exposure of healthy brain tissue. Beams are generated to avoid both eyes and are constrained to a maximum of 180 MU. The complete list of planning parameters is given in Table 7.6.

The two scenarios we consider start from very different initial conditions, but arrive at similar solutions. In Scenario 1, initially, the optic nerve is well spared, but the PTV has a low coverage; the aim is to maximise PTV coverage. Conversely, in Scenario 2 we aim to improve the sparing of the optic nerve, while, initially, the PTV has an excellent coverage.

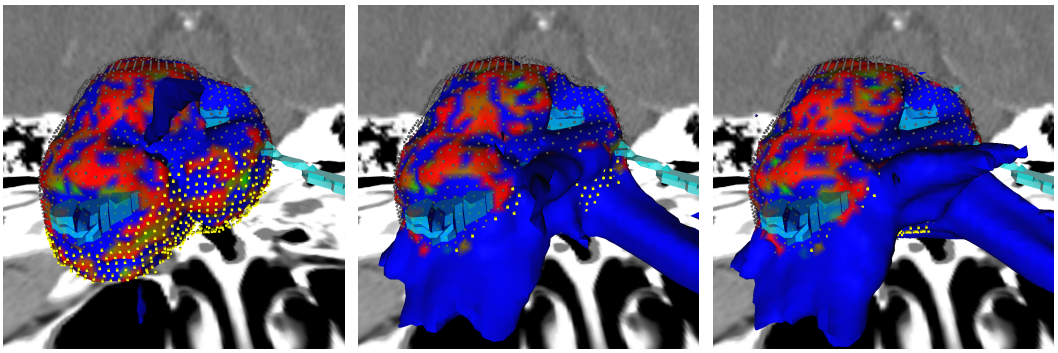
7.3.1 Scenario 1: PTV Coverage

First, starting with the bounds from Table 7.6, we run an OCO optimisation thus establishing an initial PTV coverage of 88.7 % at 18 Gy. The poor coverage is due to the very conservative OAR and shell bounds, which we need to relax if we are to improve the objective. Analysing the DVH curves in Figure 7.14, we indeed see that a large number of shell voxels are limiting. Consequently, the first trade-off we perform is to allow a flatter dose gradient through the nasal and ear cavities. We gradually relax the SHELL2 voxels to 6 Gy in the nasal cavity and to 5.5 Gy in the ear cavity (cochlea remains well within

Table 7.6: Initial planning settings for the optic nerve sheath meningioma case.

Parameter	Value	Comment
PTV UB/TX	25.70/18.00 Gy	72 % isodose
optic nerve UB	7.00 Gy	abbreviation: onerv
brainstem UB	7.00 Gy	abbreviation: bstem
eyes UB	1.50 Gy	—
SHELL1 UB	10.00 Gy	inner shell, 5 mm from PTV
SHELL2 UB	4.00 Gy	outer shell, 12 mm from PTV
beam UB	180 MU	6000 beams initial WF / 1600 beams I3D
collimators	7.5 and 10mm	diameters
optimisation grid	$1.05 \times 1.05 \times 1.05$ mm^3	PTV and optic nerve
	$1.58 \times 1.58 \times 1.58$ mm^3	brainstem, eyes
interaction grid	$1.58 \times 1.58 \times 1.58$ mm^3	grid VOI

recommended limits) and the SHELL1 voxels to 12 Gy in the nasal cavity. As a result, PTV coverage increases by an additional 3.7 % to 92.9 %.



(a) SHELL2 before relaxation (b) SHELL2 during relaxation (c) SHELL2 after all relaxations

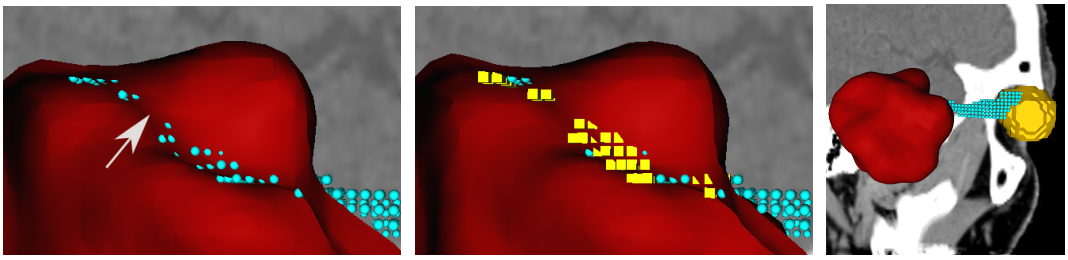
Figure 7.11: Voxels of the outer shell (gray spheres and yellow cubes) and the 4 Gy isodose surface (coloured to highlight limiting areas in red): (a) before any relaxations, (b) after outer shell and (c) after inner shell relaxations. The relaxed voxels are depicted as yellow boxes. As a result of repeated relaxation, dose fingers appear in the nasal cavity.

Figure 7.11 shows the relaxation of the outer shell and the resulting dose fingers expanding through the nasal cavity. Note how the colour of the isodose surface changes to blue, to indicate areas that are no longer limiting. Other parts of the isodose surface

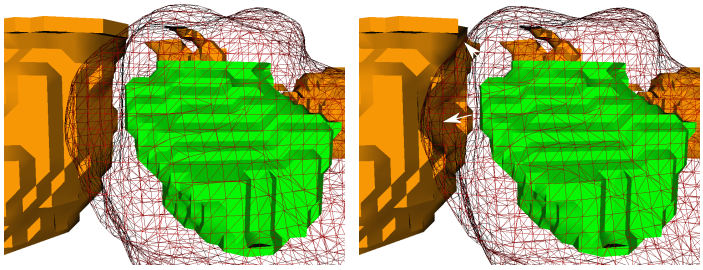
(towards the brain) remain red: the surface may not expand beyond them.

Next, we take a look at the optic nerves. Patches on the 8 Gy upper bounding surface, belonging to the nerves, are shown as limiting, some of which we carefully carve out to relax the bounds. PTV coverage increases again by 0.3 %, to 93.2 %, as dose is allowed to expand in the region of relaxation (see Figures 7.12 a, b).

Similarly, regions on the surface of the brainstem are limiting, and we relax them to 9 Gy, by carving out the respective upper bounding surface. This results in a final PTV coverage of 94.4 %, while the 7 Gy isodose surface expands to occupy the newly carved out area (see Figures 7.12 d and e, as well as the evolution of the isodose curves in Figure 7.13).

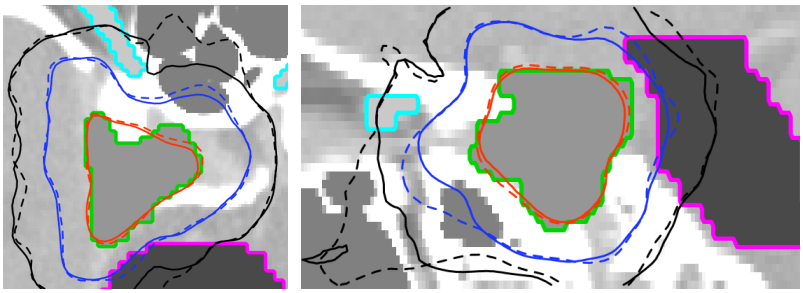


(a) Before optic nerve relaxation (b) After optic nerve relaxation (c) Context view



(d) Before brainstem relaxation (e) After brainstem relaxation

Figure 7.12: (a, b) View through the 7 Gy isodose surface (clipping plane is parallel to image plane) showing the dose extent relative to the optic nerve (cyan), (a) before and (b) after relaxation. Figure (c) shows a zoomed-out contextual view of the scene. Before relaxation, the surface barely touches the optic nerve, but as it expands during planning (in the direction indicated by the arrow), it engulfs those voxels that have been relaxed (rendered as yellow boxes). (c, d) The evolution of the 7 Gy isodose surface (red, wireframe), (c) before and (d) after relaxing the brainstem by carving the upper bounding surface (yellow) in the areas pointed to by the arrows. The isodose surface expands into the holes of the bounding surface created during relaxation. Also shown is the PTV, in green.



(a) Axial (b) Sagittal

Figure 7.13: Axial (a) and sagittal (b) slices showing the $4\,\mathrm{Gy}$, $7\,\mathrm{Gy}$, and $18\,\mathrm{Gy}$ isodose curves (black, blue, and red, respectively), before (solid) and after (dashed) I3D planning. The PTV (green), optic nerve (cyan), and brainstem (magenta) are also delineated. Notice the extension of the isodose curve, following planning, into the nasal cavity (located: top in (a); left in (b)), brainstem, and inside the PTV.

The changes in the dose shape are all the more evident on the CT slices in Figure 7.13: the prescription isodose curve extends mainly in the directions of the nasal cavity and of the brainstem. Table 7.7 details the key statistics for each of the planning steps in this scenario. Notice how the dose conformality worsens (increasing CI) as the shell bounds are relaxed and the prescription dose is allowed to extend further from the PTV. This trend continues with the relaxation of the two VOIs, optic nerves and brainstem (as slack, from the relaxation of the shells, can only now be used, when these new bounds have also been relaxed).

Table 7.7: Summary of the meningioma planning steps to improve coverage of the PTV. Metrics include coverage (CO) and conformality indices (CI), minimum (min) and maximum (max) doses.

step	PTV			optic nerve		brainstem		SHELL2
	CO (%)	CI	min (Gy)	V_7 (%)	max (Gy)	V_7 (%)	max (Gy)	V_4 (%)
OCO PTV	88.7	1.08	7.80	0.0	7.00	0.0	7.00	0.0
r. SHELL2	92.2	1.10	7.88	0.0	7.00	0.0	7.00	20.1
r. SHELL1	92.9	1.11	7.75	0.0	7.00	0.0	7.00	21.7
r. onerv	93.2	1.12	8.23	2.8	8.00	0.0	7.00	30.2
r. bstem	94.4	1.13	8.35	2.7	8.00	0.8	8.74	28.7

The DVH curves in Figure 7.14 compare the initial and final dose distributions. We

can thus see that dosimetric change, beyond the initial bounds, happens only for those voxels that were specifically selected for relaxation (target and relaxed subvolumes). On the other hand, the DVH curves for the non-target and non-relaxed subvolumes remain practically unchanged.

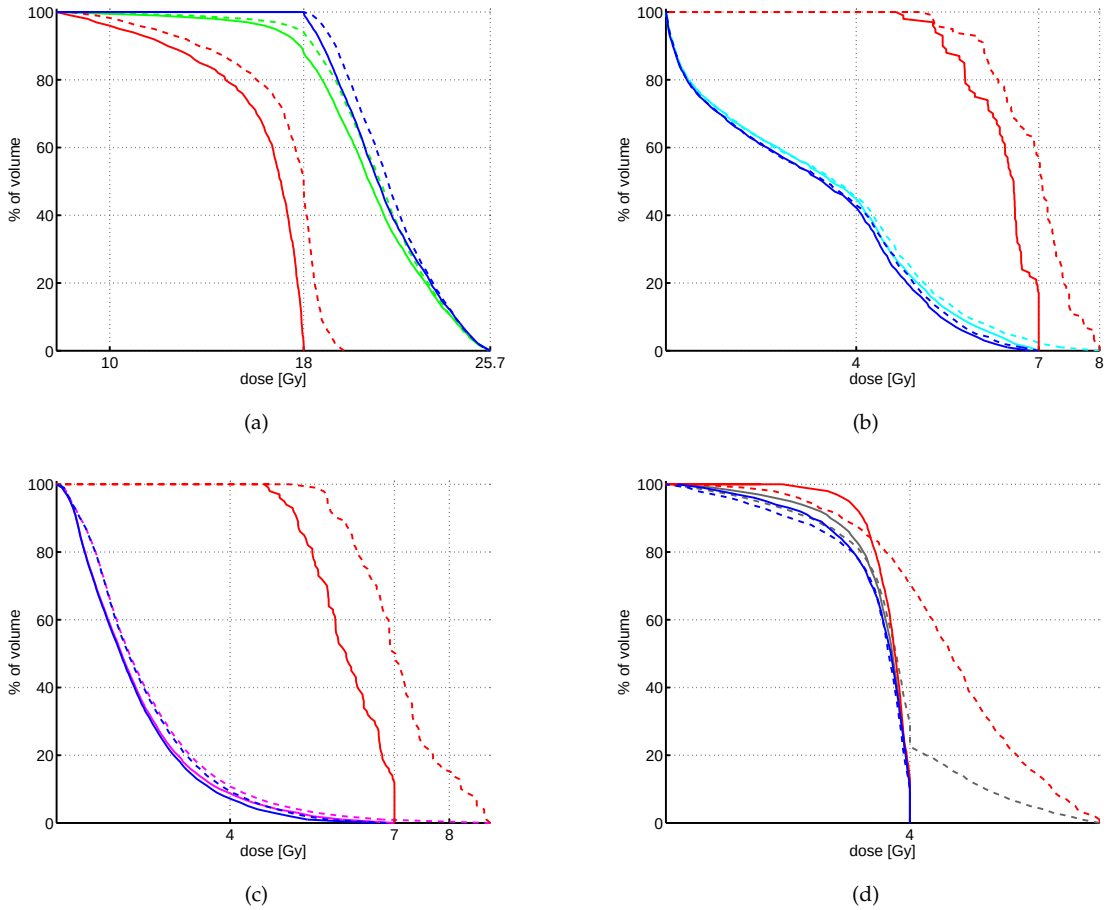


Figure 7.14: DVH curves before (solid) and after (dashed) performing local trade-offs for the ONSM case to increase PTV coverage. (a) DVHs for the complete PTV (green), and for its target (red) and non-target (blue) subvolumes. (b) DVHs for the complete optic nerve (cyan), and for its relaxed (red) and non-relaxed (blue) subvolumes. (c) DVHs for the complete brainstem (magenta), and for its relaxed (red) and non-relaxed (blue) subvolumes. (d) DVHs for the complete outer shell (magenta), and for its relaxed (red) and non-relaxed (blue) subsets of voxels. Notice how most dosimetric change is confined to the target and to the relaxed subvolumes; the DVH curves for the non-target and non-relaxed subvolumes remain almost unchanged.

7.3.2 Scenario 2: Optic Nerve Sparing

Whereas Scenario 1 started with an initial solution, in which the optic nerve was very well spared (and coming at a cost to PTV coverage), Scenario 2 starts with a good PTV coverage, coming with the cost of higher dose to the optic nerve. The goal of Scenario 2 is to improve optic nerve sparing. As we shall see, the final solution we arrive at shall be very similar to one obtained in Scenario 1.

First, we establish an initial solution by running an OCO step with relaxed optic nerve bounds (10 Gy) and slightly relaxed SHELL2 bounds (4.5 Gy). This results in excellent PTV coverage ($CO = 97.1\%$), coming at a cost of increased optic nerve exposure: more than 12 % of the optic nerve receives a dose of at least 7 Gy ($V_7^{onerv} = 12.8\%$). Dosage statistics for the other VOIs are given in the first entry of Table 7.8.

Table 7.8: Summary of the meningioma planning steps to spare the optic nerves. Last step, of establishing Pareto efficiency, is not shown, as it did not change the value of the statistics.

step	PTV		optic nerve		brainstem	SHELL2	
	CO (%)	min (Gy)	V_7 (%)	max (Gy)	max (Gy)	$V_{4.5}$ (%)	max (Gy)
OCO PTV	97.1	11.23	12.8	10.00	7.00	0.0	4.50
OME onerv	97.1	11.23	12.8	10.00	7.00	0.0	4.50
r. SHELL2	97.3	11.24	7.0	10.00	7.00	19.2	6.00
r. SHELL1	97.2	11.08	5.9	9.99	7.00	19.6	6.00
r. PTV	96.0	9.63	2.0	8.84	7.00	18.8	6.00
r. bstem	95.9	9.56	1.7	8.73	7.92	19.2	6.00

To improve optic nerve sparing, we setup an OME step aiming to reduce dose to under 7 Gy. The first optimisation step does not lead to any improvement (see second entry in Table 7.8) and we are thus obliged to investigate potential trade-offs.

Motivated by the large number of highly limiting voxels, we first relax the outer and inner shells (in the nasal and ear cavities, similarly to previous scenario). This leads to a reduction of V_7^{onerv} to just under 6 % (see third and fourth entries in Table 7.8). Next, we slightly relax areas on the PTV surface facing the target and showing up as limiting, by sculpting out the 9.5 Gy lower bounding surface; accordingly, V_7^{onerv} reaches 2 %. Lastly, we relax the brainstem by allowing 10 Gy though the few limiting areas on its surface. Comparing the last two entries in Table 7.8, we see that our last effort is futile, as the objective advances only ever so slightly. The resulting DVH curves in Figure 7.15 summarise the initial and final solutions. As trade-offs are precisely limited to parts of the

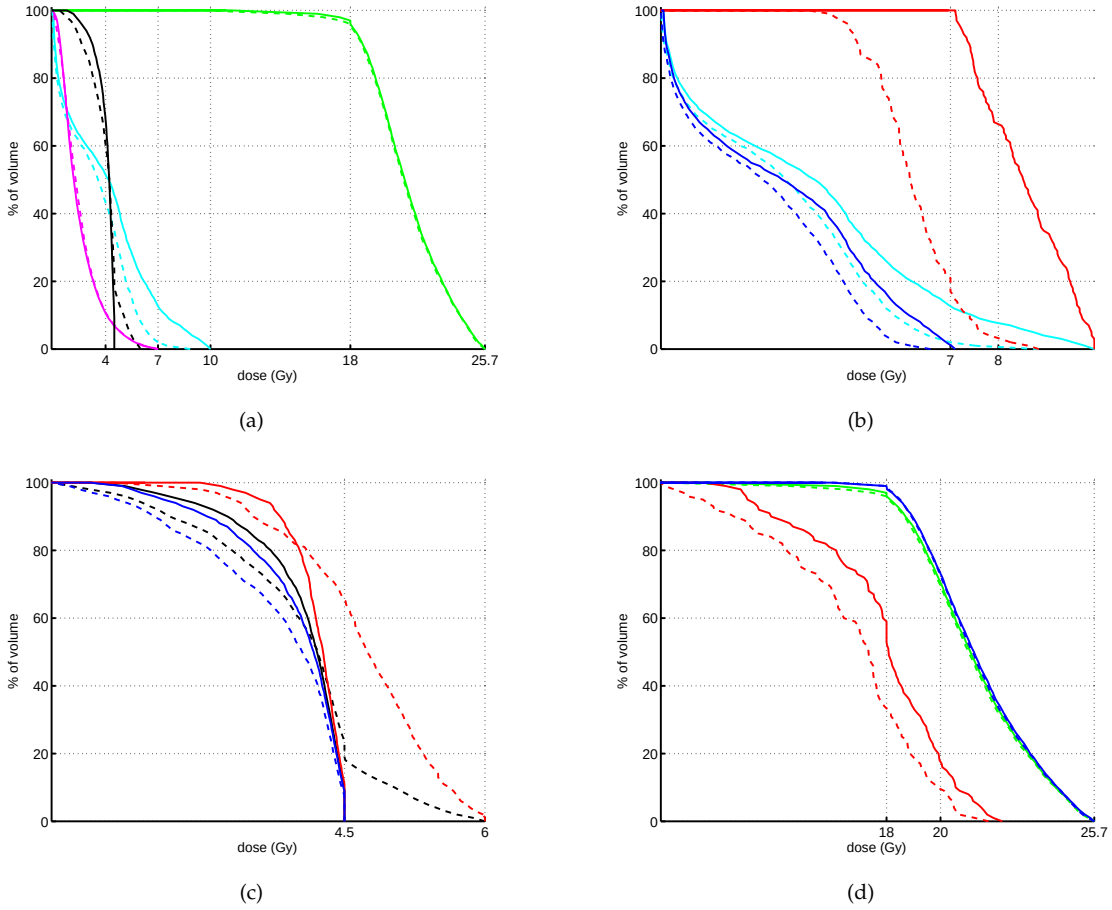


Figure 7.15: DVH curves before (solid) and after (dashed) performing local trade-offs for the ONSM case to spare the optic nerves. (a) DVHs for the PTV - green, optic nerve - cyan, brainstem - magenta, outer shell - black. (b) DVH curves for the entire optic nerve and its target (red) and non-target (blue) subvolumes. (c,d) DVH curves for the entire VOIs, SHELL2 and PTV respectively, and their relaxed (red) and non-relaxed (blue) subvolumes. Notice how dose change is limited to the relaxed subvolume; dose to the other subvolumes remains practically constant.

VOIs, the DVHs change only for the relaxation and target subvois; the DVHs of the other subvois remain practically unchanged.

7.3.3 Conclusions

Despite starting from disparate initial states, both scenarios result in similar solutions. One of the most important trade-offs for both cases has been that of allowing more dose

through the shells, in the directions of the nasal and ear cavities. Looking at the DVH curves for SHELL2 (Figures 7.14d and 7.15c), we see that most of the relaxed voxels surpass their initial bounds, as the optimiser makes use of the extra slack. This gives a sense of the influence of the outer shell on both optimisation goals. The other trade-offs, compromising either OAR sparing or PTV coverage, were evidently also necessary in achieving the optimisation goals.

Note that the apparent advantage of the solution from Scenario 2 over the one from Scenario 1 (better PTV coverage—$95.9\,\%$ vs. $94.4\,\%$—and better V_7^{onerv}—$1.7\,\%$ vs. $2.7\,\%$) can be mainly attributed on using laxer SHELL2 upper bounds ($4.5\,$Gy vs. $4\,$Gy). Indeed, comparing V_4^{SHELL2}, we see Scenario 2 allowing a lot more dose through: $60.8\,\%$ vs. $28.7\,\%$.

7.4 Spine Tumour

Stereotactic radiotherapy is often used in patients suffering from spine tumours, as either curative or palliative measure [145–148]. These cases are difficult to treat, mainly due to the closeness of the tumour to highly sensitive organs, such as the spinal cord and the esophagus.

Such a case is the one we consider next, that of a recurrent (relapse) spine tumour. Part of these results have been published in [149]. Figure 7.16a shows the PTV practically touching the spinal cord and in close proximity to the esophagus. The original plan imposed very strict dose bounds on the spinal cord, as the patient had already undergone multiple radiation regimens. These bounds hindered the dose coverage to the portion of the GTV proximate to the spine. Followup imaging revealed a relapse located inside the original GTV. This outcome motivates our main goal: to increase dose coverage of the GTV, especially in the area containing the relapse volume. In planning this case, we shall use both IDVH and I3D, and we shall compare the final results.

Obviously, knowing the location of a relapse before it occurs is a privilege reserved for fiction books and theoretical studies. We, unlike the original planners, possess this advantage. Table 7.9 lists the planning settings used in both the original plan and in our interactive approach.

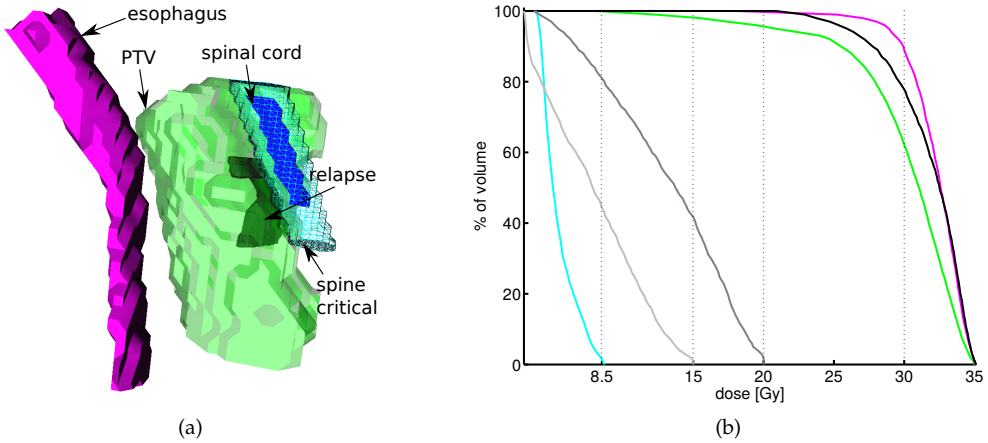
7.4.1 Initial Solution by WF Planning

First, we run similar optimisation steps to the ones used to generate the original treatment plan. These are listed in Table 7.10 and include OCO optimisations, to establish

Table 7.9: Summary of the initial planning settings for the relapse spine tumour.

	Parameter	Value	Comment
VOIs	PTV/GTV	$123.8\,\text{cm}^3/54.4\,\text{cm}^3$	—
	esophagus	$46.3\,\text{cm}^3$	abbreviation: esoph.
	spine critical	$10.1\,\text{cm}^3$	pre-irradiated, abbreviation: spncrt.
VOI bounds	GTV UB/TX	$35.00/30.00\,\text{Gy}$	upper bound / TX dose (85 % isodose)
	PTV UB/TX	$35.00/25.00\,\text{Gy}$	70 % isodose
	esophagus UB	$25.00\,\text{Gy}$	—
	spine UB	$25.00\,\text{Gy}$	—
	spine critical UB	$8.50\,\text{Gy}$	—
	SHELL1 UB	$20.00\,\text{Gy}$	inner shell, 15 mm from PTV
	SHELL2 UB	$15.00\,\text{Gy}$	outer shell, 30 mm from PTV
Beams	beam UB	$600\,\text{MU}$	3000 beams initial / 1600 beams I3D
	max. total MU	$65\,000\,\text{MU}$	—
	collimators	12.5, 20, and 30 mm	diameters
	beam node UB	$1200\,\text{MU}$	max. total MU coming out of one node
Resolution	optimisation grid	$1 \times 1 \times 1.5\,\text{mm}^3$	boost volume
		$2.9 \times 2.9 \times 3.0\,\text{mm}^3$	PTV, GTV, esoph.
		$2.9 \times 2.9 \times 1.5\,\text{mm}^3$	spn., spncrt.
	interaction grid	$2.9 \times 2.9 \times 3.0\,\text{mm}^3$	grid VOI

dose coverage, and OME optimisations, to reduce dosage to the spine, spinal channel, and esophagus. The starting solution for I3D planning is obtained by fixing the bounds of the VOIs to the values mentioned in Table 7.10. The resulting DVH curves shown in Figure 7.16b reveal a coverage of the GTV of 83.7 %, with only 55.2 % of the relapse volume receiving a dose of at least 30 Gy (also see first entry in Table 7.11).



(a) (b)

Figure 7.16: (a) The main VOIs involved in the spine case: PTV—green, relapse—black, critical spine area—cyan containing the critical spinal cord— dark blue, and the esophagus—magenta. (b) The DVH curves for the solution with which we start I3D planning: PTV—green, GTV—magenta, boost—black, critical spine—cyan, inner shell—dark gray, and outer shell—light gray. We use the same VOI colour codes across figures.

Table 7.10: The simple planning steps establishing an initial solution for the recurring spine tumour case. Each step targets a VOI and a given dose level; the result is saved by fixing the respective voxel bounds.

#	step	target VOI	target dose (Gy)	fix to (Gy)
1	OCO	PTV	25.50	25.00
2	OCO	GTV	30.50	30.00
3	OME	spncrt	0.00	8.50
4	OME	spn.	0.00	15.00
5	OME	esoph.	0.00	15.00

7.4.2 I3D Planning

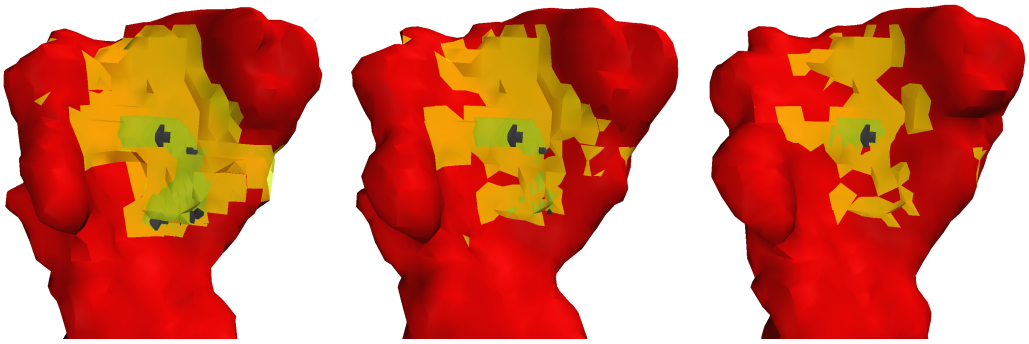
We import the previous solution and, due to runtime constraints, we reduce the number of beams from 3000 to 1600, of which 336 are weighted, as they form the solution set. First, we setup the optimisation goal by expanding the 30 Gy target isodose to cover the whole region of the GTV facing the critical spine (see Figure 7.17). As expected, the SCO step results in only a slight goal improvement. We then proceed to study the localised trade-offs that might advance our objective. Each step we discuss is summarised in Table 7.11, while the evolution of the 30 Gy isodose surface in relation to the target surface is shown in Figure 7.18.

(a) Before target selection (b) During target selection (c) Final target selection

Figure 7.17: Selection of the target area for the first I3D planning scenario of the spine case: the 30 Gy isodose surface (red), GTV (magenta), relapse volume (black), and target surface (yellow). The planner directly selects the target area with the spherical bumper (cyan, indicated by arrow the in (b)) and expands it to cover the region of interest (c).

The first trade-off is to allow a flatter dose gradient, by relaxing both shells in the directions that appear to be limiting, yet excluding those areas towards sensitive structures (such as the esophagus or the heart). We allow up to 22 Gy and 16.50 Gy through the inner and outer shells respectively. This process is shown in Figure 7.19: the planner selects the regions of relaxation directly on the isodose surface with the threshold dose equal to the upper bound of the shell. Because only a small number of shell voxels prove to be limiting, their relaxation leads to only a modest increase in target coverage, from 5.6 % to 8.1 % (entry 3 in Table 7.11).

A number of spots within the PTV and GTV appear as limiting (Figure 7.20a). Consequently, the next trade-off we perform is to allow more dose deep within the gross tumour and away from any sensitive structures. In this region, the tumour mass has replaced most of the bone tissue. We select the relaxations directly on the isodose surfaces



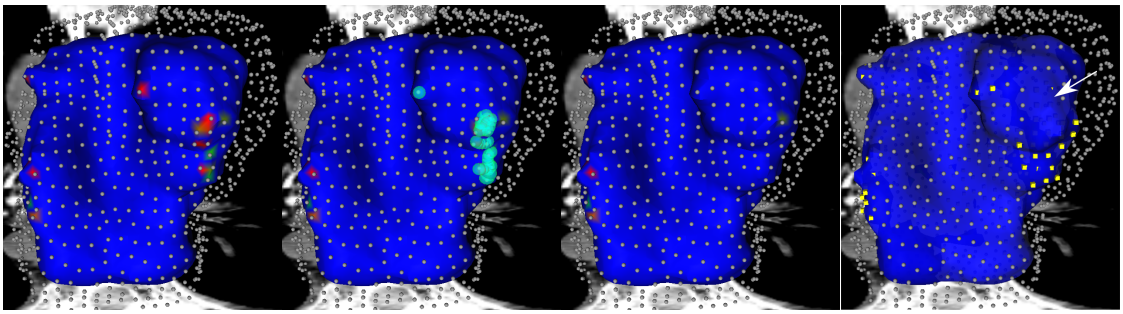
(a) Initial solution (b) After conservative relaxations (c) Final solution

Figure 7.18: Evolution of the 30 Gy isodose surface (red) as it expands to conform to the target surface (yellow, partially transparent): (a) initial solution, (b) after moderate relaxations (step 6 in Table 7.11) and (c) final solution after I3D planning. The target area covers most of the recurrent tumour (black).

(isodose values close to 35 Gy) coloured to highlight the limiting areas and we allow a dose of up to 38 Gy. As a result, the 35 Gy isodose expands within the GTV, increasing the target coverage by 25 %, to 33.1 %. The relaxation the PTV/GTV upper bounds is depicted in Figures 7.20 a–d. As more dose is allowed inside the PTV, we would expect new shell voxels to become limiting. However, this is not the case, which indicates that the shells have allow influence over the optimisation goal.

Following both shell and PTV/GTV trade-offs, the respective bounds are no longer limiting. Indeed, our I3D planning tool indicates that only the PTV/GTV lower bounds and the critical spine are highly limiting (Figure 7.21 a, b: the isodose surfaces corresponding to the respective upper bounds are depicted as cold—blue). While one could attempt to trade-off coverage in one area of the PTV/GTV for coverage in the target area, this would obviously be counter productive. Furthermore, such a trade-off is generally unlikely to bring a big improvement to the optimisation goal.

Consequently, we are left with only one further option: to relax the critical spine, as it imposes a very steep dose gradient towards the target. Analysing the surface of the critical spine (Figure 7.22a), we find regions that are highly limiting. Although only some areas and not the entire PTV facing surface of the spine are limiting, the strict bounds and the proximity to the PTV do hint at the opportunity of relaxation here. We therefore settle to allow an increased dose of maximum 10 Gy, exclusively on the surface of the critical spine and not to its interior. The motivation for relaxing the critical spine is two-fold: first, we improve the objective and, second, the resulting dose distribution may open



(a) Limiting voxels (b) Relaxation selection (c) Relaxation performed (d) Optimisation after relaxation

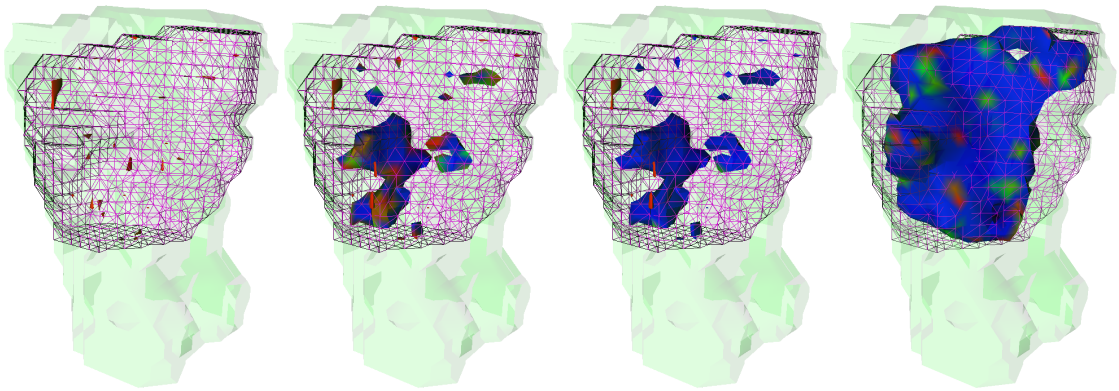
Figure 7.19: Relaxation of SHELL1 (gray spheres) by interacting with the $15\,\mathrm{Gy}$ isodose surface. The limiting areas, showing up as red in (a), are selected in (b) and, following relaxation, the surface colouring changes to indicate that region is no longer limiting (c). Following optimisation (d), the isodose surface expands in the area of relaxation (indicated by the arrow), covering some shell voxels. The relaxed voxels are drawn as yellow boxes.

new relaxation possibilities deep within the GTV, where trade-offs are more acceptable. Following relaxation, the target coverage gains $23.3\,\%$ for a total of $56.4\,\%$. On the other hand, the critical spine $V_{8.5}$ rises to almost $7\,\%$, as the respective isodose surface penetrates its surface (see its evolution in Figures 7.22d–7.22f). All other statistics are listed in entry 4 in Table 7.11.

The critical spine proved to be very limiting, as the last trade-off step led to important improvement in the objective. We therefore further relax highly limiting regions on the surface of the critical spine to $11\,\mathrm{Gy}$. Once again, the target coverage increases to $75\,\%$, while critical spine $V_{8.5}$ almost doubles, as the relaxation slack is used up.

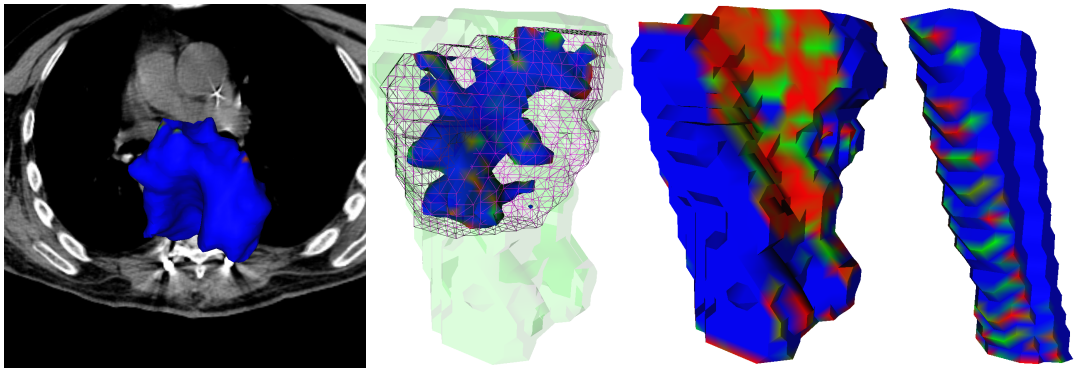
The previous two-staged relaxation of the critical spine leads to new areas within the PTV and GTV to become limiting. We choose to relax only those regions contained within the GTV to $38\,\mathrm{Gy}$ (obviously, the PTV encloses the GTV). Additionally, we allow another $1\,\mathrm{Gy}$ to those parts of the PTV/GTV that we previously relaxed, and that now are again limiting. As the objective advances, so do areas of the critical spine become limiting once more. We therefore continue with a series of alternating relaxations of the critical spine and of the PTV/GTV upper bound in the same areas as before, and to a maximum of $13\,\mathrm{Gy}$ and $40\,\mathrm{Gy}$ respectively (see entries in Table 7.11.

Finally, we establish Pareto efficiency of the solution, which results in no improved criteria and no recovered relaxation slack (compare entries (9) and (10) in Table 7.11). The dosimetric results are summarised by the DVH curves in Figure 7.23. Taking a closer look at Figure 7.23d, we can see the mean dose of the non-relaxed subvolume also increasing



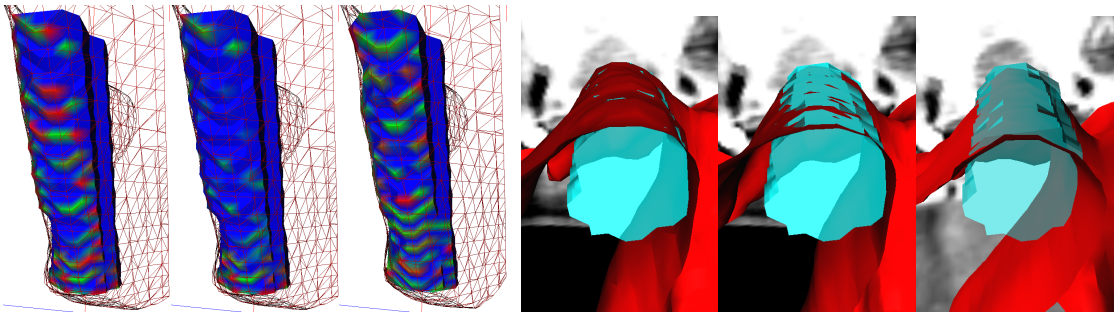
(a) (b) (c) (d)

Figure 7.20: Relaxation of the PTV (green, transparent) and GTV (magenta, wireframe) upper bounds deep within the GTV by direct interaction with the iso dose surface (limiting areas highlighted in red). (a) We select the relaxation regions on the $34.90\,\mathrm{Gy}$ isodose surface, which is made up of limiting voxels. (b) After relaxation, the isodose surface expands within the GTV. We further relax limiting areas, and the colouring changes accordingly (c). (d) The same isodose surface after I3D planning is still contained deep within the GTV.



(a) (b) (c) (d)

Figure 7.21: (a, b) After shell and PTV/GTV upper bound relaxation, the $20\,\mathrm{Gy}$ and $35\,\mathrm{Gy}$ isodose surfaces (coloured to highlight limiting areas in red) show very limited opportunities for further trade-offs. (c, d) The $20\,\mathrm{Gy}$ lower bounding ((c)—cold regions belonging to the PTV are in red) and the $9\,\mathrm{Gy}$ upper bounding ((d)—hot regions of the critical spine in red) surfaces corresponding to the PTV and critical spine respectively, are the two main limiting structures remaining.



(a) (b) (c) (d) (e) (f)

Figure 7.22: (a–c) Critical spine relaxation by sculpting out the $9.5\,\mathrm{Gy}$ upper bounding surface (limiting areas are highlighted in red): (a) before relaxation, (b) after relaxation—the colouring changes accordingly, (c) after optimisation—new areas become limiting. (d–f) The evolution of the $8.5\,\mathrm{Gy}$ isodose surface as a result of critical spine relaxation: (d) before any relaxation, (e) after conservative relaxation (step 5 in Table 7.11), and (f) after aggressive relaxation (step 6 in Table 7.11). Notice that even after aggressive relaxation, the isodose surface penetrates only superficially in the spine.

quite a bit during planning (although no voxel violates its bound of $8.5\,\mathrm{Gy}$). Such an effect has not been seen in the planning of the previous cases and it shows just how limiting the critical spine is to this particular optimisation goal. Additionally, the fact that we are increasing the dose inside an important area of the PTV, well beyond the initial values, is bound to drive up the mean dose of proximal organs.

Figure 7.24 shows the dose distribution as isodose curves overlayed on CT slices. As per our approach, the $35\,\mathrm{Gy}$ isodose curve expands exclusively within the GTV, while the $8.5\,\mathrm{Gy}$ curve cuts through only a thin part of the critical spine surface.

Table 7.11: Summary of the recurring spine tumour I3D planning steps to increase dose coverage and the result of IDVH planning: 1—initial solution, 2 — target OCO optimisation, 3—shell relaxation, 4—PTV/GTV UB relaxation, 5—spncrt. relaxation, 6—2nd spncrt. relaxation, 7—2nd PTG/GTV relaxation, 8—3rd spncrt. relaxation, 9—3rd PTG/GTV relaxation, 10—Pareto solution

	target	relapse		PTV		GTV		spncrt.		SHELL 1/2
step	CO (%)	CO	min (Gy)	CO	HI/CI	CO	V_{35} (%)	$V_{8.5}$	max(Gy)	max (Gy)
1.	0.0	55.2	16.89	90.8	1.55/1.41	83.7	0.0	1.5	8.70	20.00/15.00
2.	5.6	62.4	17.66	91.0	1.55/1.43	85.8	0.0	1.5	8.70	20.00/15.00
3.	8.1	64.0	18.09	91.1	1.55/1.48	86.7	0.0	1.4	8.70	22.00/16.50
4.	33.1	74.8	19.53	91.4	1.69/1.46	89.9	22.2	1.4	8.70	22.00/16.50
5.	56.4	87.2	22.63	92.2	1.69/1.49	91.8	19.9	6.9	10.00	22.00/16.50
6.	75.0	93.9	24.55	92.7	1.69/1.49	92.9	26.1	13.9	11.00	22.00/16.50
7.	77.4	94.9	24.89	92.7	1.73/1.49	93.4	30.6	13.9	11.00	22.00/16.50
8.	81.9	95.6	25.57	93.2	1.73/1.47	94.4	28.0	19.6	13.00	22.00/16.50
9.	87.4	97.2	25.80	93.3	1.77/1.45	95.3	38.6	19.6	13.00	22.00/16.50
10	87.4	97.2	25.80	93.3	1.77/1.45	95.3	38.6	19.6	13.00	22.00/16.50
IDVH	96.8	99.0	27.41	93.8	1.77/1.49	96.4	52.0	37.3	13.00	22.00/16.50

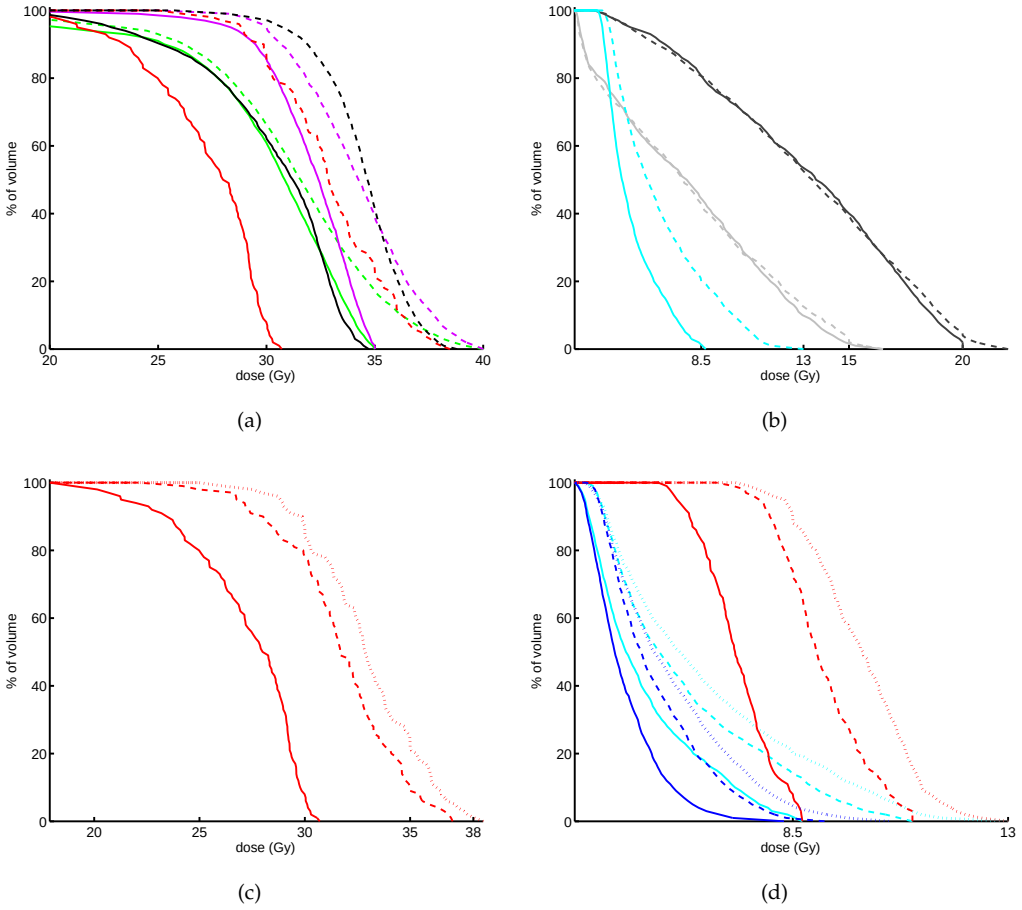
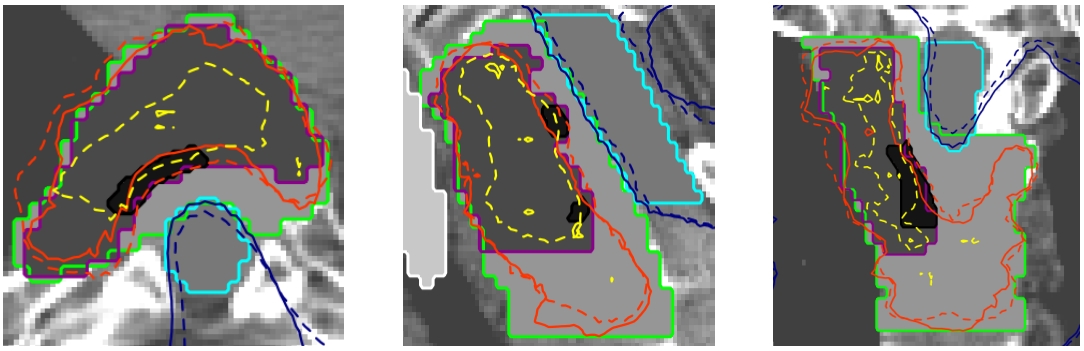


Figure 7.23: (a, b) DVH curves summarising the dose distribution for the spine case, before (solid) and after (dashed) I3D planning. (a) DVH curves for the target (red), PTV (green), GTV (magenta), and relapse volume (black). (b) DVH curves for the critical spine (cyan), and for the inner (dark gray) and outer (light gray) shells. (c) DVH curves for the target. (d) DVH curves for the critical spine (cyan) and for its relaxed (red), and non-relaxed (blue) subvolumes. In (c) and (d) the initial solution is represented by solid curves, the solution after conservative critical spine relaxation is dashed, and the final solution is dotted. After relaxation, mean dose to the non-relaxed subvolume increases, yet it does not violate the original bound of $8.5\,\%$.



(a) (b) (c)

Figure 7.24: Axial (a), sagittal (b), and coronal (c) slices showing the $8.5\,\mathrm{Gy}$, $30\,\mathrm{Gy}$, and $35\,\mathrm{Gy}$ isodose curves (blue, red, and yellow, respectively) before (solid) and after (dashed) I3D planning. Also contoured are the PTV (green), relapse (black), critical spine (cyan), and esophagus (white). Notice how the $35\,\mathrm{Gy}$ isodose is mostly contained inside the GTV and how the $8.5\,\mathrm{Gy}$ isodose barely cuts right under the critical spine surface.

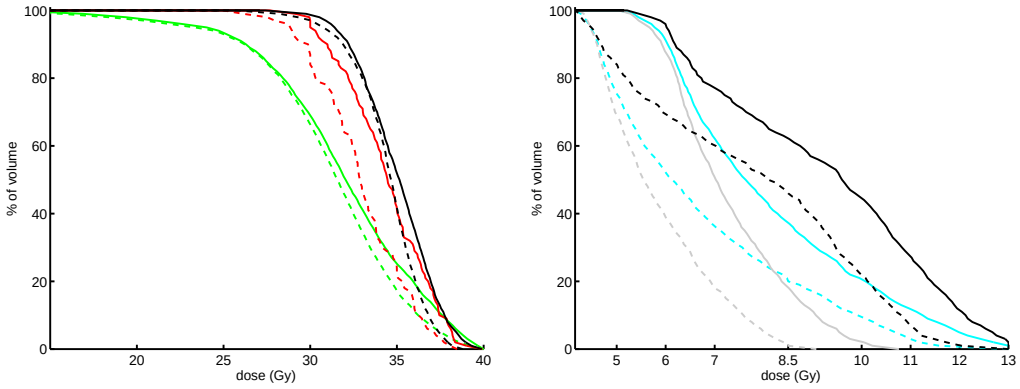
7.4.3 IDVH planning

Next, we use the IDVH tool to arrive at a comparable result, starting from the same initial solution. Note that we reuse the optimisation target we defined with the I3D tool—the area inside the GTV containing the relapse—as the IDVH tool can only use global VOI targets.

After the initial optimisation step we need to perform some trade-offs to improve the goal. First, we relax the inner and outer shells to 16.5 Gy and 22 Gy, raising target coverage from 5.6 % to 11.8 %. Second, we relax the PTV and GTV upper bounds to the conservative value of 38 Gy; target coverage increases to 42.5 %. Third, we allow 11 Gy to the critical spine, resulting in a target coverage of 74.8 %. We continue alternating PTV/GTV and critical spine relaxations, to a maximum of 40 Gy, for the former, and 13 Gy, for the latter. The final target coverage comes out at 96.8 %; dosage statistics for the this result are listed in the last entry of Table 7.11.

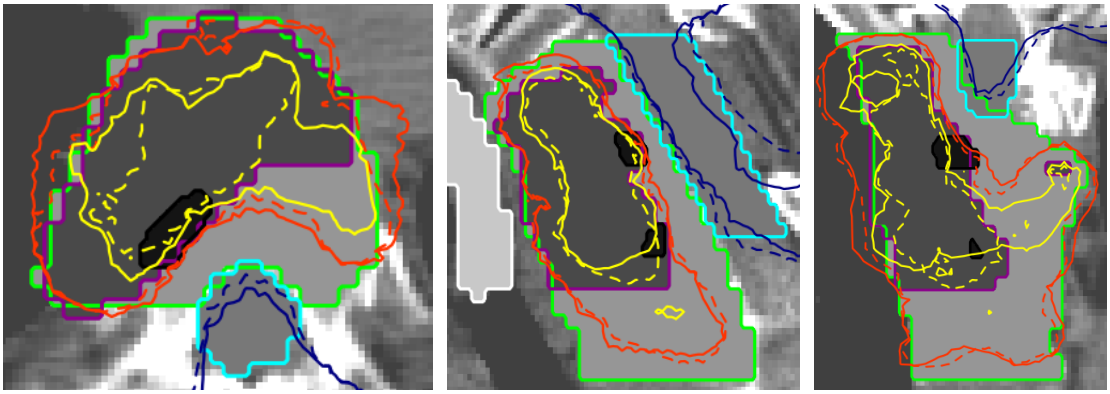
Comparing the DVH curves in Figure 7.25, we clearly see that the IDVH global relaxations lead to better target coverage (IDVH: 96.8 % versus I3D: 87.4 %). However, this comes at a cost to other criteria. First, hotspots of over 35 Gy extend beyond the GTV, well within the original PTV margins, as depicted in Figures 7.26 a and c. Second, the 8.5 Gy isodose curve extends deep into the critical spine for IDVH, while, for I3D, it only cuts through the surface (compare the extent of the 8.5 Gyisodoses in Figures 7.26 a–c). Indeed, the DVH curves in Figure 7.26b, show that relaxation of the spine during I3D planning leads to an overdosage only for its surface. The percentage of the interior of the spine, receiving a dose of at least 8.5 Gy ($V_{8.5}$) is zero for I3D and as much as 18.3 % for IDVH.

Spatial multicriteria optimisation cannot improve a plan beyond what is physically achievable by the delivery system. However, it does allow the planner to explore localised trade-offs, in an attempt to precisely form the shape of the dose distribution. Due to the fact that, within the linear programming formulation, voxel doses and bounds are coupled with each other, one trade-off may lead to new areas becoming limiting. Consequently, I3D planning resembles a greedy search for a local optimum. At each step, the planner performs the trade-off that he considers as being most advantageous. Obviously, it is impractical to analyse every limiting structure after each relaxation; instead, one must systematically consider the potential relaxations, on a per-VOI basis. After all trade-offs for that one VOI have been performed, the planner may proceed to a different VOI, including to one that has already been relaxed.



(a) (b)

Figure 7.25: DVH curves for the solutions obtained with IDVH (solid) and I3D (dashed) tools. (a) DVHs for the target (red), PTV (green), and relapse volume (black). Target coverage is almost $10\,\%$ higher with IDVH. (b) DVHs for the entire critical spine (cyan) and for its internal (gray) and surface (black) subvolumes. Notice that almost none of the internal subvolume receives a dose of over $8.5\,\mathrm{Gy}$ after I3D planning, as relaxation was limited to the surface of the critical spine.



(a) (b) (c)

Figure 7.26: Axial (a), sagittal (b), and coronal (c) slices showing the $8.5\,\mathrm{Gy}$, $30\,\mathrm{Gy}$, and $35\,\mathrm{Gy}$ isodose curves (blue, red, and yellow respectively) after IDVH (solid) and after I3D (dashed) planning. Also contoured are the PTV (green), relapse (black), critical spine (cyan), and esophagus (white). Through precise local relaxations, I3D overdosages are contained to the GTV and to the surface of the critical spine. This is not the case for IDVH: the $35\,\mathrm{Gy}$ isodose spills outside the GTV and the $7\,\mathrm{Gy}$ isodose cuts deeper into the critical spine.

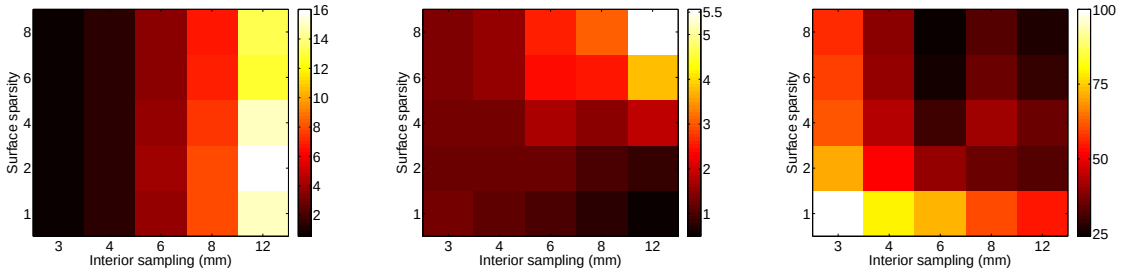
7.5 Anatomical Space Sampling

In the following sections we shall show the influence of the sparse sampling strategy from Section 6.1 on the optimisation runtime and on the dosage errors.

7.5.1 Rind And Interior Sampling

We use the two resolution approach for the PTV rind and interior of the prostate cancer case from Section 7.2. For the rind we set the sparsity factor to 1, 2, 4, 6, and 8 and we sample the interior with homogeneous voxel sizes of 3 mm, 4 mm, 6 mm, 8 mm, and 12 mm. For each parameter combination, surface sparsity and resolution, we run ten OCO optimisations, each with a different beamset consisting of 2000 beams. All other VOIs are uniformly sampled at a rate of 3 mm. As a baseline, we sample the PTV uniformly at 2 mm. We then project the results of the optimisation on a CT resolution grid and we define the following errors:

- Relative coverage error $e_{CO} = 100\left(1 - \frac{CO_k}{CO_0}\right)$, where CO_k is the PTV coverage, for trial k / baseline (0).
- Relative hotspot error $e_H = 100\frac{V^k_{41.50}}{V^0_{41.50}}$, where $V_{41.50}$ is the volume of the PTV receiving a dose over its upper bound of 41.50 Gy, for trial k / baseline (0).



(a) Hotspot error (%) (b) CO error (%) (c) Runtime (%)

Figure 7.27: (a) Mean relative hotspot errors, (b) mean relative coverage errors, and (c) mean relative runtimes for different interior PTV sampling rates and surface sparsities.

Figure 7.27 shows the variation of the mean relative errors and of the mean relative runtimes. The results reinforce our expectations: hotspot errors are highly dependent on the sampling rate of the target interior, while coverage errors are more dependent on the sparsity of the target surface. In fact, PTV coverage is maximum for low surface sparsity and low interior resolution. This result is unsurprising if we note that a reduced

number of sample points inside the PTV lead to a less constrained optimisation problem, allowing for dose distributions that create hotspots inside the PTV. At the same time, the dense surface ensures that no parts of the PTV are excluded.

The results reinforce the importance of carefully choosing the sampling resolution and the sparsity before planning.

7.5.2 Hotspot Error Compensation

We exemplify the hotspot compensation method from Section 6.1.2 on the prostate cancer case. The following results were published in [124].

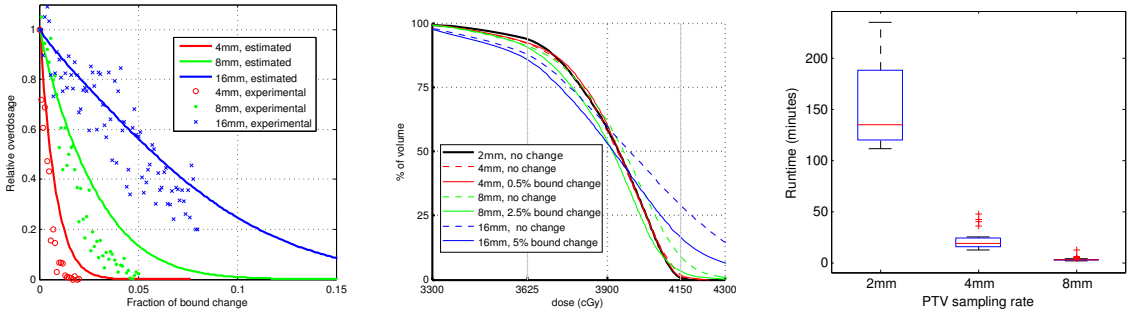
First, we estimate the reduction of the PTV overdosage due to upper bound tightening at three different uniform sampling rates: 4 mm, 8 mm, and 16 mm. For each voxel size, we randomly generate $50\,000$ beam configurations, each consisting of 260 equally weighted beams. The beam generation parameters, number of weighted beams, and collimator distribution (probability that a collimator of certain size is used), were determined by running ten simple OCO steps and averaging over the solutions. The mean collimator distribution favoured collimators with radii less than 30 mm. Estimation of other beam generation parameters, such as beam node and beam direction distributions, may improve the estimated overdosage PMF.

Second, we run a series of OCO optimisation steps with a target dose of 36.25 Gy ($80\,\%$ isodose) for which we reduce the PTV upper bound in increments of 4.15 cGy ($0.01\,\%$ of the original bound). We repeat this for each of three uniform sampling rates and we compute the corresponding dose distribution on a CT resolution grid. The relative hotspot error for a given upper bounds value b_u is then given by

$$e_H = \frac{V_{41.50}^{b_u}}{V_{41.50}^{41.50}}$$

where, $V_{41.50}^{b_u}$ is the percentage of the PTV volume receiving a dose over 41.50 Gy, when the upper bound is set to b_u.

Both estimation and experimentally measured error reductions are plotted in Figure 7.28a and show good agreement with each other. Hotspot error is reduced to $0.75\,\%$ and $2.2\,\%$, from original values of $1.71\,\%$ and $9\,\%$, by simply tightening the PTV upper bound by $0.5\,\%$ and $2.5\,\%$ for voxel sizes of 4 mm and 8 mm respectively. The DVH curves in Figure 7.28b show these error reductions and also make evident that a sampling rate of 16 mm is far too coarse to compensate through bound tightening. Our main interest in error compensation is shown in Figure 7.28c: at lower resolutions, the optimisation runtime drops dramatically.



(a) (b) (c)

Figure 7.28: Hotspot error compensation. (a) Estimated (solid curves) and experimental (markers) drop in hotspot error, as a function of upper bound tightening, for $4\,\mathrm{mm}$, $8\,\mathrm{mm}$, **and** $16\,\mathrm{mm}$ **voxel sizes (red, green, and blue respectively). (b) The PTV DVH curves before (solid) and after (dashed) hotspot compensation by upper bound reduction. (c) Runtime boxplot at different uniform sampling rates of the PTV.**

7.6 Beam Space Sampling

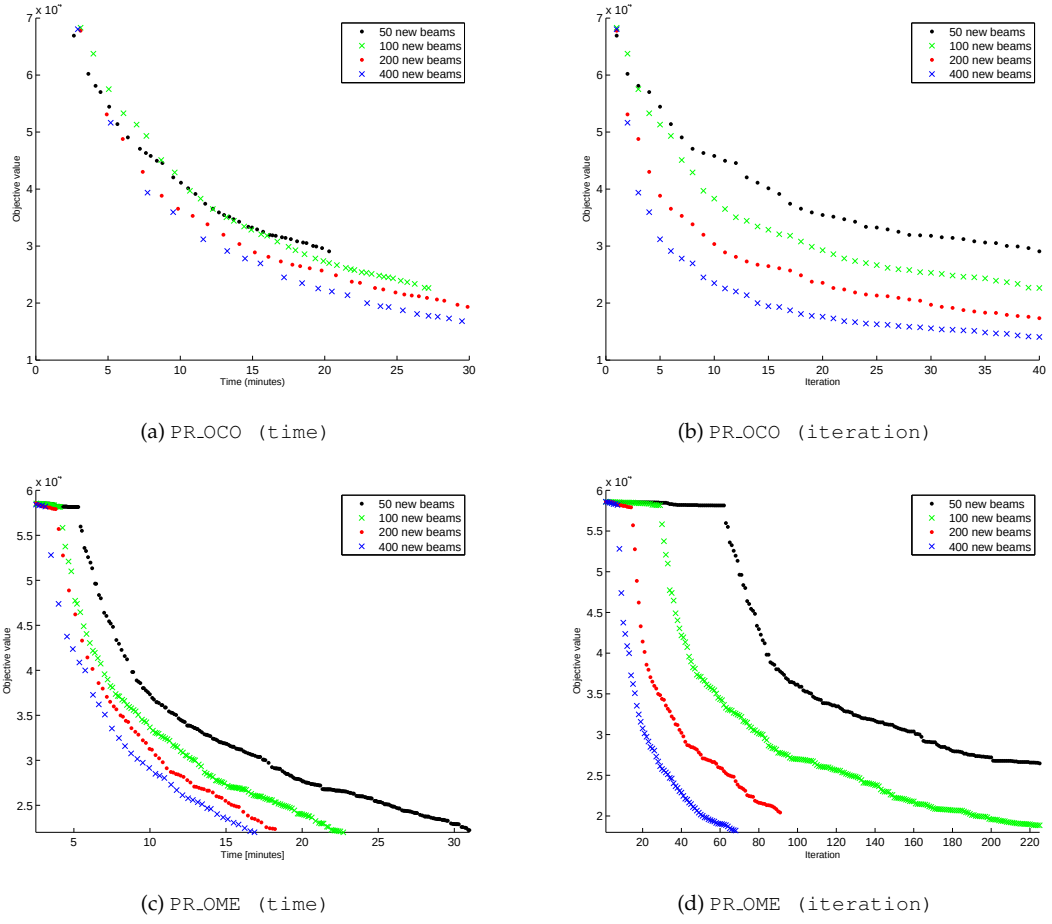
We have seen how reducing the number of voxels leads to smaller optimisation problems and thus, reduced computational overhead. Due to the same runtime considerations, we are compelled to limit the beamset size. In Section 6.2 we have proposed using beam resampling to compensate for the loss in solution quality when optimising with a reduced beamset. In this section we evaluate the possibility of integrating beam resampling into the planning workflow. We first take a look at the classical beam resampling algorithm and we demonstrate the potential speedup of our concurrent implementation. Second, we show how beam resampling can be used as a final step in the planning workflow and we discuss the implications of such an approach.

7.6.1 Random Beam Resampling

The random beam resampling algorithm from Listing 1 depends mainly on two parameters: the total number of beams entering the optimisation and the number of unweighted beams that are replaced at each iteration. Obviously, the larger the number of total beams, the larger the objective improvement per iteration will be. On the other hand, understanding the influence of the number of replaced beams is not straight forward. First, by replacing a large number of beams, the new solution will be further away from the last optimisation solution and, thus, the optimisation will take longer. Second, the time required to modify the optimiser, by removing and adding beams, increases as the number

of beams gets larger. Third, it is unclear whether resampling a small amount of beams results in the same improvement per iteration as resampling a larger amount of beams, as the optimisation problem is combinatorial in nature.

Figure 7.29 shows how the number of replacement beams influences the improvement of the objective (objective value reduction between two iterations), for two different optimisation goals of the prostate case from Section 7.2. First optimisation is an OCO step to establish initial PTV coverage, which we call `PR_OCO`. Second is an OME step to spare the rectum to 32 Gy, starting from an established dose distribution. This we denote as `PR_OME`. The total number of beams was 700 and, at each iteration, we replaced a maximum of 50, 100, 200, and 400 unweighted beams with new randomly generated ones.



(a) `PR_OCO (time)` (b) `PR_OCO (iteration)`

(c) `PR_OME (time)` (d) `PR_OME (iteration)`

Figure 7.29: The evolution of the objective value for `PR_OCO` (a, b) and `PR_OME` (c, d) during beam resampling with 50, 100, 200, and 400 new beams per iteration. Notice the plateau width for `PR_OME` is narrower for increased number of new beams.

The goals considered are representative for the types that are relevant in beam resampling. PR_OCO sees the greatest improvement per iteration in the first stages, as the lack of previous optimisation steps means no additional bounds were tightened to overly constrain the problem. Conversely, PR_OME takes place in a much more constrained space, as the previous results are preserved by tight bounds (i.e. PTV bounds are set to preserve coverage). Any new solution must obey these bounds and as such there is a plateau during the first iterations of the resampling process (see Figures 7.29 c and d). A large number of beam combinations must be tried out before a solution is found that nudges the optimisation off the plateau. Clearly, the width of the plateau is highly dependent on the number of beam combinations that were tried, as larger beamsets lead to narrower plateaus. A possible workaround to reduce this width would be to perturb the linear program, by slightly relaxing limiting bounds. However, this introduces additional complications, such as setting the magnitude and number of perturbations to meaningful values.

While the results clearly show that resampling a larger number of beams increases the objective improvement per iteration (Figures 7.29 b and d), when plotted against optimisation times (Figures 7.29 a and c), this advantage is less clear. This, again, is due to the increased optimisation time that is needed in finding a solution that is further away (in the search space) from the previous one. Still, the larger number of resampled beams performed consistently, if only slightly better.

7.6.2 Concurrent Beam Space Resampling

We next investigate the speedup potential of our concurrent beam resampling algorithm (CBR, Listing 5) over the serial implementation (RBR, Listing 1). The results have been published in [125]. We use the same scenarios for the prostate case, PR_OCO and PR_OME, for which we run concurrent beam resampling with 1, 2, 4, and 8 threads. Note that the single-threaded run coincides with the classic random beam resampling algorithm.

Figure 7.30 shows the evolution of the objective value over time and over each iteration for the different number of threads (objective value is reported for one of the threads in the thread-pool). We clearly see that the overhead of running parallel instances of the CPLEX solver actually results in a slowdown in the improvement of the objective value over time (see Figures 7.30 a and c). This slowdown is consistent across all scenarios and it increases as the number of parallel threads is raised. It is sensible to speculate, that such an effect is due to the threads competing for shared system resources, in particularly for memory access.

On the other hand, when plotted against iteration number (see Figures 7.30 b and

d) the decrease in objective value sees an obvious speedup as the number of threads increases. During the first iterations of PR_OCO, the objective improvement is similar for all thread counts. In the beginning, solutions found by the peer threads and proposed as external solutions, are almost as good as any random beamset, due to the fact that a *good* dose distribution has yet to be established. Only after the problem becomes more difficult, do random beams no longer make the cut and concurrent beam resampling starts to show a clear improvement (compare the number of iterations required to reach the same objective indicated by the horizontal line in Figure 7.30b). Nevertheless, the speedup per iteration is consistent across thread count, with 8 threads giving the best performance. For PR_OME, increasing the thread count decreases the width of the plateau in addition to speeding up objective improvement. Indeed, the plateau width for concurrent beam resampling is roughly the width of the plateau of the serial run divided to the number of threads (ideal linear speedup).

Seeing how beam resampling is a stochastic process, we wanted to analyse if the speedup per iteration remains consistent when pitted against multiple single-threaded runs. We again perform beam resampling with 1, 2, 4, and 8 threads for a spine cancer case (SP_OCO), but this time we repeat single-threaded resampling for 50 different seeds for the random beam generator. The optimisation goal is PTV coverage at $30\,\text{Gy}$, starting from an advanced solution obtained by I3D planning. Figure 7.31 shows that concurrent resampling with 8 threads reaches a coverage of $97\,\%$ in approximately $36\,\%$ of the iterations needed by the median serial run.

The individual speedups in Table 7.12 were computed as $S = N_s(o)/N_c(o)$, where N_s and N_c are the number of serial and concurrent iterations respectively, needed to achieve the objective value o. Speedup scales with the number of threads, but it quickly levels out as shown in Figure 7.32a, where the external improvement for 4 and 8 threads is very similar.

Table 7.12: Speedup of the concurrent beam resampling algorithm relative to the serial implementation ($S = N_s(o)/N_c(o)$, where $N(o)$ is the number of iterations necessary to achieve an objective at least as good as o), for reaching the objective value indicated by the horizontal lines in Figures 7.30 b and d (PR_OCO and PR_OME), and for reaching a coverage of $97\,\%$ for SP_OCO.

	PR_OCO			PR_OME			SP_OCO		
	2	4	8	2	4	8	2	4	8
speedup	1.7	2.5	4.2	1.7	2.3	3.0	1.3	1.7	2.8

The boxplots in Figure 7.32a show the absolute objective value improvement for internal and external solutions: $\Delta_{obj} = o^{prev} - o^{curr}$, where o^{curr} is the objective value obtained with either the internal or the external solution; o^{prev} is the previous objective value. As the thread count increases, the external objective improvement gets larger. This is to be expected, as there is a higher chance that a concurrently running thread will find and propose a better solution. Figure 7.32b shows the distribution of the objective improvement for external and internal beamsets. The magnitude of the objective improvement for the random (internal) beamsets follows an almost exponential distribution, as small improvements are likeliest. On the other hand, external solution improvements are more uniformly spread, with higher gains just as likely as modest ones.

Finally, Figure 7.33 shows the evolution of the objective value for different values of the improvement threshold (T_{improv} in Listing 5) for 4 threads. This threshold controls whether an external solution is accepted or whether an internal solution is generated. The higher the threshold, the larger the improvement must be such that an external solution is accepted. The results indicate that, usually, any accepted exterior solution will perform at least as good as an internal solution. It is clear then that T_{improv} should be set to 0.

On a modern multicore system, our algorithm fails to show any gains in objective value improvement over time, due to threads competing over shared resources. However, the speedup per iteration would translate into real world gains when running the algorithm on a computing cluster. A computing unit shares no resources and a solution can be quickly communicated over the network to the other units. The network communication should represent no bottleneck, as the data size of the external solution is very small.

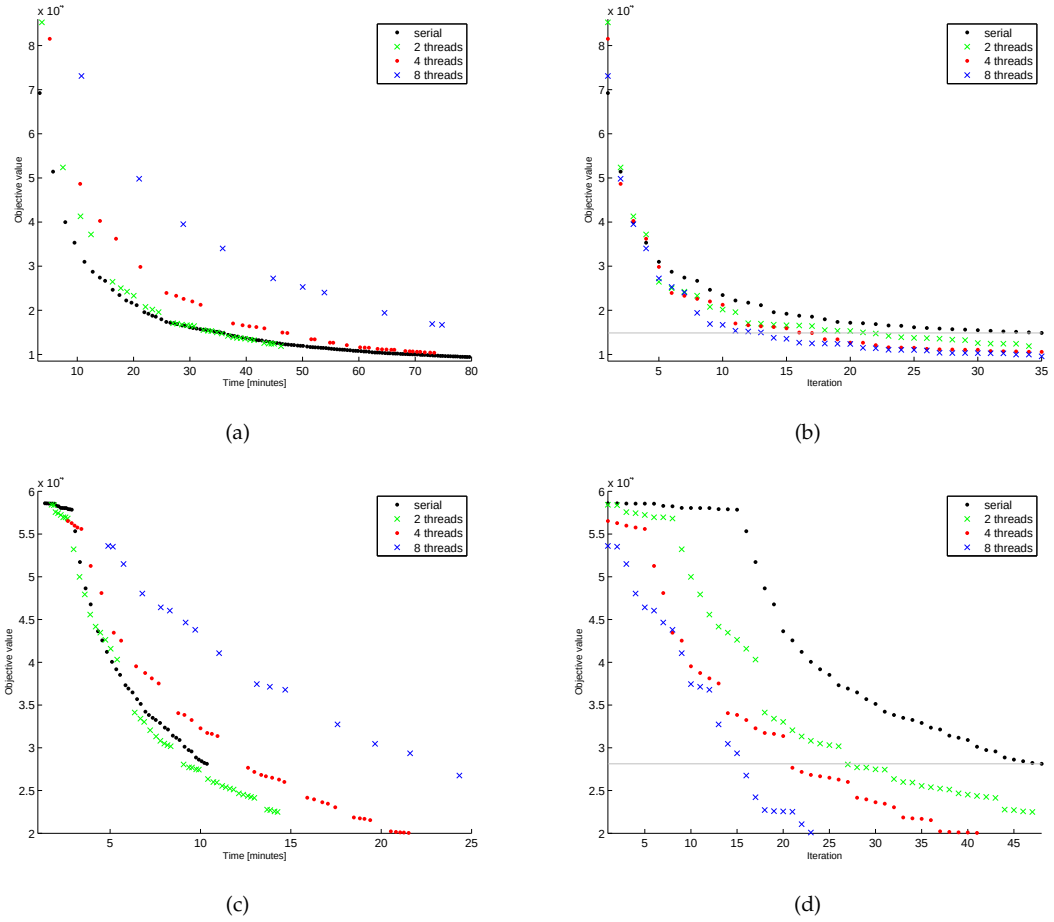
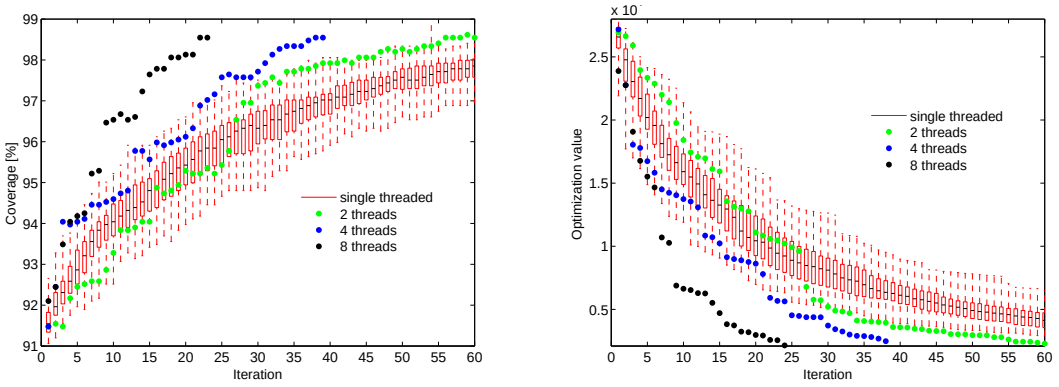
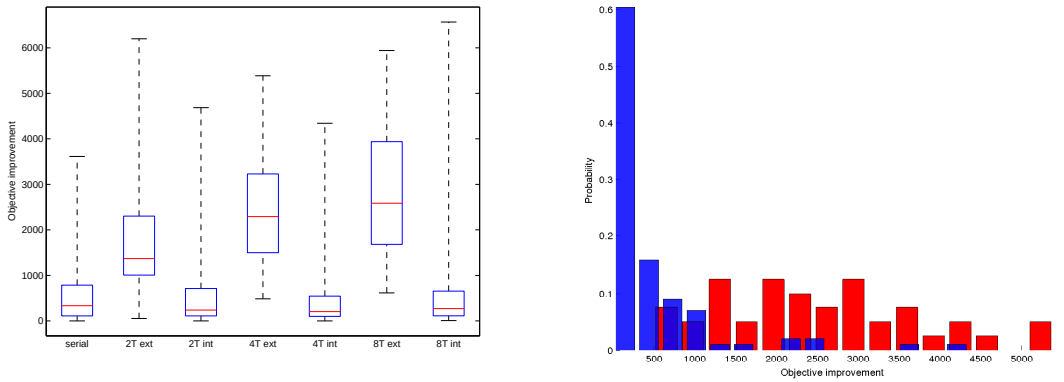


Figure 7.30: The evolution of the objective function for `PR_OCO` (a, b) and `PR_OME` (c, d) during concurrent beam resampling using 1, 2, 4, and 8 threads, plotted against time and iteration number. While the objective reduction per iteration is improved for the larger number of threads, the competition for shared resources leads to a slowdown when using multiple worker threads. Also notice in (d) the much narrower plateau for the larger number of threads, as more beam configurations are attempted simultaneously.



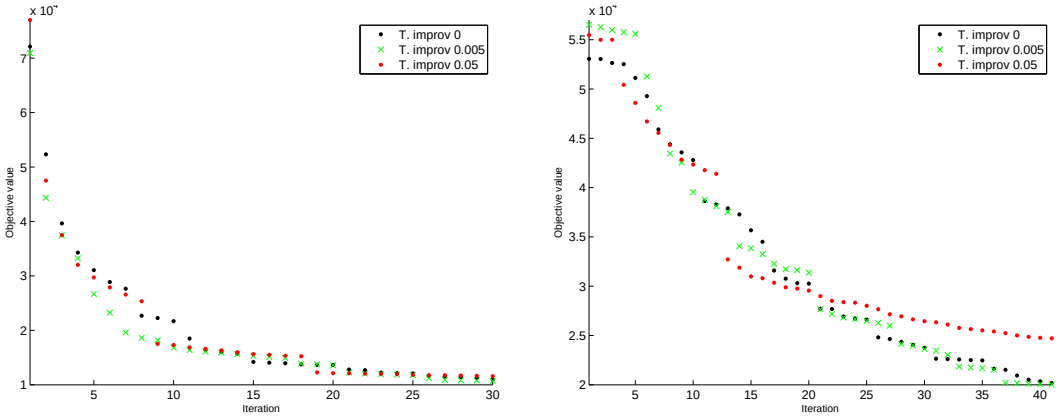
(a) (b)

Figure 7.31: The increase in target dose coverage (a) and the corresponding decrease in objective value (b) per iteration of beam resampling with 1, 2, 4, and 8 threads for SP_OCO. Note that the objective value is always non-increasing. While one would expect the target coverage to be non-decreasing, there are several instances that it does see a slight drop from one iteration to the next, due to the convex approximation of the objective.



(a) (b)

Figure 7.32: (a) The internal (*int*) and external (*ext*) objective value improvements for the concurrent beam resampling algorithm with 1, 2, 4, and 8 threads. (b) Probability mass function (PMF) of the internal (blue) and external (red) objective improvements for the concurrent beam resampling of the PR_OCO scenario. Notice how the PMF for external beamsets is a lot flatter than the PMF for internal beamsets, where the likeliest outcome is only a small improvement.



(a) (b)

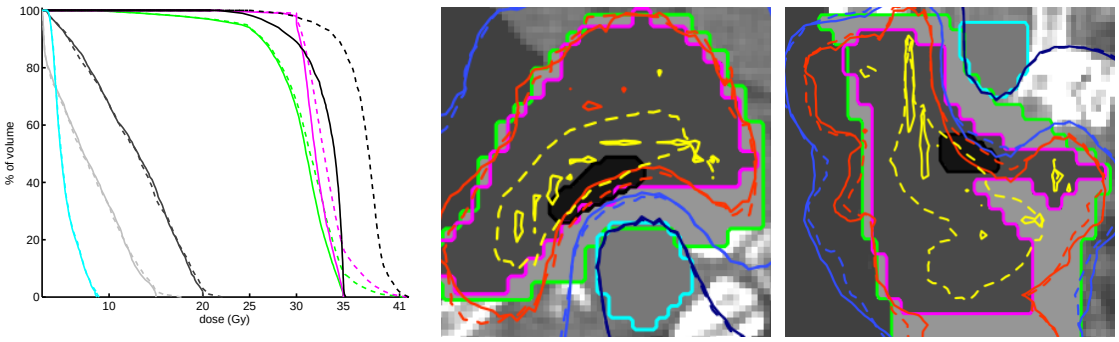
Figure 7.33: The influence of the improvement threshold on the evolution of the objective value, during concurrent beam resampling with four threads of the (a) `PR_OCO` and (b) `PR_OME` scenarios. Results indicate that a null-valued threshold gives good performance.

7.6.3 Integration of Beam Resampling with I3D Planning

Having shown the advantages of beam resampling and having established a method to reduce the time needed for its computation, we shall now give an example of integrating beam resampling within our planning workflow. Consider the spine case from Section 7.4, for which we choose a new goal: coverage of the relapse volume at $35\,\text{Gy}$. We first establish the entry solution for I3D planning by a series of beam resampling steps, with the same optimisation goals used to mirror the original plan (review Table 7.10). We import this solution into the I3D tool and we limit the beam count to 1200 (of which around 350 represent the solution) to keep computation time as low as possible. Next, we model the target dose in the shape of the relapse volume and we perform a series of relaxation steps, which eventually allow us to claim a coverage of $87.3\,\%$ for the relapse. Entries 1 to 4 in Table 7.13 summarise the interaction steps. The trade-offs were similar to the ones we already presented in Section 7.4: relaxation of the PTV and GTV upper bounds (mostly limited to the common PTV/GTV regions), relaxation of the shells (such that OARs are not affected), and relaxation of the critical spine surface. As a result, a $35\,\text{Gy}$ isodose appears deep within the PTV, engulfing much of the relapse volume (see Figures 7.34b and 7.34c). The increase in relapse volume coverage, as well as the dose effects due relaxations are summarised by the DVH curves in Figure 7.34a.

Table 7.13: Summary of the I3D planning steps for the spine case to increase dosage to the relapse volume. The first and last steps correspond to the solutions after beam resampling.

	relapse		PTV			GTV		spncrt.		SHELL1
step	CO (%)	min (Gy)	CO	HI	max (Gy)	CO	max	$V_{8.5}$ (%)	max	max
1. initial	5.8	21.68	94.6	1.55	35.00	98.4	35.00	1.2	8.70	15.00
2. r. PTV/GTV	80.7	23.06	94.8	1.83	41.26	98.9	41.00	1.1	8.72	15.00
3. r. shells	82.5	23.40	94.8	1.84	41.36	98.9	41.44	1.1	8.73	16.98
4. r. spncrt.	87.3	25.07	94.8	1.84	41.52	99.1	41.80	2.1	8.93	17.72
5. beam rsmpl.	90.0	25.88	94.8	1.84	41.46	99.1	41.56	2.0	9.00	15.00



(a) (b) (c)

Figure 7.34: DVH curves (a) and isodose curves (b, c) corresponding to the solutions before (solid) and after (dashed) I3D planning for relapse coverage. The different VOIs are depicted as follows: PTV—green, GTV—magenta, relapse—black, critical spine—cyan, inner shell—dark gray and outer shell—light gray. In (b) the isodose curves are shown as follows: $8.5\,\mathrm{Gy}$—dark blue, $20\,\mathrm{Gy}$—light blue, $30\,\mathrm{Gy}$—red and $35\,\mathrm{Gy}$—yellow. As result of I3D planning, relapse coverage increases to above $87\,\%$.

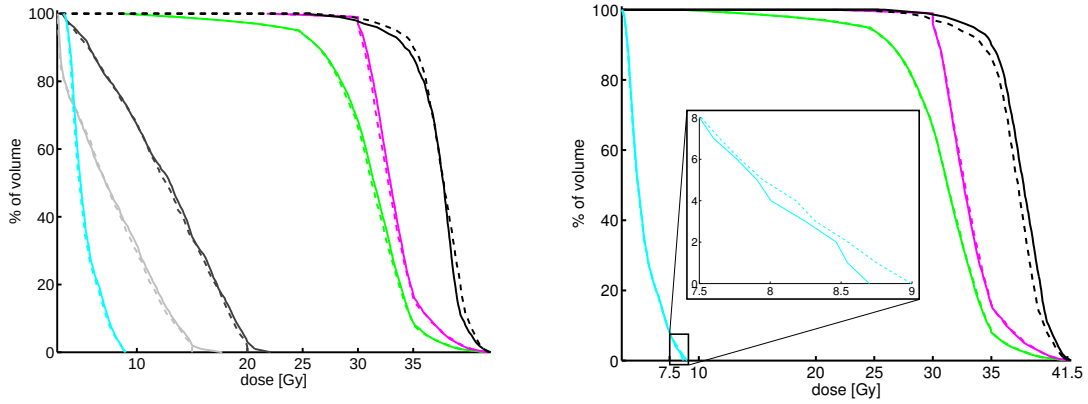
In an effort to recover solution quality, we perform beam resampling as the final planning step. Table 7.14 lists the complete optimisation goals considered during resampling. We first resample the same goal as during I3D planning, which further increases relapse coverage by $2.7\,\%$. Next, we attempt to recover the relaxations (steps 2 to 6) by starting out with the most sensitive ones. While we can barely reduce the extra dose to the critical spine, we completely recover the shell relaxations and we also slightly reduce overdosage to the target volumes: V_{35} for PTV and GTV drop from $8.8\,\%$ to $8.5\,\%$, and from $17.4\,\%$ to $16.2\,\%$, respectively. Final steps, 7 to 10, mirror those used to establish the initial dose distribution, yet fail to improve their objectives even slightly. Note that we resample these objectives last, as they had already been resampled going into I3D planning. As such, the only improvements we can expect are due to the subsequent localised trade-offs. Key dose statistics after beam resampling are listed in the last entry of Table 7.13.

The DVH curves in Figure 7.35a reveal the overall improvement brought by beam resampling, as coverage of the target increases and relaxations are recovered. Also notice how the $35\,\mathrm{Gy}$ isodose curve changes in Figure 7.36, as it expands inside the relapse volume, while at the same time it contracts in the other areas.

Although beam resampling may recover certain relaxations, there is no guarantee that this will happen and only those trade-offs that are clinically acceptable must be attempted. Indeed, if an extreme relaxation, beyond what is clinically tolerated, does prove to be recoverable, it is useless to the have performed it in the first place (although, this is

Table 7.14: The succession of optimisation goals during beam resampling of the I3D result for relapse volume coverage. Each entry shows: the optimisation step, the target VOI and dose, initial and final objective values and the number of iterations per resampling step.

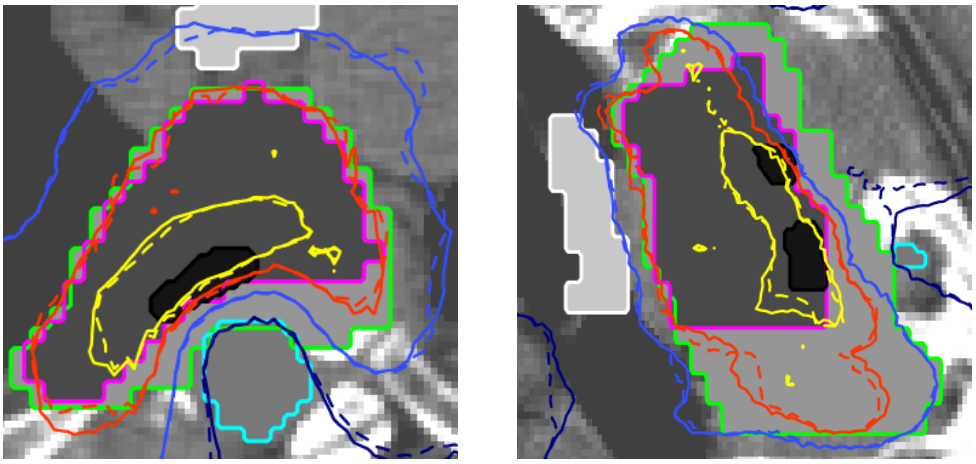
step	tg. VOI	tg. dose (Gy)	initial obj.	final obj.	# iterations
1. OCO	relapse	35.00	48 943.25	32 435.48	183
2. OME	spncrt.	7.00	21 014.94	19 281.32	91
3. OME	SHELL2	15.00	2094.15	0.00	235
4. OME	SHELL1	20.00	1651.06	0.00	302
5. OME	PTV	35.00	85 215.12	82 141.34	71
6. OME	GTV	35.00	76 996.92	76 595.10	42
7. OCO	PTV	25.00	154 163.85	154 155.38	42
8. OCO	GTV.	30.00	5783.89	5783.75	42
9. OME	esoph.	15.00	110 663.21	110 197.72	42
10. OME	spnchn.	15.00	136 916.66	136 902.61	42



(a) No resampling vs. resampling

(b) Resampling with vs. without relaxation

Figure 7.35: (a) DVH curves comparing the solution after I3D planning (solid) and after final beam resampling (dashed). Notice how the target coverage increases, while shell relaxations are completely recovered. (b) DVH curves comparing the solutions after beam resampling with (solid) and without (dashed) critical spine relaxation. Notice the trade-off of allowing only $0.25\,\mathrm{Gy}$ more to the critical spine allows for a considerable gain in relapse volume coverage. VOI colouring: relapse volume—black, PTV—green, GTV—magenta, critical spine—cyan, inner shell—dark gray, and outer shell—light gray shells.



(a) (b)

Figure 7.36: Axial (a) and sagittal (b) slices comparing the dose distributions after I3D planning (solid) and after the final beam resampling step (dashed) for the following dose values: $8.5\,\mathrm{Gy}$ **- dark blue,** $20\,\mathrm{Gy}$ **- light blue,** $30\,\mathrm{Gy}$ **- red and** $35\,\mathrm{Gy}$ **- yellow. The PTV (green), GTV (magenta), relapse (black), esophagus (white), and critical spine (cyan) are also shown.**

unknown to the user during planning). On the other hand, if the same relaxation is unrecoverable, the plan must be discarded anyway. In both cases, there is no good reason to have attempted this relaxation in the first place.

Not being able to recover certain trade-offs is indicative of their actual influence over the optimisation goal. This is the case, in the previous example, with the critical spine relaxation, which proves to be unrecoverable. To actually visualise its influence on the planning goal, we run the beam resampling script starting from a second I3D solution that we obtain with the exact same trade-offs, minus the spine relaxation. Comparing the DVH curves in Figure 7.35b, we can clearly see that just allowing the extra $0.25\,\mathrm{Gy}$ to the spine, boosts target coverage by $3.7\,\%$, from $86.3\,\%$ to $90\,\%$.

Clearly, we would like to pursue only those trade-offs that are absolutely necessary to advance the objective, thus saving us a lot of time and effort. Alas, this information is not available to us beforehand. If the final step is able to revert most of the relaxations we performed during spatial dose planning, it stands to reason that the search space was too small, and that our effort was wasted. We are then compelled to utilise a beamset that is large enough to give a good approximation of what is physically achievable given the constraints.

Beam resampling can be thought of as a drop in replacement for the final step to establish Pareto efficiency of the solution. While, theoretically, the resampled solution is

not Pareto efficient, in practice attempting to re-establish Pareto efficiency leads to no real improvement in any of the goals.

7.7 CPLEX Parameter Settings

Our efforts to keep computation time as low as possible would prove futile should the CPLEX optimiser settings not be adequately set. As already shown by Ólaffson et al. [150], the optimisation runtime is highly dependent on the parameter configuration and, as we found out, it can take values within an interval spanning more than an order of magnitude.

In this section, we investigate the parameters listed in Table 7.15. The `algorithm` parameter sets the actual search algorithm to *primal simplex*, *dual simplex*, or *interior point*. We chose not to use the interior point setting, as we consistently observed worse runtimes when solving a linear program, than with any of the simplex algorithms. Forcing `predual` to `on`, the primal problem is reduced in size by a *presolve* step, then internally converted into the dual problem and passed on to the optimisation algorithm for solving. *Presolving* identifies and removes unused variables and constraints and can considerably reduce the size of the LP. Due to its low overhead and high improvement potential, we always presolve our problems. Finally, the `pricing` parameter sets the heuristic used to move to a new vertex on the LP polyhedron.

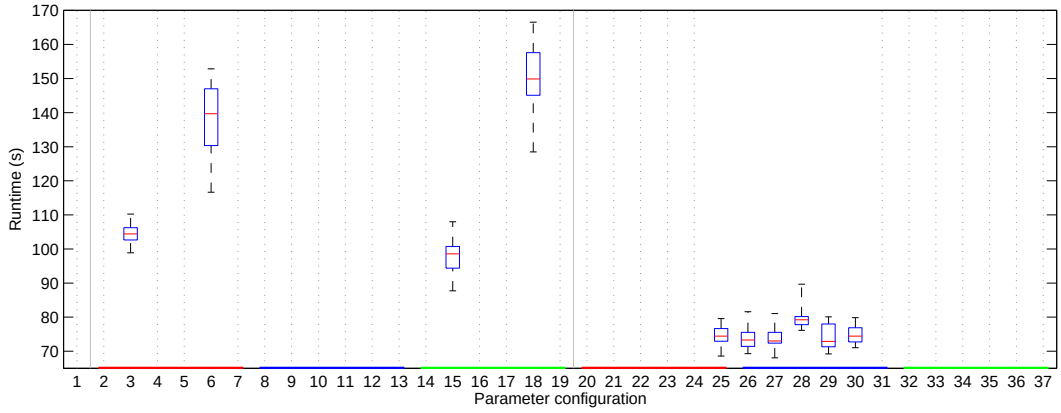
The performance of the optimiser is highly dependent on the problem that is being solved. Consequently, we would like to identify the parameter configuration that leads to the lowest runtime given a linear program. For each parameter configuration in Table 7.15 we run three optimisation problems, which are representative for our planning process: (1) an OCO step, (2) an OME step, and (3) a succession of relaxation and OME optimisation steps, simulating I3D planning. The resulting runtimes for all three scenarios are plotted in Figure 7.37. First thing we notice is the tremendous runtime variance given different parameter configurations, which is consistent with the observations made by Ólaffson et al. [150]. Improper optimiser settings result in computation times that are over an order of magnitude worse than the fastest time. Second, there exists no one parameter configuration that performs consistently well across all test scenarios, although there are configurations that performed consistently bad. Third, although Scenarios 2 and 3 share the same optimisation goal, the results show a large discrepancy between settings that we would have expected to perform equally well for both. A closer inspection reveals that the CPLEX optimiser is unable to use advanced start for some configurations. This implies that reoptimisation after constraint relaxation takes as long

Table 7.15: Parameter-value tuples controlling the execution of the CPLEX optimiser. A detailed explanation of each of the parameter can be found in [151]. The index corresponds to the parameter configuration used in the experiments.

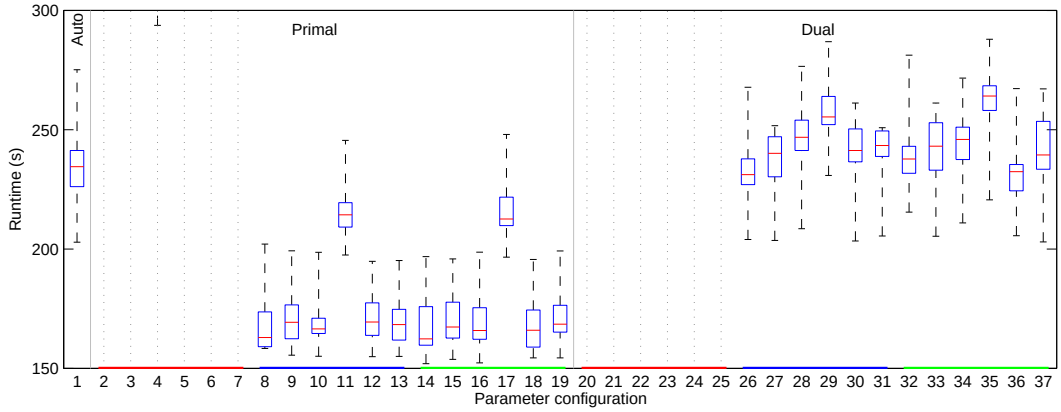
Index	Algorithm	Predual	Pricing
1	auto	auto	auto
2	primal	off	reduced cost
3			auto
4			devex
5			steepest-edge
6			initial norms
7			full
8		auto	reduced cost
9			auto
10			devex
11			steepest-edge
12			initial norms
13			full
14		on	reduced cost
15			auto
16			devex
17			steepest-edge
18			initial norms
19			full
20	dual	off	auto
21			full
22			steepest-edge
23			full steepest
24			initial norms
25			devex
26		auto	auto
27			full
28			steepest-edge
29			full steepest
30			initial norms
31			devex
32		on	auto
33			full
34			steepest-edge
35			full steepest
36			initial norms
37			devex

as a normal uninitialised optimisation step, as the optimiser does not explore the search space starting from the previous solution.

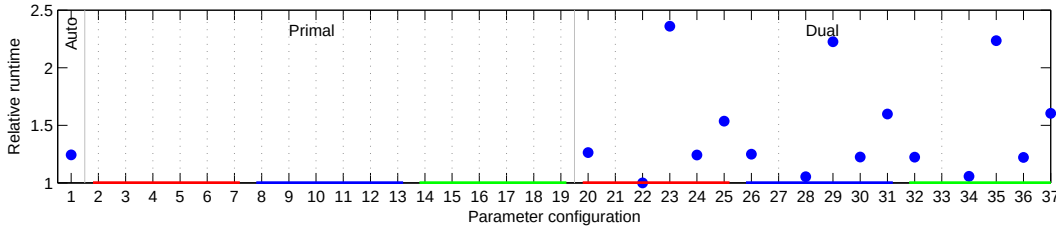
Interestingly, although the automatic setting performed a lot worse for Scenarios 1 and 2, it performed well during Scenario 3. This seems to indicate that automatic settings may be well suited for I3D planning, which we could repeatedly observe while planning other cases. Unfortunately, present results are insufficient to answer which settings, if any, can be used to guarantee consistently short runtimes. As such, we can only recommend further investigation in this direction.



(a)



(b)



(c)

Figure 7.37: Optimisation runtimes for different CPLEX parameter settings, each denoted by the indices along the x-axis (same indices as used in Table 7.15). (a, b) Boxplot of the runtimes for all ten beamsets during Scenarios 1 and 2. (c) Total optimisation runtime for Scenario 3, plotted relative to the lowest total runtime. The colouring along the x-axis signifies the `predual` setting: `off` (red), `auto` (blue) and `on` (green). The dotted vertical lines denote configurations for which runtimes are too large to fit within the y-axis limits.

8 Conclusion

We have started out with the goal of creating a tool that would give the planner direct control over the shape of the dose distribution. We intended this tool to augment the already existing planning system and, as such, we identified two important requirements: ease of use and compatibility with the established planning workflow. We tackled ease of use by employing simplified graphics, a familiar control scheme, natural virtual interactions, and by striving to keep computation time as close as possible to interactive rates. In maintaining compatibility, we introduced local dose shape as an additional planning criterion alongside conventional criteria. Furthermore, we positioned the *Interactive3D* tool as a drop-in addition to the planning system.

As we expanded on the topics of interest, we formulated a number of five hypotheses, which we supported by the experimental results from Chapter 7. In this Chapter, we tie-up the hypotheses by referring to the results, we state our contribution to the state of the art, we acknowledge the limitations of this work, and, finally, we propose several directions for future research.

8.1 Contribution to Knowledge

In considering local dose shape as additional planning criteria, we proposed an extension to the multicriteria optimisation workflow which allowed us to come up with Pareto efficient solutions and, thus, fulfilling Hypothesis 1. In Section 7.1, we constructed a simplified planning problem showing, first, how the I3D tool fits into the planning workflow and, second, how the solution found through our approach compares to the solutions obtained by two conventional methods. Displayed in the two-objective Pareto space, the I3D solution is clearly dominated by the Pareto fronts generated by the conventional methods. However, the Pareto space hides all locality information and, as such, local dose shape criteria cannot be represented. In our example, a seemingly bad sparing of the shell, obtained by Interactive3D, may actually be acceptable, as the dose bounds are only violated within the nasal cavity, which, in turn, allows for good PTV coverage. Furthermore, solutions giving rise to numerically equivalent sparing of the shell (equal objective

values), may lead to very different dose distributions. For the conventional approaches, dose fingers appeared randomly around the shell, while in the case of Interactive3D, these formed only within the area for which trade-offs were deemed as acceptable.

The same Pareto plots make obvious another point, which is mirrored throughout other planning scenarios. As the improvement per interaction step is markedly smaller for I3D than for either IDVH or WF, it is clear that our tool must be part of a larger planning system. Established tools, such as WF and IDVH, in our case, must be used to find a good approximation of the treatment plan. Only then, is the effort of finely forming the dose distribution tractable.

Second, in Sections 7.2 through 7.4 we presented a number of different planning scenarios for three real cancer cases, in which we demonstrated the use of the two I3D operations for dose shaping: *setting the optimisation goal* and *trade-off selection*. In this respect, dose shaping is actually a process of balancing local trade-offs, as we can almost never realise a local dose goal without relaxing some constraint. For example, in Section 7.2.2, in an effort to spare a sensitive portion of the rectum, we were compelled to accept a reduced dose to a portion of the PTV facing the rectum and to allow slightly more dose to the bladder. Further, in Section 7.2.3, as we attempted to create a boost volume inside one of the lobes of the prostate, we again allowed more dose to the bladder and to the shells, and still we completely avoided compromising the rectum. In another scenario involving a spine cancer case, Section 7.4.2, we were able to increase coverage of the target area (enclosed within the PTV) by sacrificing dose homogeneity only in areas contained within the GTV, and by slightly compromising the very conservative dose bounds on the proximal OARs. For the optic meningioma case of Section 7.3, we started out from two very different scenarios: one in which excellent OAR sparing meant the PTV coverage was lacking, and, the other one, in which excellent PTV coverage was achieved at the cost of subjecting the optic nerve to higher doses. In both cases we shifted the dose to achieve similar end results, wherein PTV coverage and optic nerve sparing were well balanced. Note that we cannot make any assumption on the clinical outcome of the resulting dose distributions, as our aim was to demonstrate the dose shaping, which is attainable using our technique. In describing the steps for interactive planning, we show the fulfilment of Hypotheses 2 and 3—the complete representation of the optimisation problem through visual elements and the representation of the basic optimisation operations through interaction upon these elements.

Interactivity is central to our planning method and, in an effort to improve interaction quality, we concentrated on reducing the computation complexity of the optimisation problem. In a third part of experimental results (Sections 7.5 and 7.6), we evaluated

our methods for improving sampling of the anatomical and beam spaces. As expected, preferentially oversampling portions of the VOIs that are likely to be exposed to high dose gradients (e.g. VOI surfaces facing other VOIs), allows us to preserve most of the solution quality (Hypothesis 4), for which we would have required a much finer grid, had we used homogeneous sampling. In most cases, the computational cost of using such a fine resolution would have been prohibitive. Furthermore, for I3D planning, there is a need to also supersample the areas on interaction in advance. Due to the fact that these areas are typically located in portions of volumes that are likely to experience high dose gradients, most of these areas will already have been oversampled. On the other hand, subsampling other regions (e.g. VOI interiors) does introduce artefacts. This is why, in Section 6.1.2, we compensated aliasing errors inside the PTV, by tightening the voxel bounds according to a model that predicts the reduction in hotspot magnitude.

Optimising with a limited beam set introduces another artefact of its own: the solution is often not as good as it could be, if a larger beam set would be used. In fact, even if computation time would not be an issue, the solution obtained with the maximum amount of beams we can fit in the internal memory can generally be improved upon. Consequently, Section 6.2 introduced one last, automated step to planning, in which we improve each objective and we recover as much as we can from all relaxations, through beam resampling. Here, we also attempted to accelerate resampling, by taking advantage of the multicore architecture of most processors today. Unfortunately, we could achieve no such speedup, as peer threads compete for memory access on the shared memory bus. However, we did see an important objective improvement per iteration for multiple running threads. It stands to reason then that, as also stated in Hypothesis 5, the *Concurrent Beam Resampling* algorithm could take advantage of the computing architecture of a multiprocessor system—where each computing unit possesses its own private working memory—to deliver real world acceleration of the resampling process.

Despite our best efforts to accelerate computation, we are ultimately limited by the performance of the CPLEX optimiser. As noted in [150], optimisation runtime is highly dependent on the parameter settings. In the final part of our experiments, we varied the CPLEX parameter values, in search of setting configurations that would perform consistently well across different scenarios. Our attempt resulted in negative results: we could identify no such configuration, nor could we identify which configurations could be used optimally given a certain optimisation goal. This result does not mean that such a parameter set does not exist, as it may very well do, but if it does, it definitely not trivial to find. During interactive planning, our experiments suggest that the automatic settings may provide suitable performance.

8.2 Theoretical Implications and Limitations of Interactive3D

Interactive three-dimensional planning opens up new possibilities, as it allows the user to consider the locality of the goals and of the trade-offs. This means that there is now an even larger number of optimisation criteria to take into account. Thus, this new link added to the planning toolchain further increases the complexity of an already highly sophisticated process [152]. The user still has to use established methods to quickly identify a solid initial solution. Only then, does it make sense to tweak the plan using Interactive3D. Typically, at this point, a lot of constraints are rigid and most goals are suitably satisfied. At the same time, areas taken into consideration with our tool are relatively small. Consequently, the implied improvements are as well limited when compared to those obtained by earlier steps.

When shifting dose around, one must remember that "there is no such thing as a free lunch" [153]. To achieve a dose goal, the user must typically agree to a number of localised trade-offs. While we attempt to make evident those trade-offs that may result in an improvement, what we actually do, is to hide those trade-offs that are not worthwhile considering from the current solution. Ultimately, the user has to decide which relaxations to perform. This decision is influenced by a number of factors, such as risk assessment and opportunity of improving the goal. First, risk assessment is a particularly difficult issue and ongoing research is trying to predict complications as a result of therapy [8, 128, 129]. Currently, metastudies, such as the Emami paper [154] and the QUANTEC review [7, 144], guide clinicians in assessing the probability of complications. However, we are a long way away from predicting the clinical effects of local dose modifications. Second, the planner has to actually perform a trade-off to assess its real potential for advancing the goal and, because constraints are coupled, previously non-limiting voxels may become limiting following relaxation in other areas (which we could see during the relaxations steps in Section 7.4.2). Consequently, despite our guideline for systematically exploring the available trade-offs, interactive three-dimensional planning remains a time consuming trial and error process.

8.3 Recommendation for Future Research

The I3D tool fulfils our expectations for optimised dose shape manipulation, yet there are a number of interesting problems we did not address. First, users would profit from improved scoring of potential relaxations, in a manner similar to sensitivity analysis. A lot of planning time is currently lost on relaxing bounds that, although appear limiting,

do not advance the objective. One idea would be to concurrently solve optimisation problems, for which the top most limiting clustered voxels would be relaxed. Each solution would give an idea of how much the objective is dependent on the respective cluster. Multiple threads would start from the current master solution and run in the background, while the user is navigating the virtual scene. Another approach would be to vary the beam weights in a manner that is more similar to direct planning, as proposed by Kamerling et al. [106] or by Otto [107]. Although not optimal, the resulting dose distribution would satisfy the main goal, and one could analyse which constraints had to be violated and to what degree. Note that a lot of care should be taken, as there may exist a solution that is optimal, and for which totally different relaxations may be needed.

Second, optimisation runtime is the major hurdle to real-time optimal deformation of the dose distribution. Any acceleration of the linear optimiser would certainly improve planning, either because less time would be required, or because the planner would be incentivized to try out a more diverse set of trade-offs.

Third, while mouse and keyboard interaction with a three-dimensional scene projected on a two-dimensional display is simple and familiar, it is certainly not optimal. One possible enhancement would be to use a virtual reality headset (e.g. Oculus Rift [155, 156]), together with a natural input modality (such as body, hand or tool tracking, 6-D mice, etc.). Such a visualisation and control scheme could allow superior insight into the optimisation problem and better orientation within the rendered scene. On the other hand, simulator sickness is a still an issue for a number of users [157] and most VR systems are still quite complicated. Hence, the benefits may not be worth the trouble.

Fourth, we have speculated in Section 7.6 that concurrent beam resampling would provide a speedup over classical beam resampling when implemented on a multiprocessor system. Future experiments should provide a definitive test of our hypothesis.

Finally, for beam resampling, we randomly generate replacement beams, in an effort to identify a superior solution. It may prove advantageous to use a heuristic to score each beam, and replace only those beams that are least useful. Such heuristics could be similar to those used in *iterative beam planning* [10]. In fact, combining this with Concurrent Beam Resampling, multiple iterative beam planing algorithms could be solved in parallel threads with solutions being proposed to the peer threads. Future studies would then have to look into whether beams provided by peer threads are more useful than beams that are heuristically higher rated.

Bibliography

[1] Sherry L. Murphy, Jiaquan Xu, and Kenneth D. Kochanek. Deaths: Final data for 2010. Technical Report 4, Volume 61, U.S. Department of Health and Human Services, 2013.

[2] A. S. Ahmad, N. Ormiston-Smith, and P. D. Sasieni. Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960. *British Journal of Cancer*, feb 2015.

[3] Cancer survival for common cancers, 2015. Cancer survival for common cancers.

[4] Rajamanickam Baskar. Cancer and radiation therapy: Current advances and future directions. *International Journal of Medical Sciences*, 9(3):193–199, 2012.

[5] Eric J. Hall and Amato J. Giaccia. Dose-response relationships for model normal tissues. In *Radiobiology for the Radiologist*. Lippincott Williams & Wilkins, 2006.

[6] Wolfgang C. Schlegel, Thomas Bortfeld, and Anca Ligia Grosu. *New Technologies in Radiation Oncology*. Springer, 2006.

[7] Lawrence B. Marks, Ellen D. Yorke, Andrew Jackson, Randall K. Ten Haken, Louis S. Constine, Avraham Eisbruch, Søren M. Bentzen, Jiho Nam, and Joseph O. Deasy. Use of normal tissue complication probability models in the clinic. *International Journal of Radiation Oncology, Biology, Physics*, 76(3):S10–S19, 2010.

[8] Mike Partridge and Steve Webb. Modeling late rectal toxicities based on a parameterized representation of the 3D dose distribution. https://publications.icr.ac.uk/10527/, April 2011.

[9] Thomas Dirscherl, Mark Rickhey, and Ludwig Bogner. Feasibility of TCP-based dose painting by numbers applied to a prostate case with 18f-choline PET imaging. *Zeitschrift für Medizinische Physik*, 22(1):48–57, 2012.

[10] Alexander Schlaefer. *Computer assisted planning for robotic radiosurgery*. PhD thesis, Universität zu Lübeck, Berlin, 2007.

[11] T. Bortfeld and W. Schlegel. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Physics in Medicine and Biology*, 38(2):291–304, 2011.

[12] David Craft. Multi-criteria optimization methods in radiation therapy planning: a review of technologies and directions, 2013.

[13] David Craft, Tarek Halabi, and Thomas Bortfeld. Exploration of tradeoffs in intensity-modulated radiotherapy. *Physics in Medicine and Biology*, 50(24):5857–5868, December 2005. PMID: 16333160.

[14] Jan J Wilkens, James R Alaly, Konstantin Zakarian, Wade L Thorstad, and Joseph O Deasy. IMRT treatment planning based on prioritizing prescription goals. *Physics in medicine and biology*, 52(6):1675–1692, March 2007. PMID: 17327656.

[15] K.-H. Küfer, M. Monz, A. Scherrer, P. Süss, F. Alonso, A.S.A. Sultan, Th. Bortfeld, D. Craft, and Chr. Thieke. Multicriteria optimization in intensity modulated radiotherapy planning. Technical Report 77, Fraunhofer (ITWM), 2005.

[16] A. Scherrer, T. Grebe, F. Yaneva, K.-H. Küfer. Interactive DVH-based planning of intensity modulated radiation therapy (IMRT). *Berichte des Fraunhofer ITWM*, 222, 2012.

[17] Cristian Cotrutz and Lei Xing. IMRT dose shaping with regionally variable penalty scheme. *Medical*

physics, 30(4):544–551, April 2003. PMID: 12722806.

[18] K. S. Clifford Chao, Angel I. Blanco, and James F. Dempsey. A conceptual model integrating spatial information to assess target volume coverage for IMRT treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, 56(5):1438–1449, 2003.

[19] Alan E. Nahum and Julien Uzan. (Radio)Biological optimization of external-beam radiotherapy. *Computational and Mathematical Methods in Medicine*, 2012:e329214, 2012.

[20] Cristian Cotrutz and Lei Xing. Using voxel-dependent importance factors for interactive DVH-based dose optimization. *Physics in medicine and biology*, 47(10):1659–1669, May 2002. PMID: 12069085.

[21] Yong Yang and Lei Xing. Inverse treatment planning with adaptively evolving voxel-dependent penalty scheme. *Medical Physics*, 31(10):2839–2844, 2004.

[22] Pavel Lougovski, Jordan LeNoach, Lei Zhu, Yunzhi Ma, Yair Censor, and Lei Xing. Toward truly optimal IMRT dose distribution: Inverse planning with voxel-specific penalty. *Technology in cancer research & treatment*, 9(6):629–636, December 2010. PMID: 21070085 PMCID: PMC3057528.

[23] Chuan Wu, Gustavo H Olivera, Robert Jeraj, Harry Keller, and Thomas R Mackie. Treatment plan modification using voxel-based weighting factors/dose prescription. *Physics in medicine and biology*, 48(15):2479–2491, August 2003. PMID: 12953910.

[24] Harold Elford Johns and John Robert Cunningham. *Physics of Radiology, Fourth Edition*. Charles C Thomas Pub Ltd, Springfield, Ill., U.S.A, 4th edition, February 1983.

[25] D J Brenner and E J Hall. Fractionation and protraction for radiotherapy of prostate carcinoma. *International Journal of Radiation Oncology, Biology, Physics*, 43(5):1095–1101, 1999.

[26] Martin Jermann. Particle therapy patient statistics. Statistics, Particle Therapy Co-Operative Group, 2014.

[27] Teruhiko Terasawa, Tomas Dvorak, Stanley Ip, Gowri Raman, Joseph Lau, and Thomas A. Trikalinos. Systematic review: Charged-particle radiation therapy for cancer. *Annals of Internal Medicine*, 151(8):556–565, 2009.

[28] Beam guidance to the three threapy rooms. Universtätsklinikum Heidelberg, Ion-Beam Therapy Center, 2015. Available at http://www.klinikum.uni-heidelberg.de/HIT-Bildergalerie.114799.0.html.

[29] Gary A. Ezzell, James M. Galvin, Daniel Low, Jatinder R. Palta, Isaac Rosen, Michael B. Sharpe, Ping Xia, Ying Xiao, Lei Xing, Cedric X. Yu, IMRT subcommitte, and AAPM Radiation Therapy committee. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT subcommittee of the AAPM radiation therapy committee. *Medical Physics*, 30(8):2089–2115, 2003.

[30] Christian Fiandra, Andrea R. Filippi, Paola Catuzzo, Angela Botticella, Patrizia Ciammella, Pierfrancesco Franco, Valeria C. Borca, Riccardo Ragona, Santi Tofani, and Umberto Ricardi. Different IMRT solutions vs. 3d-conformal radiotherapy in early stage hodgkin's lymphoma: dosimetric comparison and clinical considerations. *Radiation Oncology*, 7(1):186, November 2012.

[31] Chi Lin, Sarah S. Donaldson, Jane L. Meza, James R. Anderson, Elizabeth R. Lyden, Christopher K. Brown, Karen Morano, Fran Laurie, Carola A. Arndt, Charles A. Enke, and John C. Breneman. Effect of radiotherapy techniques (IMRT vs. 3D-CRT) on outcome in patients with intermediate-risk rhabdomyosarcoma enrolled in COG d9803 - a report from the children's oncology group. *International Journal of Radiation Oncology Biology Physics*, 82(5):1764–1770, April 2012.

[32] Erjona Bakiu, Ervis Telhaj, Elvisa Kozma, Ferdinand Ruşi, and Partizan Malkaj. Comparison of 3D CRT and IMRT tratment plans. *Acta Informatica Medica*, 21(3):211–212, 2013.

[33] Ryan Foster, Jeffrey Meyer, Puneeth Iyengar, David Pistenmaa, Robert Timmerman, Hak Choy, and Timothy Solberg. Localization accuracy and immobilization effectiveness of a stereotactic body frame for a variety of treatment sites. *International Journal of Radiation Oncology, Biology, Physics*, 87(5):911–

916, 2013.

[34] John R. Adler, Jr., Steven D. Chang, Martin J. Murphy, James Doty, Paul Geis, and Stephen L. Hancock. The Cyberknife: A frameless robotic system for radiosurgery. *Stereotactic and Functional Neurosurgery*, 69(1):124–128, 1997.

[35] Wenyu Cheng and John R. Adler. An overview of cyberknife radiosurgery. *Chinese Journal of Clinical Oncology*, 3(4):229–243, 2006.

[36] W. Kilby, J. R. Dooley, G. Kuduvalli, S. Sayeh, and C. R. Maurer. The CyberKnife robotic radiosurgery system in 2010. *Technology in Cancer Research & Treatment*, 9(5):433–452, 2010.

[37] H. W. Hamacher and K. H. Küfer. Inverse radiation therapy planning – a multiple objective optimization approach. *Discrete Applied Mathematics*, 118(1):145–161, 2002.

[38] Werner Bär, Marco Schwarz, Markus Alber, Luc J. Bos, Ben J. Mijnheer, Coen Rasch, Christoph Schneider, Fridtjof Nüsslin, and Eugene M. F. Damen. A comparison of forward and inverse treatment planning for intensity-modulated radiotherapy of head and neck cancer. *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology*, 69(3):251–258, 12 2003.

[39] Achim Schweikard, Alexander Schlaefer, and John R Adler. Resampling: an optimization method for inverse planning in robotic radiosurgery. *Medical physics*, 33(11):4005–4011, November 2006. PMID: 17153380.

[40] O. Belohlavek, I. Carrio, M. Danna, E. Deniaud-Alexandre, T. Inoue, M. Macmanus, C. Messa, U. Nestle, K. Rosenzweig, and S. Schipani. The role of PET/CT in radiation treatment planning for cancer patient treatment. Technical Report IAEA-TECDOC-1603, International Atomic Energy Agency, 10 2008. ISBN 978-92-0-110408-3.

[41] P Metcalfe, G P Liney, L Holloway, A Walker, M Barton, G P Delaney, S Vinod, and W Tome. The potential for an enhanced role for MRI in radiation-therapy treatment planning. *Technology in cancer research & treatment*, 12(5):429–446, October 2013. PMID: 23617289.

[42] W C Liu, M Schulder, V Narra, A J Kalnin, C Cathcart, A Jacobs, G Lange, and A I Holodny. Functional magnetic resonance imaging aided radiation treatment planning. *Medical physics*, 27(7):1563–1572, July 2000. PMID: 10947259.

[43] Jenghwa Chang, Alex Kowalski, Bob Hou, and Ashwatha Narayana. Feasibility study of intensity-modulated radiotherapy (IMRT) treatment planning using brain functional MRI. *Medical dosimetry: official journal of the American Association of Medical Dosimetrists*, 33(1):42–47, 2008. PMID: 18262122.

[44] Peroni Marta, Patete Paolo, Ghielmetti F.,Casolino D., Ongania E., Casolino D.S., Fariselli L., Baroni Guido. VMAT inverse planning including DTI tractography fiber bundles as organs at risk: a feasibility study. In *MICCAI 2012 Workshop*, 2012.

[45] C C Ling, J Humm, S Larson, H Amols, Z Fuks, S Leibel, and J A Koutcher. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *International journal of radiation oncology, biology, physics*, 47(3), June 2000. PMID: 10837935.

[46] Søren M Bentzen and Vincent Gregoire. Molecular imaging-based dose painting: a novel paradigm for radiation therapy prescription. *Seminars in radiation oncology*, 21(2):101–110, April 2011. PMID: 21356478.

[47] Joe H. Chang, Daryl Lim Joon, Sze Ting Lee, Sylvia J. Gong, Nigel J. Anderson, Andrew M. Scott, Ian D. Davis, David Clouston, Damien Bolton, Christopher S. Hamilton, and Vincent Khoo. Intensity modulated radiation therapy dose painting for localized prostate cancer using 11C-choline positron emission tomography scans. *International Journal of Radiation Oncology, Biology, Physics*, 83(5):e691–e696, August 2012.

[48] Norbert Hodapp. ICRU report 83, prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). Technical Report 83, International Commission on Radiation

Units and Measurements (ICRU), 2010.

[49] Anders Ahnesjö and Maria Mania Aspradakis. Dose calculations for external photon beams in radiotherapy. *Physics in Medicine and Biology*, 44(11):R99–R155, November 1999.

[50] Andrés Gómez, Carlos Fernández Sánchez, José Carlos Mouriño Gallego, Javier López Cacheiro, Francisco J. González Castaño, Daniel Rodríguez-Silva, Lorena Domínguez Carrera, David González Martínez, Javier Pena García, Faustino Gómez Rodríguez, Diego González Castaño, and Miguel Pombar Cameán. Monte carlo verification of IMRT treatment plans on grid. *Studies in Health Technology and Informatics*, 126:105–114, 2007.

[51] Accuray. Monte Carlo dose calculation algorithm for the CyberKnife robotic radiosurgery system. Whitepaper, Accuray Incorporated, 2010.

[52] Accuray. Spine treatment plan image, 2015. Available at http://www.accuray.com/pressroom/virtual-press-kit/images-logos.

[53] Mattias Ehrgott and Mena Burjony. Radiation therapy planning by multicriteria optimisation. In *Proceedings of the 36th Annual Conference of the Operational Research Society of New Zealand*, 2001.

[54] P. Mayles, A. Nahum, and J. C. Rosenwald, editors. *Handbook of Radiotherapy Physics: Theory and Practice*. CRC Press, 1st edition, 2007.

[55] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[56] H. Edwin Romeijn, James F. Dempsey, and Jonathan G. Li. A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, 49(10):1991, May 2004.

[57] Alexander V. Lotov and Kaisa Miettinen. Visualizing the pareto frontier. In Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Roman Slowiński, editors, *Multiobjective Optimization*, number 5252 in Lecture Notes in Computer Science, pages 213–243. Springer Berlin Heidelberg, 2008.

[58] A Schlaefer and A Schweikard. Stepwise multi-criteria optimization for robotic radiosurgery. *Medical Physics*, 35(5):2094–2103, May 2008. PMID: 18561685.

[59] Aswin L. Hoffmann, Dick den Hertog, Alex Y. D. Siem, Johannes H. A. M. Kaanders, and Henk Huizenga. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. *Physics in Medicine and Biology*, 53(22):6345, November 2008.

[60] Karl-Heinz Küfer, HorstW. Hamacher, and Thomas Bortfeld. A multicriteria optimization approach for inverse radiotherapy planning. In Wolfgang Schlegel and Thomas Bortfeld, editors, *The Use of Computers in Radiation Therapy*, pages 26–28. Springer Berlin Heidelberg, 2000.

[61] C Cotrutz, M Lahanas, C Kappas, and D Baltas. A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. *Physics in medicine and biology*, 46(8):2161–2175, August 2001. PMID: 11512617.

[62] Michael Lahanas, Eduard Schreibmann, and Dimos Baltas. Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Physics in medicine and biology*, 48(17):2843–2871, September 2003. PMID: 14516105.

[63] David L. Craft, Tarek F. Halabi, Helen A. Shih, and Thomas R. Bortfeld. Approximating convex pareto surfaces in multiobjective radiotherapy planning. *Medical Physics*, 33(9):3399–3407, 2006.

[64] Christian Thieke, Karl-Heinz Küfer, Michael Monz, Alexander Scherrer, Fernando Alonso, Uwe Oelfke, Peter E Huber, Jürgen Debus, and Thomas Bortfeld. A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. *Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology*, 85(2):292–298, November 2007. PMID: 17892901.

[65] M Monz, K H Küfer, T R Bortfeld, and C Thieke. Pareto navigation: algorithmic foundation of interactive multi-criteria IMRT planning. *Physics in medicine and biology*, 53(4):985–998, February 2008. PMID:

18263953.

[66] Matthias Ehrgott and Ines Winz. Interactive decision support in radiation therapy treatment planning. *OR Spectrum*, 30(2):311–329, apr 2008.

[67] David Craft and Michael Monz. Simultaneous navigation of multiple pareto surfaces, with an application to multicriteria IMRT planning with multiple beam angle configurations. *Medical physics*, 37(2):736–741, February 2010. PMID: 20229883.

[68] Kyung-Wook Jee, Daniel L McShan, and Benedick A Fraass. Lexicographic ordering: intuitive multicriteria optimization for IMRT. *Physics in medicine and biology*, 52(7):1845–1861, April 2007. PMID: 17374915.

[69] T Long, M Matuszak, M Feng, B A Fraass, R K Ten Haken, and H E Romeijn. Sensitivity analysis for lexicographic ordering in radiation therapy treatment planning. *Medical physics*, 39(6):3445–3455, June 2012. PMID: 22755724.

[70] Massimiliano Caramia and Paolo Dell'Olmo. *Multi-objective management in freight logistics increasing capacity, service level and safety with optimization algorithms*. Springer, London, 2008.

[71] I. Das and J. E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural optimization*, 14(1):63–69, August 1997.

[72] R. T. Marler and J. S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, April 2004.

[73] Rasmus Bokrantz. Distributed approximation of pareto surfaces in multicriteria radiation therapy treatment planning. *Physics in medicine and biology*, 58(11):3501–3516, June 2013.

[74] Aswin L. Hoffmann, Alex Y. D. Siem, Dick den Hertog, Johannes H. A. M. Kaanders, and Henk Huizenga. Derivative-free generation and interpolation of convex pareto optimal IMRT plans. *Physics in Medicine and Biology*, 51(24):6349, December 2006.

[75] J I Serna, M Monz, K H Küfer, and C Thieke. Trade-off bounds for the pareto surface approximation in multi-criteria IMRT planning. *Physics in medicine and biology*, 54(20):6299–6311, October 2009. PMID: 19809122.

[76] David Craft and Thomas Bortfeld. How many plans are needed in an IMRT multi-objective plan database? *Physics in Medicine and Biology*, 53(11):2785, June 2008.

[77] Alexander Schlaefer, Oliver Blanck, and Achim Schweikard. Interactive multi-criteria inverse planning for robotic radiosurgery. In *XVth International Conference on the Use of Computers in Radiation Therapy (ICCR 2007)*, June 2007.

[78] A Niemierko and M Goitein. The influence of the size of the grid used for dose calculation on the accuracy of dose estimation. *Medical physics*, 16(2):239–247, April 1989. PMID: 2541326.

[79] James F. Dempsey, H. Edwin Romeijn, Jonathan G. Li, Daniel A. Low, and Jatinder R. Palta. A fourier analysis of the dose grid resolution required for accurate IMRT fluence map optimization. *Medical Physics*, 32:380, 2005.

[80] H. Li and J. Dempsey. A fourier analysis of the dose grid resolution requirement in proton therapy IMRT fluence-map optimization. *Medical Physics*, 32(6):2027–2027, 2005.

[81] Heeteak Chung, Hosang Jin, Jatinder Palta, Tae-Suk Suh, and Siyong Kim. Dose variations with varying calculation grid size in head and neck IMRT. *Physics in Medicine and Biology*, 51(19):4841–4856, October 2006. PMID: 16985274.

[82] Humberto Rocha, Joana M. Dias, Brigida C. Ferreira, and Maria do Carmo Lopes. Influence of sampling in radiation therapy treatment design. In Beniamino Murgante, Osvaldo Gervasi, Andrés Iglesias, David Taniar, and Bernady O. Apduhan, editors, *Computational Science and Its Applications - ICCSA 2011*, number 6784 in Lecture Notes in Computer Science, pages 215–230. Springer Berlin Heidelberg,

January 2011.

[83] Michael C. Ferris, Rikhardur Einarsson, Ziping Jiang, and David Shepard. Sampling issues for optimization in radiotherapy. *Annals of Operations Research*, 148(1):95–115, November 2006.

[84] A Niemierko and M Goitein. The use of variable grid spacing to accelerate dose calculations. *Medical physics*, 16(3):357–366, June 1989. PMID: 2544792.

[85] A Niemierko and M Goitein. Random sampling for evaluating treatment plans. *Medical physics*, 17(5):753–762, October 1990. PMID: 2233561.

[86] Alexander Scherrer, Karl-Heinz Küfer, Thomas Bortfeld, Michael Monz, and Fernando Alonso. IMRT planning on adaptive volume structures – a decisive reduction in computational complexity. *Physics in Medicine and Biology*, 50:2033–2053, May 2005.

[87] Benjamin C Martin, Thomas R Bortfeld, and David A Castanon. Accelerating IMRT optimization by voxel sampling. *Physics in Medicine and Biology*, 52(24):7211–7228, December 2007. PMID: 18065835.

[88] John D. Fenwick and Juan Pardo-Montero. Numbers of beam angles required for near-optimal IMRT: theoretical limits and numerical studies. *Medical Physics*, 38(8):4518–4530, 2011.

[89] George Starkschall and R. Alfredo C. Siochi. *Informatics in Radiation Oncology*. CRC Press, 2013.

[90] Charles D. Hansen and Chris R. Johnson. *Visualization Handbook*. Academic Press, Amsterdam ; Boston, 1st edition, December 2004.

[91] Bernhard Preim and Dirk Bartz. *Visualization in Medicine: Theory, Algorithms, and Applications*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.

[92] Charles M. Washington and Dennis T. Leaver. *Principles and Practice of Radiation Therapy*. Elsevier Health Sciences, 2015.

[93] Benjamin Fahimian, Andrew Ceballos, Silvan Türkcan, Daniel S. Kapp, and Guillem Pratx. Seeing the invisible: direct visualization of therapeutic radiation beams using air scintillation. *Medical Physics*, 41(1):010702, 2014.

[94] Lesley A. Jarvis, Rongxiao Zhang, David J. Gladstone, Shudong Jiang, Whitney Hitchcock, Oscar D. Friedman, Adam K. Glaser, Michael Jermyn, and Brian W. Pogue. Cherenkov video imaging allows for the first visualization of radiation therapy in real time. *International Journal of Radiation Oncology, Biology, Physics*, 89(3):615–622, 2014.

[95] J. Remy, M. Remy-Jardin, D. Artaud, and M. Fribourg. Multiplanar and three-dimensional reconstruction techniques in CT: impact on chest diseases. *European Radiology*, 8(3):335–351, 1998.

[96] Xiao Hui Wang, Janet E. Durick, Amy Lu, David L. Herbert, Saraswathi K. Golla, Kristin Foley, C. Samia Piracha, Dilip D. Shinde, Betty E. Shindel, Carl R. Fuhrman, Cynthia A. Britton, Diane C. Strollo, Sherry S. Shang, Joan M. Lacomis, and Walter F. Good. Characterization of radiologists' search strategies for lung nodule detection: Slice-based versus volumetric displays. *Journal of Digital Imaging*, 21(Suppl 1):39–49, October 2008.

[97] Klaus Engel. *Real-time volume graphics*. A K Peters, Wellesley, Mass., 2006.

[98] Hubert Nguyen. *GPU Gems 3*. Addison-Wesley Professional, 2007.

[99] Gordon Kindlmann. Transfer functions in direct volume rendering: Design, interface, interaction. *Course notes of ACM SIGGRAPH*, 2002.

[100] J. Kniss, G. Kindlmann, and C. Hansen. Multidimensional transfer functions for interactive volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):270–285, 2002.

[101] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, August 1987.

[102] Will Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit, Third Edition*. Kitware Inc., 3rd

edition, August 2004.

[103] Sebastiaan Breedveld, Pascal R. M. Storchi, Marleen Keijzer, Arnold W. Heemink, and Ben J. M. Heijmen. A novel approach to multi-criteria inverse planning for IMRT. *Physics in Medicine and Biology*, 52(20):6339, October 2007.

[104] Laura LeGault and Michael C. Ferris. An interactive GUI tool for radiotherapy treatment planning, June 2009.

[105] Tiberiu Viuleț, Norman Rzezovski, and Alexander Schlaefer. Towards interactive planning for radiotherapy by three-dimensional iso-dose manipulation. In *International Journal of Computer Assited Radiology and Surgery*, volume 6, page S57, Berlin, Germany, June 2011. Springer.

[106] C. Kamerling, P. Ziegenhein, H. Heinrich, and U. Oelfke. Isodose curve manipulation for interactive dose shaping. *Medical Physics*, 39(6):3979–3980, June 2012.

[107] Karl Otto. Real-time interactive treatment planning. *Physics in Medicine and Biology*, 59(17):4845, 2014.

[108] M. Christopher, H. Roger, and H. David. 3D sculpting and auto-stereoscopic display of ICRU-50 volumes and anatomy in conformal radiotherapy treatment planning. In *Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meetring of the Biomedical Engineering Society*, volume 2, pages 1165 vol.2–, 1999.

[109] Ida Olofsson, Karljohan Lundin, Matthew Cooper, Per Kjall, and Anders Ynnerman. A haptic interface for dose planning in stereo-tactic radio-surgery. In *Proceedings of the Information Visualisation, Eighth International Conference*, pages 200–205, Washington, DC, USA, 2004. IEEE Computer Society.

[110] Alexander Schlaefer, Hiroya Shiomi, Oliver Blanck, and Achim Schweikard. Radiosurgery: identification of efficient treatment beams guided by autostereoscopic visualization. *GMS CURAC*, 1, October 2006.

[111] Jason Geng. Volumetric 3D display for radiation therapy planning. *Journal of Display Technology*, 4:437–450, December 2008.

[112] R.A Ruddle. The effect of environment characteristics and user interaction on levels of virtual environment sickness. In *IEEE Virtual Reality, 2004. Proceedings*, pages 141–285, March 2004.

[113] Accuray. Iris variable aperture collimator for the CyberKnife robotic radiosurgery system: Design, beam properties and clinical benefits. Whitepaper, Accuray Incorporated, 2008. Available at http://www.cyberknife.com.tr/images/yayin/Iris_Whitepaper.pdf.

[114] Daryl P. Nazareth, Stephen Brunner, Matthew D. Jones, Harish K. Malhotra, and Mohammad Bakhtiari. Optimization of beam angles for intensity modulated radiation therapy treatment planning using genetic algorithm on a distributed computing platform. *Journal of Medical Physics / Association of Medical Physicists of India*, 34(3):129–132, 2009.

[115] Tarek Halabi. *Multi-criteria Optimization with Dose-Volume Criteria*. PhD disertation, University of Massachusetts Amherst, 2007.

[116] Faiz M Khan. *The physics of radiation therapy*. Lippincott Williams & Wilkins, Philadelphia, 2010.

[117] Maciej Pech, Jens Ricke, Max Seidensticker, Grzegorz Staskiewicz, Gero Wieners, Konrad Mohnike, Ricarda Ruhl, Johannes Steinberg, Peter Wust, and Peter Seidensticker. Assessment of the tolerance dose of the hepatic reticulo-endothelial system (RES) after single fraction HDR-irradiation: an in-vivo study employing SSPIO. *International journal of radiation biology*, 84(10):830–837, October 2008. PMID: 18979318.

[118] Jonathan David Pfautz. *Depth Perception in Computer Graphics*. PhD disertation, University of Cambridge, May 2000.

[119] Steve Eddins. Rainbow Color Map Critiques: An Overview and Annotated Bibliography, April 2015.

[120] Mathias Hummel, Christoph Garth, Bernd Hamann, Hans Hagen, and Kenneth I. Joy. IRIS: illustrative

rendering for integral surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1319–1328, November 2010.

[121] Cass Everitt. Interactive order-independent transparency. *White paper, nVIDIA*, 2(6):7, 2001.

[122] Louis Bavoil and Kevin Myers. Order independent transparency with dual depth peeling. *NVIDIA OpenGL SDK*, 2008.

[123] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, 2004.

[124] Tiberiu Viuleţ and Alexander Schlaefer. *Error compensated sparse optimization for fast radiosurgery treatment planning*, volume 39. Medical Physics, August 2012.

[125] Tiberiu Viuleţ, Oliver Blanck, and Alexander Schlaefer. Parallel optimization of beam configurations for CyberKnife treatments. *Medical Physics*, 41(6):283–283, 2014.

[126] Leonard N. Chen, Simeng Suy, Sunghae Uhm, Eric K. Oermann, Andrew W. Ju, Viola Chen, Heather N. Hanscom, Sarah Laing, Joy S. Kim, Siyuan Lei, Gerald P. Batipps, Keith Kowalczyk, Gaurav Bandi, John Pahira, Kevin G. McGeagh, Brian T. Collins, Pranay Krishnan, Nancy A. Dawson, Kathryn L. Taylor, Anatoly Dritschilo, John H. Lynch, and Sean P. Collins. Stereotactic body radiation therapy (SBRT) for clinically localized prostate cancer: the Georgetown University experience. *Radiation Oncology*, 8(1):58, March 2013. PMID: 23497695.

[127] Alan J. Katz, Michael Santoro, Fred Diblasio, and Richard Ashley. Stereotactic body radiotherapy for localized prostate cancer: disease control and quality of life at 6 years. *Radiation Oncology*, 8(1):118, May 2013. PMID: 23668632.

[128] Claudio Fiorino, Cesare Cozzarini, Vittorio Vavassori, Giuseppe Sanguineti, Carla Bianchi, Giovanni Mauro Cattaneo, Franca Foppiano, Alessandro Magli, and Anna Piazzolla. Relationships between DVHs and late rectal bleeding after radiotherapy for prostate cancer: analysis of a large group of patients pooled from three institutions. *Radiotherapy and Oncology*, 64(1):1–12, July 2002.

[129] Florian Buettner, Sarah L Gulliford, Steve Webb, Matthew R Sydes, David P Dearnaley, and Mike Partridge. Assessing correlations between the spatial distribution of the dose to the rectal wall and late rectal toxicity after prostate radiotherapy: an analysis of data from the MRC RT01 trial (ISRCTN 47772397). *Physics in Medicine and Biology*, 54(21):6535–6548, November 2009.

[130] Suki Gill, Jessica Thomas, Chris Fox, Tomas Kron, Aldo Rolfo, Mary Leahy, Sarat Chander, Scott Williams, Keen Hun Tai, Gillian M. Duchesne, and Farshad Foroudi. Acute toxicity in prostate cancer patients treated with and without image-guided radiotherapy. *Radiation Oncology*, 6(1):145, October 2011. PMID: 22035354.

[131] Debra E Freeman and Christopher R King. Stereotactic body radiotherapy for low-risk prostate cancer: five-year outcomes. *Radiation Oncology*, 6(1):3, January 2011.

[132] Alan M Nieder, Michael P Porter, and Mark S Soloway. Radiation therapy for prostate cancer increases subsequent risk of bladder and rectal cancer: a population based cohort study. *The Journal of urology*, 180(5):2005–2009; discussion 2009–2010, November 2008. PMID: 18801517.

[133] Sarah L Gulliford, Kerwyn Foo, Rachel C Morgan, Edwin G Aird, A Margaret Bidmead, Helen Critchley, Philip M Evans, Stefano Gianolini, W Philip Mayles, A Rollo Moore, Beatriz Sánchez-Nieto, Mike Partridge, Matthew R Sydes, Steve Webb, and David P Dearnaley. Dose-volume constraints to reduce rectal side effects from prostate radiotherapy: evidence from MRC RT01 trial ISRCTN 47772397. *International journal of radiation oncology, biology, physics*, 76(3):747–754, March 2010. PMID: 19540054.

[134] Oscar Acosta, Gaël Dréan, Juan David Ospina, Antoine Simon, Pascal Haigron, Renaud de Crevoisier. Explaining relationships between local dose and rectal toxicity in prostate cancer radiotherapy with voxel-based population analysis. *Proceedings of Image-Guided and Multimodal Dose Planning in Radiation Therapy*, August 2012.

[135] Oscar Acosta, Gael Drean, Juan D. Ospina, Antoine Simon, Pascal Haigron, Caroline Lafond, and Renaud de Crevoisier. Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. *Physics in Medicine and Biology*, 58(8):2581, April 2013.

[136] Maximilian Niyazi, Peter Bartenstein, Claus Belka, and Ute Ganswindt. Choline PET based dose-painting in prostate cancer–modelling of dose effects. *Radiation oncology (London, England)*, 5:23, 2010. PMID: 20298546 PMCID: PMC2848061.

[137] D. B. Fuller, C. Lee, S. Hardy, and H. Jin. Virtual HDRsm CyberKnife prostate treatment: Toward the development of non-invasive HDR dosimetry delivery and early clinical observations. *International Journal of Radiation Oncology, Biology, Physics*, 69(3):S358–S359, November 2007.

[138] Siavash Jabbari, Vivian K. Weinberg, Tania Kaprealian, I-Chow Hsu, Lijun Ma, Cynthia Chuang, Martina Descovich, Stephen Shiao, Katsuto Shinohara, Mack Roach III, and Alexander R. Gottschalk. Stereotactic body radiotherapy as monotherapy or Post-External beam radiotherapy boost for prostate cancer: Technique, early toxicity, and PSA response. *International Journal of Radiation Oncology, Biology, Physics*, 82(1):228–234, January 2012.

[139] Florian Roser, Makoto Nakamura, Rosa Martini-Thomas, Madjid Samii, and Marcos Tatagiba. The role of surgery in meningiomas involving the optic nerve sheath. *Clinical neurology and neurosurgery*, 108(5):470–476, July 2006. PMID: 16191463.

[140] Giuseppe Minniti, Maurizio Amichetti, and Riccardo M Enrici. Radiotherapy and radiosurgery for benign skull base meningiomas. *Radiation Oncology*, 4(1):42, October 2009.

[141] Roberto Pacelli, Laura Cella, Manuel Conson, Fausto Tranfa, Diego Strianese, Raffaele Liuzzi, Raffaele Solla, Antonio Farella, Marco Salvatore, and Giulio Bonavolontà. Fractionated stereotactic radiation therapy for orbital optic nerve sheath meningioma - a single institution experience and a short review of the literature. *Journal of radiation research*, 52(1):82–87, 2011. PMID: 21293073.

[142] Pantaleo Romanelli, Livia Bianchi, Alexander Muacevic, and Giancarlo Beltramo. Staged image guided robotic radiosurgery for optic nerve sheath meningiomas. *Computer aided surgery: official journal of the International Society for Computer Aided Surgery*, 16(6):257–266, 2011. PMID: 21991922.

[143] Alexander Schlaefer, Tiberiu Viuleţ, Alexander Muacevic, and Christoph Fürweger. Multicriteria optimization of the spatial dose distribution. *Medical Physics*, 40(12):121720, 2013.

[144] Søren M. Bentzen, Louis S. Constine, Joseph O. Deasy, Avi Eisbruch, Andrew Jackson, Lawrence B. Marks, Ten Haken, Randall K, and Ellen D. Yorke. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. *International Journal of Radiation Oncology, Biology, Physics*, 76(3):S3–S9, March 2010.

[145] Simon S Lo, Arjun Sahgal, Jian Z Wang, Nina A Mayr, Andrew Sloan, Ehud Mendel, and Eric L Chang. Stereotactic body radiation therapy for spinal metastases. *Discovery medicine*, 9(47):289–296, April 2010. PMID: 20423672.

[146] Seil Sohn and Chun Kee Chung. The role of stereotactic radiosurgery in metastasis to the spine. *Journal of Korean Neurosurgical Society*, 51(1):1–7, January 2012. PMID: 22396835 PMCID: PMC3291699.

[147] Beant Gill, Eric Oermann, Andrew Ju, Simeng Suy, Xia Yu, Jennifer Rabin, Christopher Kalhorn, Mani N Nair, Jean-Marc Voyadzis, Keith Unger, Sean P Collins, K W Harter, and Brian T Collins. Fiducial-free CyberKnife stereotactic body radiation therapy (SBRT) for single vertebral body metastases: acceptable local control and normal tissue tolerance with 5 fraction approach. *Frontiers in oncology*, 2:39, 2012. PMID: 22645718 PMCID: PMC3355827.

[148] Qi Wang, Yongchun Song, Hongqing Zhuang, Xiaoguang Wang, Fengtong Li, Yang Dong, Jingsheng Wang, and Zhiyong Yuan. Robotic stereotactic irradiation and reirradiation for spinal metastases: safety and efficacy assessment. *Chinese medical journal*, 127(2):232–238, January 2014. PMID: 24438609.

[149] Tiberiu Viuleţ, Oliver Blanck, and Alexander Schlaefer. Analysis of spatial trade-offs for a spinal SRS case. *Medical physics*, 2014. Medical physics [0094-2405] Viulet, T J.:2014 Bd.:41 iss:6 S.:231 -231.

[150] Arinbjörn Ólafsson and Stephen J Wright. Linear programing formulations and algorithms for radiotherapy treatment planning. *OPTIMIZATION METHODS AND SOFTWARE*, 21(2):201—231, 2006.

[151] IBM. *IBM ILOG CPLEX Optimization Studio CPLEX User's Manual*. IBM, 2011.

[152] B. Fraass, K. Doppke, M. Hunt, G. Kutcher, G. Starkschall, R. Stern, and J. Van Dyke. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. *Medical Physics*, 25(10):1773–1829, 1998.

[153] Milton Friedman. *There's No Such Thing as a Free Lunch*. Open Court Publishing Co ,U.S., 1977.

[154] B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson. Tolerance of normal tissue to therapeutic irradiation. *International Journal of Radiation Oncology, Biology, Physics*, 21(1):109–122, 1991.

[155] Bochao Li, Ruimin Zhang, and Scott Kuhl. Minication affects action-based distance judgments in oculus rift HMDs. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, pages 91–94. ACM, 2014.

[156] Mary K. Young, Graham B. Gaylor, Scott M. Andrus, and Bobby Bodenheimer. A comparison of two cost-differentiated virtual reality systems for perception and action tasks. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, pages 83–90. ACM, 2014.

[157] Richard Yao, Tom Heath, Aaron Davies, Tom Forsyth, Nate Mitchell, and Perry Hoberman. Oculus VR best practices guide, 2014.

List of Figures

List of Tables