Scenario Dose Calculation for Robust Optimization in Proton Therapy Treatment Planning

Johan Sundström

Master's thesis
2021:E32

LUND UNIVERSITY

Faculty of Engineering
Centre for Mathematical Sciences
Mathematics

CENTRUM SCIENTIARUM MATHEMATICARUM

Abstract

Robust optimization plays a vital role in proton therapy treatment planning. A monoenergetic beam of protons has a finite range and delivers most of its dose at some specific depth in the patient. This allows the design of treatments that mainly damage the tumor and not any surrounding tissue. However, this property also makes proton therapy sensitive to uncertainties in the patient model used for treatment planning. For example, the patient might be slightly misaligned during the treatment and the delivered dose then deviates from the plan. Robust optimization samples such uncertainties into scenarios and simultaneously optimizes the treatment for all scenarios.

This thesis studies an approximate scenario dose calculation method with the potential to improve the computational efficiency of robust optimization in proton therapy treatment planning. The method uses the concept of water-equivalent path lengths to calculate scenario doses as range-corrected deformations of the nominal dose. The method is implemented in a commercial treatment planning system and compared to both a Monte Carlo baseline and an approximate alternative. Compared to the approximate alternative, the proposed method handles a wider class of scenarios which also includes patient setup rotations and anatomical changes.

The results indicate that the proposed method offers no improvements in terms of accuracy. However, when used in the context of robust optimization, no significant degradation of the optimized treatment plan quality is observed in three realistic patient cases. This indicates that the method might be useful when efficiency is the bottleneck. The method is not optimized for performance within the scope of this thesis. However, based on the results and a theoretical analysis, the largest performance gains can be expected for pencil beam scanning plans with many spots per beam.

Keywords: *cancer*, *radiation therapy*, *proton therapy*, *treatment planning*, *robust optimization*, *dose calculation algorithms*, *scenarios*, *patient setup*, *range error*, *organ motion*, *pencil beam scanning*, *computational efficiency*, *range-corrected dose-mapping*.

Preface

This master's thesis is written as a degree project within the Engineering Physics program at the Faculty of Engineering (LTH), Lund University. The project was carried out as a collaboration between LTH and RaySearch Laboratories AB during the spring of 2021.

I want to express my gratitude to RaySearch Laboratories for providing such a stimulating environment to work in and an interesting topic to investigate. Particularly, I am grateful to my RaySearch supervisor Björn Andersson who has provided excellent guidance throughout the entire project, helped me navigate in the world of radiotherapy and C++, and not hesitated to discuss both the core ideas and the tiniest details of the implementation. I am also grateful to Albin Fredriksson who initiated the project, and to the countless people at RaySearch who have provided feedback.

Furthermore, I want to thank my LTH supervisors Carina Geldhauser and Pontus Giselsson for helpful feedback on content and format of the report, and for their continuous support throughout.

Stockholm, June 2021
Johan Sundström

Abbreviations

Abbreviation	Definition
APM	Analytical probabilistic modelling
BEV	Beam's eye view
CRS	Compressed row-storage (matrix)
CT	Computed tomography
CTV	Clinical target volume
DVH	Dose-volume histogram
HU	Hounsfield unit
IMPT	Intensity-modulated proton therapy
INV	Inverse square law factor
MC	Monte Carlo
MCS	Multiple Coulomb scattering
MU	Monitor unit
OAR	Organ at risk
PB	Pencil beam (algorithm)
PBS	Pencil beam scanning
PTV	Planning target volume
QA	Quality assurance
RBE	Relative biological effectiveness
RC	Ray casting
RMSE	Root-mean-square error
ROI	Region of interest
RSP	Relative stopping power
SQP	Sequential quadratic programming
TPS	Treatment planning system
VSAD	Virtual source-axis distance
WEPL	Water-equivalent path length
WET	Water-equivalent thickness

Contents

1

Introduction

1.1 Background

Cancer ranks among the leading causes of death in most countries. Too many people are familiar with the tragedies it may cause. Luckily, cancer treatment is an active field of research and treatments continuously improve following technical advances.

One of the most common and effective types of cancer treatment is radiotherapy, in which ionizing radiation is used to kill cancer cells. Cancer cells are extra sensitive to radiation, but unfortunately radiation damages healthy tissue as well. Thus, the planning of radiotherapy treatments is always a balancing act between sufficiently irradiating the tumor and saving healthy tissue from damage. The main quantity of interest in radiotherapy treatment planning is the *absorbed dose*, which is the amount of radiation energy deposited per unit mass of the absorbing material.

There are many different treatment machines and radiation modalities with various strengths and weaknesses. This thesis deals with *external beam proton therapy* in which a narrow beam of accelerated protons is scanned across the tumor. Proton therapy allows treating a very specific volume of the patient, meaning that the dose is confined to a small volume. By confining the dose to a small volume around the tumor, the surrounding healthy tissue can be spared from unnecessary damage.

This strength of proton therapy is also a weakness. When the dose is confined to a small volume, the treatment becomes very sensitive to, *e.g.*, any small misalignments of the patient. The depth at which the protons deposits most of the dose depends heavily on the electron density of the material traversed by the beam. For example, the shape of the dose cloud changes dramatically if the beam unexpectedly passes through some bone, making proton therapy sensitive to such setup errors.

In proton therapy treatment planning, scenario-based *robust optimization* is used to ensure that the treatment is robust against small uncertainties, *e.g.*, in the patient position or in the beam range. Such optimization finds a treatment that works well for all *scenarios* in a set of sampled realizations of these uncertainties. To run robust optimization in a reasonable amount of time, computerized treatment planning systems rely on dose calculation algorithms that are both accurate and efficient.

This thesis investigates approximate scenario dose calculation algorithms. Specifically, the algorithm described in [Park et al., 2012] is of central interest. This algorithm approximates the dose in various scenarios by deforming the accurately calculated dose from a *nominal scenario*. The deformations are based on the concept of *water-equivalent path lengths*. Such approximate scenario dose calculations could, if accurate and quick enough, be used to make robust optimization more efficient. This, in turn, could allow for including a larger number of scenarios in optimization and allow the treatment planner to spend more time tuning the formulation of the optimization problem for best results.

1.2 Aims and objectives

The general aim of this thesis is to investigate and compare alternative scenario dose calculation methods for robust optimization in proton therapy treatment planning. The reason for investigating this is that current algorithms are sometimes prohibitively slow and memory inefficient. Advancements in the field could lead to higher quality treatments for patients and to better efficiency at clinics. Ideally, we would like to answer the question:

> Can a scenario dose calculation algorithm in the style of [Park et al., 2012] improve robust optimization efficiency in proton therapy treatment planning?

However, this question cannot be expected to be answered fully within the scope of the thesis. Judging the usability of dose calculation methods is, in the end, a clinical decision. Furthermore, all possible implementation variants cannot be tested. The aim is instead sub-divided into a few research questions of smaller scope.

Research questions

The research questions are as follows:

I Can a scenario dose calculation algorithm in the style of [Park et al., 2012] be incorporated into the robust optimization procedure of a commercial treatment planning system?

II Why does robust optimization efficiency depend on the choice of scenario dose calculation algorithm?

III Which scenario dose calculation algorithms are already available, and how can these be fairly evaluated and compared with the proposed method?

IV What are the strengths and weaknesses in terms of accuracy and efficiency of the proposed method compared to the algorithms that already exist?

V How could the proposed method be improved?

where the proposed method refers to the implementation in I.

Limitations

Optimizing the efficiency of the proposed implementation is beyond the scope of this thesis. Instead, optimizing and assessing the method in terms of accuracy is of main concern. Accuracy is the more fundamental limitation of the two. If the method proves to be accurate enough, it can later be optimized for performance.

1.3 Methodology

To answer the research questions II, III, and IV, this thesis first provides a brief literature review of robust optimization for treatment planning and of dose calculation algorithms. Answering research questions I, IV, and V requires implementing the method. The implementation is developed iteratively during the project, regularly gauging the effect of changes by evaluating the performance on a few small test cases. At the end of the project, the best implementation in terms of accuracy is evaluated by re-running dose calculations and robust optimizations on a set of realistic patient test cases. The patient test cases are selected after discussion with experts within the field to be challenging for the proposed method.

1.4 Contributions

The main contribution of this thesis is that a new method for approximate scenario dose calculation is implemented and evaluated. The method extends a previously published method to handle a wider class of scenarios. It is very efficient in terms of the number of full Monte Carlo simulations that has to be run, relying only on simulation results for the nominal scenario.

The method is here evaluated against a Monte Carlo baseline as suggested by [Park et al., 2012] who evaluated it against a pencil beam baseline. Furthermore, the method is incorporated in a robust optimization procedure. This was mentioned as a potential use case in the original article, but such results has to the best of our knowledge not been published before.

1.5 Outline of the report

This thesis is about scenario dose calculation algorithms for proton therapy treatment planning. The discussion uses many concepts and technical terms which the reader is not expected to know unless already familiar with radiotherapy treatment planning. Thus, this report features three chapters attempting to cover the necessary background. Abbreviations are summarized in a table on page 7.

- Chapter 2 covers general concepts in radiotherapy such as *dose* and *treatment planning*. It also includes a description of proton therapy delivery techniques. The discussion of proton therapy physics introduces the highly relevant concepts of *range* and *water-equivalent path lengths*.
- Chapter 3 introduces the concept of inverse treatment planning, which means optimizing treatments using mathematical optimization. Specifically, scenario-based robust optimization is discussed as a way of accounting for model uncertainties. The mathematical notation introduced in this chapter is used throughout the report.
- Chapter 4 uses concepts from the two previous chapters when describing various dose calculation algorithms. This includes both the accurate Monte Carlo simulations and the scenario dose approximation algorithms used in the experiments.
- In the experiments, a specific algorithm is implemented and compared to alternatives. Chapter 5 describes the technical details about the implementation and the experiments are described in Chapter 6. Chapter 7 presents and discusses the results. Finally, conclusions are summarized in Chapter 8.

Readers familiar with radiotherapy treatment planning may skip large parts of Chapters 2-4 and still appreciate the rest of the report. However, it is not recommended to skip past Section 3.2 which introduces some mathematical notation, or Section 4.3 which sheds some light on the scenario dose approximation methods. Unless already familiar with robust optimization, glancing through Section 3.3 is also recommended.

2

Radiation therapy

This chapter introduces general concepts in radiation therapy and treatment planning and may be skipped by anyone well familiar with the field.

2.1 General concepts

Radiation

Radiation is energy in motion, emitted and transmitted in the form of particles or waves. One common example is electromagnetic radiation, which includes radio waves, visible light, and the high-energetic gamma rays used in photon therapy. Such radiation always travels at the speed of light. Particles with rest mass traveling at high speeds, such as accelerated electrons, protons or heavier ions, is another type of radiation [Canfield Upton et al., 2021].

Radiation therapy, or radiotherapy, uses *ionizing radiation*. Such radiation has enough energy to ionize atoms or molecules by knocking away some of their electrons. These ionization events may damage biological tissue directly or indirectly, which is turned to an advantage when treating cancer.

Biological effects of radiation

A central concept in this thesis is *absorbed dose*, or dose for short. The purely physical dose quantity, the *physical dose*, is a measure of how much radiation energy is absorbed per mass unit of the absorbing material. Physical dose is measured in the derived SI-unit Gray (Gy), where $1\,\mathrm{Gy} = 1\,\mathrm{J/kg}$.

The extent of radiation damage to biological tissue depends on the physical dose. However, different types of radiation have different *relative biological effectiveness* (RBE) relative to photon radiation. For instance, it is commonly approximated that 1 Gy of proton radiation has similar biological effect as 1.1 Gy of photon radiation, implying a constant RBE of 1.1 for protons. Under this approximation, 1.0 cGy of physical dose from proton radiation may also be expressed as an equivalent RBE dose 1.1 cGy(RBE). More refined RBE models may depend also on the absorbing tissue, the energy of radiation, and the temporal pattern of radiation delivery [Paganetti, 2018].

When radiation travels through biological tissue, it interacts with the atoms and molecules of the cells. In most of these interactions, the affected atoms and molecules temporarily reach excited states and the added energy is soon transferred to heat. What causes most biological effects, though, is ionization events that directly or indirectly damage the DNA.

DNA is a large and double-helix shaped molecule that encodes all genetic information a cell needs to function, grow and reproduce. Damaging the DNA means altering the chemical structure of this molecule which might corrupt the important information it carries. Cells have significant capabilities to repair such damage via various mechanisms and lasting damage occur only when the repair fails.

A high speed electron used in radiotherapy traverses a cell in approximately 10^{-14} seconds. During this *physical phase* of radiant exposure, the primary excitation and ionization events occur. Such events create 'free radicals' which are broken molecules with unpaired electrons. The free radicals are highly reactive and during the *chemical phase*, which lasts about 1 ms, they engage in a series of chemical reactions that might involve and damage the DNA molecule [Joiner and Kogel, 2018]. Such damage is an indirect effect of radiation, as differentiated from direct effects caused by the primary energy deposition. See Figure 2.1 for an illustration.

As described above, the physical and chemical phase extend over a quite short period of time. The subsequent *biological phase*, however, may continue for many years. The biological phase starts with enzymatic reactions acting on the radiation imposed chemical damage. Starting there, the cell

is usually successful in repairing any DNA damage. On the occasion it fails, the result might sooner or later be cell death [Joiner and Kogel, 2018].

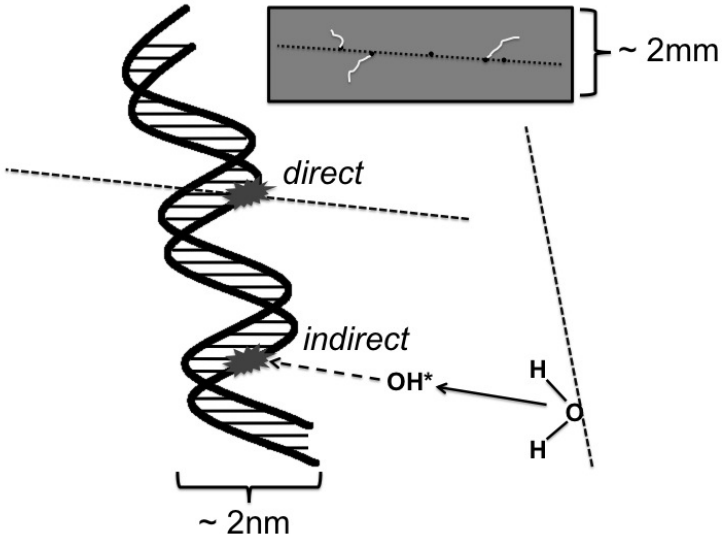


Figure 2.1 Figure 22:1 from [Paganetti, 2018] reproduced with permission[1]. The figure illustrates the difference between direct and indirect DNA damage caused by particles traveling along the dashed lines.

The curative effects of radiotherapy is attributed to the fact that sufficiently high doses of radiation cause tumor cell death. For healthy tissue, on the other hand, cell death remains an unwanted side effect. Patients undergoing radiotherapy may experience mild, severe or even lethal side effects if normal tissue receive too high doses of radiation. Early side effects may appear weeks or months after the exposure. Late side effects, such as secondary tumors, may show up after many years. With side effects in mind, designing radiotherapy treatments becomes a balancing act where efficient tumor control must be weighed against the side effects that come with higher doses.

The role of radiotherapy

There is unfortunately no universal cure for cancer. Instead, many different techniques are used to treat patients suffering from the disease. Some treatments are curative, meaning that they aim to cure the patient. Others are palliative, *i.e.*, devoted to ease the symptoms.

Cancer treatments can be divided into surgery, radiotherapy, and systemic treatments, where systemic treatments include chemotherapy, targeted therapy, hormonal therapy, and immunotherapy. Each treatment modality has different strengths and may be better suited for some cancer sites than others. Often a combination is used to treat each patient and the prevalence of different treatment modalities varies across cancer sites [Miller et al., 2019].

About half of cancer patients receive radiotherapy at some point [Joiner and Kogel, 2018]. Radiotherapy is used both in curative and palliative care, as sole treatment or in combination with other modalities. For instance, it can be used in combination with surgery. The purpose could be to shrink the tumor in order to make it operable, or to kill any cancer cells remaining after operation. Radiotherapy can also be combined with some systemic treatments to make them more effective.

Delivery techniques

There are three main types of delivery techniques in radiation therapy. Most common is *external beam therapy*, in which the radiation originates from an external source and reaches the patient's body in the form of beams. In *brachytherapy*, sealed radioactive sources are physically inserted into the patient's body. In *radionuclide therapy*, radioactive substances are instead introduced into the patient's body by injection or ingestion. These substances are designed to localize to specific locations in the body, making it a targeted therapy. From now on, this thesis deals exclusively with external beam therapy.

[1] Permission conveyed through Copyright Clearance Center, Inc.

External beam modalities. Different types of radiation sources are used in external beam therapy. *Photon* therapy uses electromagnetic radiation in the form of X-rays and gamma rays. *Electron* therapy uses beams of accelerated electrons. Finally, particle therapy uses beams of *neutrons*, *protons* or other heavier ions.

Clinical reasons for considering different modalities are twofold. Firstly, radiation from different modalities have different biological impact, *i.e.*, the RBE depends on the modality. Secondly, the dose distributions that can physically be achieved in the patient also depends on the modality.

2.2 Proton therapy

Since this thesis investigates a dose calculation method with intended applications in proton therapy, a section devoted to proton therapy is warranted. Central concepts introduced here include *pencil beam scanning* (PBS), *relative stopping power*, *range*, and *water-equivalent path length* (WEPL). These concepts are needed to describe the dose calculation method central to this thesis.

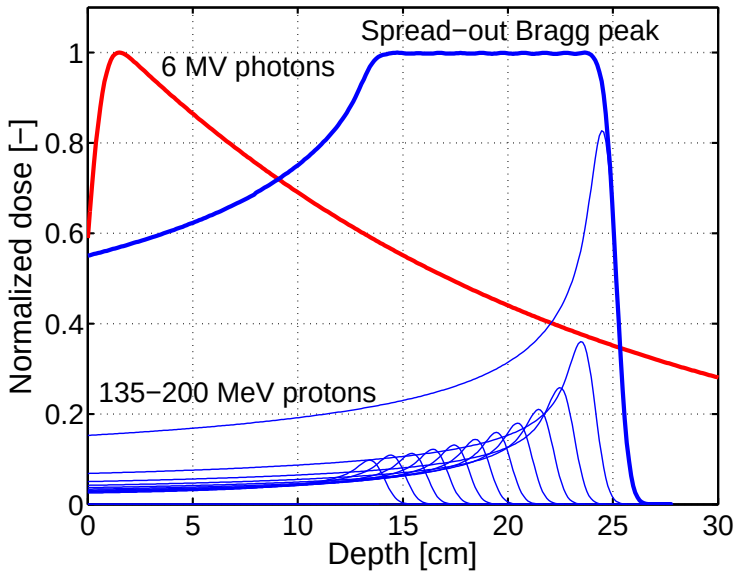


Figure 2.2 Depth-dose curves along the beam central axis. When appropriately superposing Bragg peaks from various monoenergetic proton beams, a spread-out Bragg peak of uniform dose with a sharp distal falloff appears deep in the patient. Figure reproduced with permission from [Bokrantz, 2013].

Pros and cons of proton therapy

There exist more than 100 proton therapy clinics worldwide, and more facilities are under construction [PTCOG, 2021]. Figure 2.2 illustrates the most important argument in favor of proton therapy over other modalities. A monoenergetic proton beam delivers most of its dose at a specific depth in the patient, referred to as the *Bragg peak*. Photon beams, on the other hand, penetrate straight through the patient with the dose decaying in an exponential manner beyond the peak. Both modalities can be used to treat deep tumors. However, taking advantage of the shape of the proton Bragg curve, it is possible to create treatment plans with as high dose values within the tumor and lower levels in most of the healthy tissue surrounding it. Thus, proton therapy has the potential to decrease the side effects of radiotherapy.

The sharp Bragg peak is a strength of proton therapy but also its weakness. When conforming the dose very small volume in the patient, it also becomes extremely important to ensure that this volume actually covers all of the tumor as intended. Even a small misalignment of the patient might lead to the proton beams traversing tissue of different material composition (*e.g.* bone instead of lung), which can dramatically affect the locations of the Bragg peaks. Photon therapy is not nearly as sensitive to density heterogeneities, and consequently it is a lot easier to ensure *robustness* of photon treatment plans than it is for proton treatment plans.

Some other disadvantages of proton therapy include bulky and expensive treatment machines and comparatively long treatment delivery times [Lomax, 2018]. However, this might change as the

technology matures. For the interested reader, [Mohan et al., 2017] provides an overview of current limitations of proton therapy and efforts needed to overcome them.

Proton therapy physics

The proton as a particle. Protons are subatomic particles. Any given atom has at least one proton in its nucleus. In fact, the number of protons in the nucleus is what defines the elements of the periodic table. Protons have a positive, elemental electrical charge $1\,e$ and a rest mass of $1.67 \times 10^{-27}\,\mathrm{kg}$, which is approximately 1836 times more than the rest mass of an electron.

Proton interactions. The fact that protons are electrically charged allows them to interact with external *electromagnetic fields*. This is utilized in the design of proton therapy treatment machines which use electric fields to accelerate the protons and magnetic fields to deflect them in various directions. Beam energies are usually expressed in units of electronvolt (eV), corresponding to the energy gained by a proton when accelerated through a 1 V potential difference. Uncharged particles, like neutrons, cannot be accelerated in the same way and require other machine designs. The positive charge of protons also affects how protons interact with matter, which in turn determines where and how a proton beam deposits dose in the patient.

Protons traveling through a medium interact both with the atomic nuclei and the electrons surrounding them. However, given that the nuclei occupy roughly 10^{-15} of the volume of the atoms, interactions with electrons are far more likely to occur, making Coulomb scattering by atomic electrons the main mechanism of energy loss [Krane, Halliday, et al., 1988]. Given the comparably small mass of electrons, many thousands of such scattering events are needed to make the proton lose all its initial energy. In some of these interactions, the electron receives enough energy to become a secondary ionizing particle. In each individual interaction with an electron, the proton may be deflected by some small angle. The effect of many subsequent such events is a broadening of the beam which is treated rigorously by multiple Coulomb scattering (MCS) theory.

Interactions with nuclei are less common, but occasionally a proton passes close enough to some nucleus for an interaction to occur. Since a nucleus has a comparably large mass, such interactions may deflect the proton from its otherwise straight path. When interacting with a nucleus, the proton may simply deflect due to the repulsive Coulombic force (Rutherford scattering), or it may engage in a nuclear reaction in which the nucleus absorbs the primary proton and proceeds to emit some secondary particles in various directions [Newhauser and Zhang, 2015]. Probabilities of nuclear reactions are usually called *cross-sections*.

Stopping power and range. The fate of an individual proton is determined by the interactions that happen to occur along its path. For a beam of protons containing a very large number of particles, it is meaningful to consider the *average* particle behavior. For example, the *range* of a beam is defined as the depth at which half of the surviving primary particles have stopped. In clinical settings, the range is more commonly defined based on dose levels. According to a common definition, the range equals the distance to the distal point on the Bragg curve (beyond the Bragg peak) where the the dose level is 80 % of the Bragg peak dose [Paganetti, 2018].

Traveling through a medium, the average proton experiences a retarding force due to Coulombic interactions with electrons that eventually forces it to lose all its initial kinetic energy. The magnitude of the energy loss per unit length

$$S = -\frac{dE}{dx} \tag{2.1}$$

is called the *stopping power*. The stopping power is proportional to the mass density ρ of the absorbing medium, and often the *mass stopping power* S/ρ is considered instead. The mass stopping power depends also on other qualities of the absorbing material as well as the energy of the particle. According to the Bethe-Block formula, mass stopping power can be calculated as

$$\frac{S}{\rho} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A}\frac{z^2}{\beta^2}\left(\ln\frac{2m_e c^2\beta^2}{1-\beta^2} - \beta^2 - \ln I - \frac{\delta}{2} - \frac{C}{Z}\right) \tag{2.2}$$

where N_A is Avogadro's constant, r_e is the classical electron radius, m_e is the electron rest mass, c is the speed of light, Z is the atomic number and A the atomic weight of the absorbing material, z is the charge of the proton, $\beta = v/c$ where v is the velocity of the proton, I is the mean excitation energy of the atomic electrons, and finally δ and C are part of correction terms relevant only at high and low proton energies respectively [Newhauser and Zhang, 2015].

From the complicated expression in (2.2), one can observe that the stopping power is actually proportional to the *electron density* $\rho_e = N_A \rho Z/A$ of the absorbing material rather than just the mass density. This is reasonable since the dominating mechanism for energy loss is Coulombic interactions with electrons.

As reflected by β in (2.2), the stopping power also depends on the particle energy. When the proton energy decreases, the stopping power increases. For a beam of protons with initial energy E, the range R can be calculated by an integral

$$R(E) = \int_E^{\varepsilon} \left(\frac{dE'}{dx} \right)^{-1} dE' \tag{2.3}$$

where $\varepsilon > 0$ is some small final energy [Paganetti, 2018]. The shape of the Bragg curve (Figure 2.2) is a consequence of how stopping power varies with the proton energy. Specifically, the rate of energy loss is much higher when the velocity of the proton reaches zero which explains the rise of the Bragg peak.

Water equivalent paths. Protons interact with human tissues in a similar fashion as with water. In clinical practice, it is common to specify the penetrating capacity of a beam by its range in water. In this context, the water-equivalent thickness (WET) of some slab of material is interpreted as the thickness of a slab of water having the same stopping effect on the radiation.

To define WET, we need the concept of *relative stopping power* (RSP), defined as the quotient between the stopping power in a material and in water. Ignoring the final two correction terms in (2.2) and grouping all β-dependence into a function $f(\beta)$, the RSP is well approximated by

$$\frac{S_m}{S_w} = \frac{\rho_e^m}{\rho_e^w} \frac{(f(\beta) - \ln I_m)}{(f(\beta) - \ln I_w)} \tag{2.4}$$

where m is short for material and w for water [Abbema et al., 2018]. The RSP of a given material thus mainly depends on its electron density relative to water. To some degree it also depends on the material-dependent mean ionizing potential, I, which prevents the energy-dependence from canceling.

For a thin slab of material m with physical thickness t_m, the water-equivalent thickness (WET) t_w is given by

$$t_w = t_m \frac{S_m}{S_w} \tag{2.5}$$

which, as follows from the weak energy dependence in (2.4), depends only slightly on the proton's energy when entering the slab.

The RSP is also used to define the related concept of water-equivalent path length (WEPL). Ignoring the energy-dependence of RSP, the WET of several consecutive slabs can be calculated by summing over the individual slabs. A more physical definition would account for the energy loss over each slab and adjust the RSP accordingly. However, keeping the approximation of energy-independence, the WEPL along some curve l in a heterogeneous medium can be defined as a curve integral over a constant RSP scalar field

$$\text{WEPL} = \int_l \frac{S_{m(p)}}{S_w} \, ds \tag{2.6}$$

where the material m is a function of the point $p \in \mathbb{R}^3$ in space. This definition avoids making WEPL an energy-dependent quantity. However, to evaluate the integral, one still has to choose some energy for the RSP calculations with different choices yielding somewhat different results.

Clinical proton beams

In proton therapy, the beams reaching the patient are generated using one of two techniques. The first uses *passive scattering*. In this technique, an initially narrow and monoenergetic beam is scattered to achieve a wide lateral beam profile. By placing physical collimators and range compensators in the beam path, the scattered beam is then shaped according to the clinical needs. Most new treatment centers built today, however, use another technique called *pencil beam scanning* (PBS).

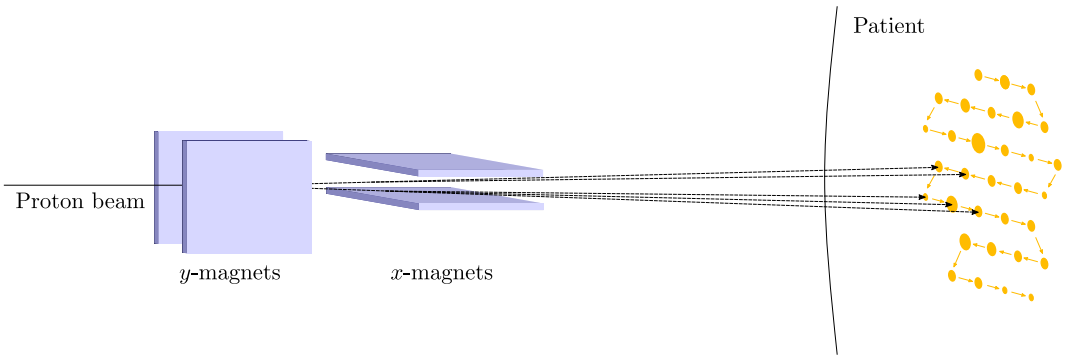


Figure 2.3 In some PBS treatment machines, two pairs of magnets are used to control the beam direction. The beam can then be scanned across the tumor while also varying the energy of the proton beam.

Pencil beam scanning. Pencil beam scanning (PBS) is sometimes referred to as intensity-modulated proton therapy (IMPT) or spot scanning. The technique is quite general and can be used for other charged particles than protons as well. As compared to passive scattering techniques, PBS does not rely on passive scattering to widen the narrow and monoenergetic beam from the accelerator. Instead, the beam is *scanned* across the target volume, irradiating only part of the volume at each instant as illustrated in Figure 2.3.

During the scanning process, properties of the beam such as the energy might also be varied. Adjusting the beam energy allows the Bragg peak to be placed at different depths in the patient. Recalling Figure 2.2 on page 16, the Bragg peak is where most of the dose is deposited.

PBS is a highly flexible delivery technique thanks to the high degree of beam control. The ability to aim the beam in various directions allows for shaping the beam in the lateral direction. By varying the beam energy, Bragg peaks can be placed at varying depths which allows the dose to be shaped in the depth direction. The machine can effectively paint a dose distribution of almost any shape in the patient by superposing a large number of Bragg peaks.

Often, the combinations of beam directions and energies achievable by a PBS treatment machine are discretized into *spots*. The spots are organized in energy layers, and within an energy each spot corresponds to a specific beam direction. Figure 2.4 illustrates a scanning path between spots also seen in Figure 2.3.

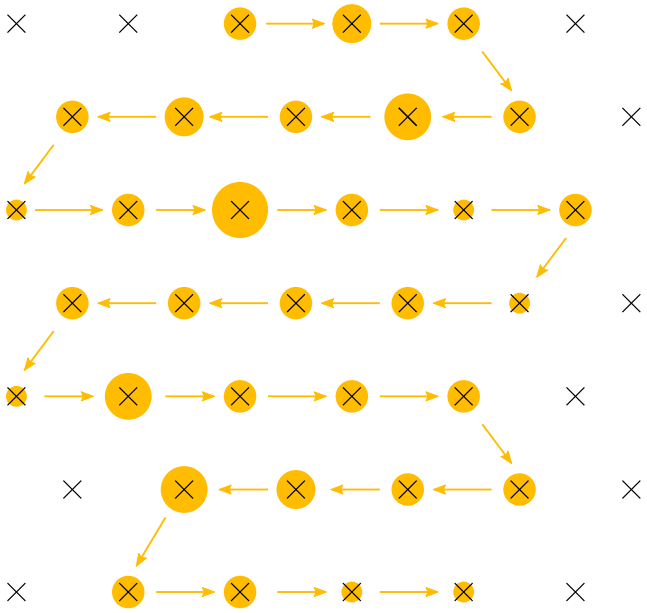


Figure 2.4 Beam's eye view (BEV) of the scanning between spots within an energy layer. The spot weights are proportional to the number of protons delivered by the beam in each spot and are here illustrated as discs of varying area.

2.3 Treatment planning

The treatment planning process

The goal of radiation therapy treatment planning is to find the best possible treatment for a given patient. The planning process usually involves many steps and requires collaboration between several professions. As illustrated in Figure 2.5, treatment planning is usually an iterative process. While sharing key elements, the exact procedure might vary significantly between clinics, treatment techniques, and cancer sites. Keeping that in mind, this section does not provide a comprehensive description of all possible clinical workflows but is intended to put treatment planning in a clarifying context for readers who are less familiar with the topic.

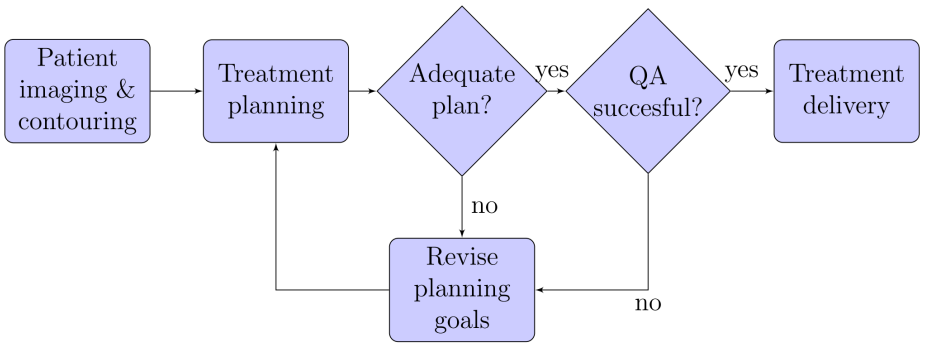


Figure 2.5 Treatment planning is usually an iterative procedure in which the planning goals might need to be revised several times. Figure reproduced with permission from [Andersson, 2017].

Initial phase. Radiation therapy treatment planning is initiated when an oncologist, a physician specialized in cancer care, prescribes radiation therapy as part of the treatment for some patient. See Section 2.1 for an overview of the clinical usage of radiation. At this stage, the patient has received the cancer diagnosis and probably undergone a medical imaging exam during which the tumor(s) have been located.

Clinical goals. Since radiation damages all cells, one overarching goal of treatment planning is to adapt the treatment such that the radiation is optimally confined to the patient-specific tumor location, or *target*. For the patient, both under-dosage and over-dosage may be lethal. Too much radiation results in severe damage to previously healthy tissue. Too little results in ineffective tumor control. Thus, an optimal treatment plan ensures that all cancer cells are sufficiently irradiated while surrounding, healthy tissue is spared from unnecessary damage. Unfortunately, the two goals are usually in conflict and some trade-offs must be made.

A physician is responsible for making the trade-off between target coverage and healthy tissue sparing and formulates a set of clinical goals for the treatment to this end. The clinical goals are based on clinical evidence and experience and for example usually include some constraints on the dose delivered to the patient. This could include a minimum dose level in the target and maximum dose levels in neighboring organs at risk (OARs). Often, the physician is involved in delineating the targets and OARs in the planning image of the patient.

Configuring the treatment machine. Given the clinical goals, the question remains how the treatment machine should be configured to achieve the desired outcome. This is the job of the treatment planner. The complexity of the task depends on the degrees of freedom in the treatment setup, which in turn depends on the delivery technique, modality of radiation, and ultimately the available parameters of the treatment machine. For most treatments, the treatment planner makes use of an advanced computer program called a treatment planning system (TPS) to complete the task. The TPS includes models of both the patient and the treatment machine and allows for simulation of hypothetical treatments. This allows the planner to simulate several treatments and adjust parameters until the treatment looks good in simulation. Once the planner is satisfied with a simulated treatment, the physician either approves the plan or suggests modifications. Treatment planning is often an iterative procedure that requires tight collaboration between professions.

QA and delivery. Once the the treatment plan is approved by the physician it is sent to a physicist. The role of the physicist is to review the plan and to perform quality assurance (QA). The QA might

include delivering the treatment to a water phantom, then measuring the delivered dose and ensuring than the measured dose corresponds close enough to the dose calculated by the TPS. Once the plan has passed the QA, the treatment can be delivered to the patient. For fractionated treatments, the treatment is delivered at several occasions, *fractions*, over the course of several days or weeks.

Treatment planning systems

The computerized TPS has a central role in the radiation therapy planning process. Reduced to its core functionality, the TPS allows the user to import anatomical data and facilitates the design of a personalized radiation treatment for each patient. In the process, the TPS simulates delivery of hypothetical treatments and presents the resulting dose distributions to the user. The user in turn makes changes to the plan or proceeds to get it approved for actual delivery. When the plan is approved, the TPS outputs detailed treatment instructions to the treatment machine.

To run treatment simulations, the TPS needs an accurate model of both the treatment machine and the patient. Through calibration procedures, medical physicists ensure that the treatment machine model corresponds well enough to reality before it is put to clinical use. In this calibration procedure, the machine is fed various instructions and the physical output (such as the dose in water phantom) is measured. The medical physicists compare these measurements to the model predictions and adjust parameters of the model until it is accurate enough. The patient is modeled based on medical images as described in the next section. TPS dose calculations are described in detail in Chapter 4.

Patient models

An accurate 3D model of the patient as positioned in the treatment machine is a necessary requirement when planning and simulating radiotherapy treatments. Different medical imaging techniques can be used to this end. Common examples are computed tomography (CT) scans and magnetic resonance imaging (MRI).

Treatment position. Before acquiring the images, the patient is fixated in a position that has been selected for the treatment. It is important that the position is both comfortable and reproducible since the treatment might be delivered over many fractions, *i.e.*, on many separate occasions. The planning is usually done based on images taken before the treatment has started, and before each delivery the patient must resume the exact same position with aid from the medical personnel. To reduce uncertainty in the patient position, different fixation equipment might be used to physically secure the patient in place. This is especially important in proton therapy due to the sharp dose falloff in proton beams which makes the delivery sensitive to setup errors [Paganetti, 2018]

Medical imaging. The medical images are imported in the TPS and used to reconstruct a 3D model of the patient. Since different imaging techniques are based on different physical interactions, different image modalities provide different information about the patient.

In constructing the 3D model, the images have two main purposes. The first is to allow for identification and location of different structures in the body. Based on the images, it must be possible to delineate volumes of interest such as tumors and organs since these volumes are part of the clinical goal definitions. The second purpose is to create an accurate 3D representation of the body in terms of material. Material properties such as mass density and elemental composition determine the behavior of the radiation in the body. An accurate material representation is thus necessary input to any accurate dose calculation. See for instance the discussion in Section 2.2 on how the proton stopping power depends on material properties. Usually, material properties are derived from the CT images. These images display radiodensity in terms of Hounsfield units (HU) for the X-rays used in imaging. Converting this information to RSP for protons usually requires fitting some curve to experimental data [Paganetti, 2018]. Errors in this HU-RSP conversion is one source of uncertainty in TPS-calculated dose distributions.

Image registration. If several images are used to define the patient model, one must consider the problem of image registration, *i.e.*, the process of aligning images from different measurements in a shared frame of reference. Such an alignment is necessary in order to locally combine information between the images.

Assuming that the patient remains exactly still between the measurements, a linear transformation (translation, scaling and rotation) may suffice to align the images. In a more realistic scenario where the patient might move slightly–for example by breathing–deformable registration is necessary. In deformable registration, one of the images is allowed to be locally warped in addition to the global, linear transformation. Unfortunately, deformable registration is an inherently ill-posed

problem [Paganetti, 2018]. Nevertheless, algorithms have been developed that take this challenge into account and the functionality is available in some TPS.

Segmentation. An important aspect of the patient model is the segmenting of the full volume into smaller regions of interest (ROIs). This can be done manually, automatically, or with a combination of automatic and manual tools. Segmentations of different ROIs may be defined on different images, and are then mapped to the same space using the image registration transformations.

3

Inverse treatment planning

Treatment planning is the process of translating the clinical goals prescribed by a physician into an actual, deliverable treatment plan. See Section 2.3 for context. To keep the discussion relevant to the thesis, this chapter focuses exclusively on inverse planning for PBS proton therapy.

3.1 Definition

In inverse treatment planning, the treatment planner specifies the desired *outcome* of a treatment and leaves the computerized treatment planning system (TPS) to figure out the optimal treatment parameters to match. This should be compared to forward planning, in which the dosimetrist manually specifies the *treatment parameters* and the TPS is simply used to simulate the outcome. Figure 3.1 illustrates the two situations.

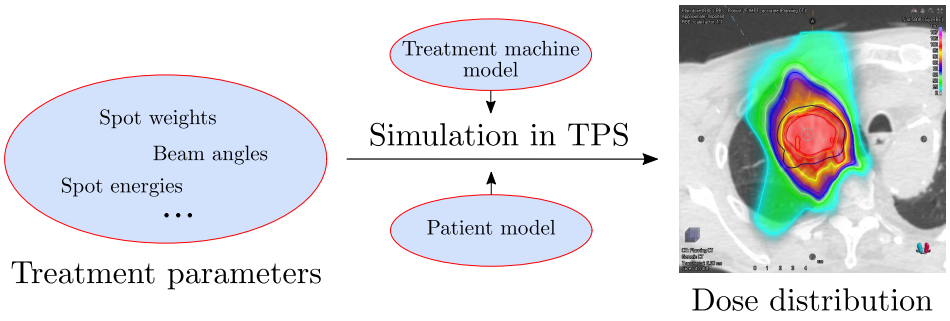


Figure 3.1 In *forward planning*, the planner specifies the desired treatment in the treatment parameter domain (left) and uses the TPS to calculate the resulting dose distribution (right). In *inverse planning*, the planner instead specifies the desired treatment in the dose domain (right) and asks the TPS to find a set of treatment parameters (left) that would result in the desired outcome. This requires the TPS to, in a sense, invert the order of cause and effect in the physics simulation. Hence the name inverse planning.

3.2 Mathematical framework

The patient. The patient is modeled as a three-dimensional volume in the physical space which is discretized into a finite set of *voxels* indexed by $\mathcal{V}$. The voxels are small, box-shaped, volume elements that typically are a few millimeters wide. A full patient volume might then consist of approximately $10^6 - 10^7$ voxels.

Regions of interest (ROIs), such as target volumes or organs, are subvolumes of the patient volume. Each ROI is modeled as a subset $\mathcal{R} \subset \mathcal{V}$ of the voxel grid indices paired with a volume vector $v = (v_i)_{i \in \mathcal{R}}$ containing the ROI volume within each voxel. Element v_i in v equals the volume of voxel i if the voxel is fully contained within the ROI. For voxels at the ROI boundary, v_i has a value of something in between zero and the full voxel volume. The derived relative volume vector is defined as $r = (r_i)_{i \in \mathcal{R}}$ with $r_i = v_i / \sum_i v_i$.

The beams. According to the treatment plan, the patient is irradiated by one or several proton beams which are here indexed by $\mathcal{B}$. Geometrically, each beam is associated with a source position outside the patient and an *isocenter* usually located somewhere in a patient target volume. The

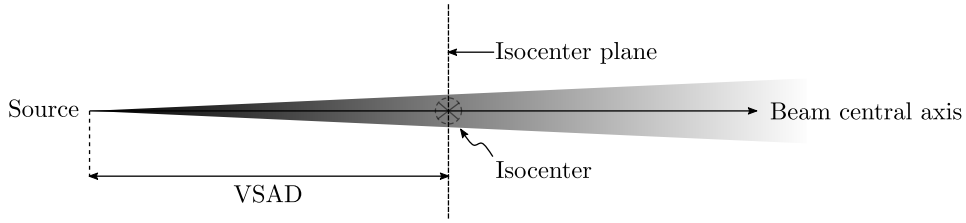


Figure 3.2 Illustration of geometrical beam concepts such as isocenter, isocenter plane, (virtual) source, beam central axis, and virtual source-axis distance (VSAD). The source is usually mounted on a gantry that can be rotated around the isocenter.

isocenter defines the origin of the beam's eye view (BEV) coordinate system. The affine plane that intersects the isocenter and is perpendicular to the beam central axis is called the *isocenter plane*. See Figure 3.2 for an illustration.

Conveniently, the spots associated with each beam $b \in \mathcal{B}$ are defined in the isocenter plane. The spots are grouped by energy layers $l \in \mathcal{L}_b$, and for each energy layer l, the spots represent a unique geometrical location in the isocenter plane. However, the number and location of spots may vary between beams and energy layers. Thus the spots are indexed by a beam- and energy-dependent index set $\mathcal{T}_{bl}$[2]. We find that the total number N of spots is given by

$$N = \sum_{b \in \mathcal{B}} \sum_{l \in \mathcal{L}_b} |\mathcal{T}_{bl}| \tag{3.1}$$

where $|\cdot|$ denotes the cardinality of a set.

Each spot $t \in \mathcal{T}_{bl}$ has a corresponding monitor unit (MU) value x_{blt} which is a machine-specific measure of radiation output. Monitor units can be defined in several ways. Often 1 MU is calibrated to correspond to the amount of some machine-controllable quantity (*e.g.* irradiation time) that results in 1 cGy of dose in a reference setting [Holmes et al., 2013]. Sometimes, the spot MUs are expressed as fractions of the total beam MU and are then referred to as *spot weights*. The spot MU variables are collected in a column vector $x = (x_{blt})_{(b,l,t) \in \mathcal{B} \times \mathcal{L}_b \times \mathcal{T}_{bl}}$. For convenience, we from now on enumerate these by $1 \dots N$ and write $x = (x_k)_{k \in 1 \dots N}$.

The dose. The dose is discretized to the same set of voxels as the patient. Within each voxel, the dose is considered constant, and the full dose distribution is represented by a column vector $d \in \mathbb{R}^{|\mathcal{V}|}$. We assume a linear relationship between the dose d and optimization variables x given by

$$d = Dx, \tag{3.2}$$

where $D \in \mathbb{R}^{|\mathcal{V}| \times N}$ is called the *dose deposition matrix*. An element D_{ij} in the matrix can be interpreted as the dose deposited to voxel i per unit MU of the scanned proton beam while set to the position and energy corresponding to spot j. Calculating the elements of D amounts to solving a radiation transport problem using models of the beam and the patient. Dose calculation algorithms are the subject of Chapter 4.

3.3 Optimization for treatment planning

The inverse treatment planning problem is naturally treated as a mathematical optimization problem. The treatment parameters are considered optimization variables and the objective is a function of the delivered dose. In this section, the mathematical formulation of the PBS treatment planning problem is formalized.

Objectives and constraints

In inverse planning, objectives and constraints are generally specified as functions of the dose and thus only indirectly of the optimization variables. This is useful for the treatment planner since the physical dose relates more closely to the biological effects of radiation than does, *e.g.*, the spot weights. In proton therapy it is usually assumed that the relative biological effectiveness (RBE) is a constant factor of 1.1. Under this approximation it is sufficient to formulate objectives in terms of

[2] where notation $\mathcal{T}$ is used to avoid confusion with the index set $\mathcal{S}$ for scenarios introduced later.

the physical dose [Paganetti, 2018]. One might also abstract further, create a more advanced RBE model, and specify objectives in terms of the RBE dose. For the purpose of this thesis, it is sufficient to discuss physical dose objectives.

Commonly, the full optimization problem is constructed from a set of objectives $\{f_i(d)\}_{i\in\mathcal{O}}$ and a set of dose-constraints $\{c_j(d) \leq 0\}_{j\in\mathcal{C}}$ specified by the treatment planner. The numbers of such objectives and constraints vary from case to case and we here index them by $\mathcal{O} = 1,\ldots,|\mathcal{O}|$ and $\mathcal{C} = 1,\ldots,|\mathcal{C}|$ for generality.

Each dose-objective and dose-constraint is specified on a specific ROI. Below, a selection of such objectives is presented. The selection corresponds to the objectives used in the experiments presented later in this thesis.

Uniform dose. The uniform dose objective is a convex and smooth function that puts a quadratic penalty on the deviation of dose from a prescribed dose level $d_p \in \mathbb{R}$

$$f_{\text{uniform}}(d) = \sum_{i\in\mathcal{R}} r_i(d_i - d_p)^2. \tag{3.3}$$

Minimum and maximum dose. The minimum and maximum dose objectives also put quadratic penalties on the deviation of dose from a prescribed dose level. However, the minimum dose objective $f_{\min}$ penalizes only *negative* deviations from $d_{\min} \in \mathbb{R}$, whereas the maximum dose objective $f_{\max}$ penalizes only *positive* deviations from $d_{\max} \in \mathbb{R}$. The functions are expressed as

$$f_{\min}(d) = \sum_{i\in\mathcal{R}} r_i((d_{\min} - d_i)_+)^2 \tag{3.4}$$

$$f_{\max}(d) = \sum_{i\in\mathcal{R}} r_i((d_i - d_{\max})_+)^2 \tag{3.5}$$

where the positive part operation is defined as $(\cdot)_+ = \max\{\cdot, 0\}$. Both functions are convex and continuously differentiable. The two objectives can be used on the same ROI simultaneously and then collapses to the uniform dose objective if the two objectives are given equal weights and $d_{\min} = d_{\max}$.

Dose falloff. To encourage the optimizer to find a dose that confines well to the target volume, one might add a *dose falloff* objective. This objective is similar to the maximum dose objective, but replaces $d_{\max}$ with a spatially varying function $\Delta(\delta_i)$, where $\delta_i \in \mathbb{R}$ is defined as the shortest distance from voxel i to the target volume. In RayStation[3], the dose falloff objective is defined as

$$f_{\text{falloff}}(d) = \sum_{i\in\mathcal{R}} r_i \left(\left(\frac{d_i - \Delta(\delta_i)}{d_{\text{high}}} \right)_+ \right)^2 \tag{3.6}$$

where the dose falloff function depends on a few parameters. We note that the falloff objective is convex and continuously differentiable in the dose. The full expression of $\Delta(\delta)$ is given by

$$\Delta(\delta) = \begin{cases} \infty & \delta = 0 \\ d_{\text{low}} + (d_{\text{high}} - d_{\text{low}})(1 - \delta/\delta_{\text{low}}) & \delta \in (0, \delta_{\text{low}}) \\ d_{\text{low}} & \delta \geq \delta_{\text{low}} \end{cases} \tag{3.7}$$

where the dose level d_{high} reflects the desired maximum dose near the target, and d_{low} the desired maximum dose at equal or greater distances than δ_{low} [Bokrantz and Fredriksson, 2014].

Constraints. All objective functions introduced above may also be used to define constraints. However, in clinical treatment planning, objectives are often in direct conflict with each other and hard constraints are used with caution. For example, sufficient target coverage might in some situations be impossible to achieve without simultaneously irradiating an adjacent OAR more than desired. Imposing hard constraint in these situations limits the set of feasible solutions too much and it is often necessary to reformulate such hard constraints as objectives instead [Paganetti, 2018]. For example, a hard box-constraint on the dose in each voxel of a ROI

$$d_{\min} \leq d_i \leq d_{\max}, \quad \forall i \in \mathcal{R} \tag{3.8}$$

[3] A commercial TPS developed by RaySearch Laboratories AB.

might better be replaced by a combination of minimum dose (3.4) and maximum dose (3.5) objectives.

In terms of constraints, the full optimization problem also includes some domain constraints due to limitations of the treatment machine. For example, negative MUs are not physically possible. In addition, the treatment planner might specify linear constraints on x to limit, *e.g.*, the spot MU per fraction. Together, these constraints are expressed as[4] $Kx \leq m$ where $K \in \mathbb{R}^{N \times N}$ and $m \in \mathbb{R}^N$. Assuming non-negative MU is the only hard machine constraint, the physical constraints reduce to $x \geq 0$.

Nominal optimization problem

We first formulate the optimization problem ignoring any model uncertainties.

MCO formulation. The nominal treatment planning problem can be formulated as a constrained multi-criteria optimization (MCO) problem

$$\begin{aligned} \underset{x \in \mathbb{R}^N}{\text{minimize}} \quad & \{f_i(d)\}_{i \in \mathcal{O}} \\ \text{subject to} \quad & d = Dx, && \text{(dose deposition)} \\ & c_j(d) \leq 0, \quad j \in \mathcal{C}, && \text{(dose constraints)} \\ & Kx \leq m && \text{(physical constraints)} \end{aligned} \tag{3.9}$$

where the set of objectives $\{f_i\}_{i \in \mathcal{O}}$ are functions of the dose d, which is linear in the optimization variables x and constrained by a set of dose constraints $\{c_j\}_{j \in \mathcal{C}}$. The physical constraints $Kx \leq m$ further limit the feasible set.

Weighted sum formulation. There exist several approaches to handle the trade-offs inherent to MCO problems [Bokrantz, 2013]. One intuitive approach is to assign *a priori* weights $w_i \geq 0$ to the individual objectives f_i and define a new objective as their weighted sum

$$f = \sum_{i \in \mathcal{O}} w_i f_i \tag{3.10}$$

which reduces the MCO problem (3.9) to the single-criteria optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^N}{\text{minimize}} \quad & f(d) \\ \text{subject to} \quad & d = Dx, && \text{(dose deposition)} \\ & c_j(d) \leq 0, \quad j \in \mathcal{C}, && \text{(dose constraints)} \\ & Kx \leq m && \text{(physical constraints)} \end{aligned} \tag{3.11}$$

A second approach is to optimize the objectives $\{f_i\}_{i \in \mathcal{O}}$ individually and instead do some trade-off between the individual optima $\{x_i^*\}_{i \in \mathcal{O}}$ in the solution domain. For simplicity, the weighted sum formulation is used in this thesis.

Optimization under uncertainty

As always when modeling real-life with mathematics, the model is an imperfect representation of reality. Such uncertainty is often accounted for by introducing safety margins in the calculations.

The nominal optimization problem implicitly assumes zero uncertainty in the delivered dose given a fix set of treatment parameters. This assumption is unfortunately unreasonable in practice. The dose calculation is based on models of the patient, the treatment machine, and of the physical interactions involved–models that never correspond exactly to reality. Optimal solutions to the nominal optimization problem may thus result in sub-optimal treatments in practice.

To improve the robustness of optimized treatment plans, efforts have been made to include uncertainty in the optimization. The idea is to ensure good plan quality even if the model deviates slightly from reality. The interested reader may refer to [Unkelbach et al., 2018] for a review of robust planning in radiotherapy.

[4] The notation $Kx \leq m$ is chosen over of the more common $Ax \leq b$ to avoid confusion with beam indices and dose-mapping matrices introduced later.

Sources of uncertainty. In the nominal optimization problem (3.11), there are three sources of uncertainty [Unkelbach et al., 2018]:

1. Uncertainty in the dose deposition matrix D caused by
   a) any intrinsic uncertainties in the dose calculation algorithm.
   b) uncertainties in the mathematical formulation of the radiation transport problem that the dose calculation algorithm is used to solve, *i.e.*, uncertainties in the model inputs to the dose calculation. This includes patient model uncertainties such as the exact patient treatment position, the exact shape and location of organs, and the mass density and material composition of tissues.
2. Uncertainty in the machine output, *i.e.*, in the spot MUs x.
3. Uncertainty in the mathematical definition of an ideal dose distribution and the proper choice of objective function f to match.

This thesis deals with uncertainties of type 1. Uncertainties of type 2 can be considered small thanks to verification measurements during machine QA [Unkelbach et al., 2018]. Uncertainties of type 3 are out of scope since they do not affect dose calculation.

The dose as a random variable. The dose d is modeled as a random variable to acknowledge uncertainty in the dose deposition matrix. To make optimization tractable, the underlying (type 1b) uncertainties are discretized into a finite set of *scenarios*, each corresponding to a realization of the uncertainties, indexed by the set $\mathcal{S} = 1, \ldots, |\mathcal{S}|$ [Fredriksson, 2013]. The dose d_s in a specific scenario $s \in \mathcal{S}$ is then modeled as a deterministic function of the optimization variables x

$$d_s = d(x; s) = D_s x \tag{3.12}$$

where (3.2) implies the second equality. An important observation is that new dose deposition matrices D_s must be calculated for each added scenario. This imposes computational limits on the number of scenarios that can be considered.

Analytical probabilistic modeling (APM) is an alternative way of including uncertainty in the calculations without sampling the underlying uncertainties into scenarios [Bangert et al., 2013]. With APM, it is possible to calculate closed-form expressions of the mean and covariance of the dose distribution which can then be used in probabilistic treatment planning optimization or evaluation. This can be more computationally efficient than sampling based approaches [Wahl et al., 2017].

One major drawback of APM is that it relies on a functional approximation of the dose calculation algorithm. To calculate the m-th moment of an element D_{ij} in the dose deposition matrix D, one must evaluate the integral

$$\mathbb{E}[D_{ij}^m] = \int D_{ij}(y)^m p(y)\, dy \tag{3.13}$$

where y denotes the underlying uncertain parameters and $p(y)$ their probability density. Finding an analytical expression for $D_{ij}(y)$ is far from trivial. In [Bangert et al., 2013], such expressions are derived for setup and particle range uncertainties in a pencil beam dose calculation. An extension to deforming geometries (*e.g.*, anatomical deformations) is also claimed to be analytically feasible. However, APM does not generalize easily to non-pencil beam algorithms such as Monte Carlo-algorithms, and the extension to anatomical deformations remains an open question [Unkelbach et al., 2018].

Scenarios. In this thesis, scenarios are defined as sampled realizations of the (type 1b) uncertainties that affect the inputs to dose calculation. Each scenario $s \in \mathcal{S}$ is assigned a probability $\pi_s \geq 0$ such that $\sum_{s \in \mathcal{S}} \pi_s = 1$. The uncertainties are mainly related to two kinds of errors in the patient model:

- Errors in the **patient setup** relative to the beam. In combination, these correspond to a rigid geometrical transformation of the patient parameterized by *patient setup shifts* and *patient setup rotations*.
- Errors in the **patient image** and material properties derived thereof. Such errors include organ motion and patient weight loss, but also errors in the CT-HU to RSP conversion which affects the calculated beam range. The scenarios considered in this thesis are: *anatomical changes/organ motion* (new patient image), and systematic *density shifts* (uniform scaling of the density).

The scenario with no model errors is referred to as the *nominal scenario* with index $s_0 \in \mathcal{S}$.

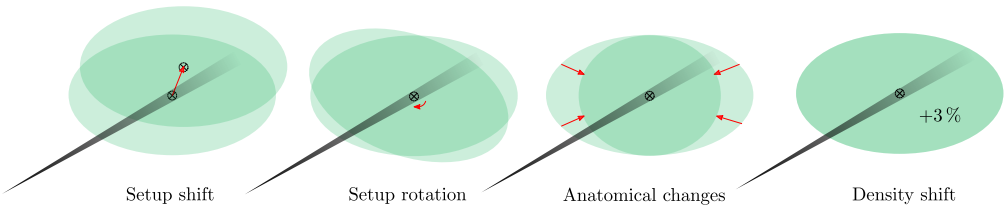


Figure 3.3 Illustration of four types of patient model errors as deviations from a nominal scenario. In a patient setup shift or setup rotation scenario, the patient moves relative to the beam. Anatomical changes are errors in the patient model that cannot be explained by a setup shift or setup rotation. Density shifts involve a uniform scaling of the density of the patient.

Treatment course. During the course of a fractionated treatment, many different error scenarios may occur. When optimizing the full treatment, it is thus relevant to consider how errors correlate.

Some errors are *interfractional* in nature. For example, it might be reasonable that the patient is positioned 0.5 cm to the left on one day and 0.5 cm to the right on another. Other errors are better modeled as *systematic* errors across all fractions. There might also be *intrafractional* errors, for example caused by breathing motion.

Distinguishing between systematic, interfractional, and intrafractional uncertainties is important when designing an objective function that takes the uncertainties into account. However, it only affects the probability of the different scenarios to occur and co-occur. Calculation of the dose deposition matrices D_s is not affected. Since calculation of D_s is the main concern in this thesis, only systematic errors are considered from now on.

Objective function. When the dose is a random variable, the objective f in (3.11) also becomes a random variable. Before optimizing, we must then apply some cost functional on f to retrieve a real-valued objective. One approach is to consider variations of minimax stochastic programming such as expected value, worst case, or conditional value at risk (CVaR) optimization [Fredriksson, 2012]. With sampled scenarios, expected value optimization becomes

$$\underset{x}{\text{minimize}} \quad \sum_{s\in\mathcal{S}} \pi_s f(d_s(x)) \tag{3.14}$$

where outcomes are weighted by their probability to occur. Worst case optimization

$$\underset{x}{\text{minimize}} \quad \max_{s\in\mathcal{S}} f(d_s(x)) \tag{3.15}$$

ignores the probabilities π_s and optimizes the outcome in the worst scenario. CVaR optimization interpolates between the two approaches and is interpreted as optimizing the expected outcome in a given fraction $0 < \alpha \leq 1$ of the worst scenarios. For details, see [Fredriksson, 2012].

Robust optimization formulation. Regardless of the choice of cost functional, the new objective is a real-valued function of the scenario doses. For generality, we denote the accumulated objective $g(\{d_s\}_{s\in\mathcal{S}})$. The optimization problem becomes

$$\begin{aligned}
&\underset{x\in\mathbb{R}^N}{\text{minimize}} && g(\{d_s\}_{s\in\mathcal{S}}) && \\
&\text{subject to} && d_s = D_s x, \quad s\in\mathcal{S}, && \text{(scenario dose deposition)} \\
&&& c_j(d_s) \leq 0, \quad (j,s)\in\mathcal{C}\times\mathcal{S}, && \text{(robust dose constraints)} \\
&&& Kx \leq m && \text{(physical constraints)}
\end{aligned} \tag{3.16}$$

where $\mathcal{S}$ enumerates the scenarios and $\mathcal{C}$ enumerates the dose constraints.

Robust optimization in RayStation

Robust optimization formulation. In RayStation, robustness may be required only for some objectives and constraints. Recalling the weighted sum formulation (3.11), we partition the set of objectives into robust objectives $\{f_i\}_{i\in\mathcal{O}_r}$ and non-robust objectives $\{f_i\}_{i\in\mathcal{O}_{nr}}$ such that $\mathcal{O}_r \cup \mathcal{O}_{nr} = \mathcal{O}$ and $\mathcal{O}_r \cap \mathcal{O}_{nr} = \emptyset$. The index set $\mathcal{C}$ for the dose-constraints is similarly partitioned into $\mathcal{C}_r$ and $\mathcal{C}_{nr}$. For the robust objectives, the accumulated objective is formulated as in worst case optimization (3.15)

but with the maximum operator approximated by a weighted power mean operator in order to make the objective smooth and more suitable for continuous optimization. The full robust optimization problem takes the form

$$\begin{aligned}
\underset{x\in\mathbb{R}^N}{\text{minimize}} \quad & \sum_{i\in\mathcal{O}_{nr}} w_i f_i(d_{s_0}) + \left(\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\Big(\sum_{i\in\mathcal{O}_r} w_i f_i(d_s)\Big)^p\right)^{1/p} \\
\text{subject to} \quad & d_s = D_s x, \quad s\in\mathcal{S}, && \text{(scenario dose deposition)} \\
& c_j(d_{s_0}) \le 0, \quad j\in\mathcal{C}_{nr}, && \text{(non-robust dose constraints)} \\
& c_j(d_s) \le 0, \quad (j,s)\in\mathcal{C}_r\times\mathcal{S}, && \text{(robust dose constraints)} \\
& Kx \le m && \text{(physical constraints)}
\end{aligned} \tag{3.17}$$

where $p = 8$ is empirically chosen. Recalling earlier given definitions, w_i are positive importance weights of the dose-based objectives f_i. Such objectives are described in Section 3.3 and are selected by the treatment planner. This also applies to the dose-constraint functions c_j. The scenario doses d_s for scenarios $s\in\mathcal{S}$ are connected to the optimization variables x via the dose deposition matrices D_s as described in (3.12). The final linear constraints model physical limitations of the treatment machine together with limits on x specified by the treatment planner.

Optimization algorithm. The optimization algorithm in RayStation is inspired by, *e.g.*, [Gill et al., 2005] and [Philip et al., 2001] and implemented as a general-purpose *sequential quadratic programming* (SQP) algorithm using *Broyden–Fletcher–Goldfarb–Shanno* (BFGS) updates. The algorithm is line-search based and iteratively solves problems such as (3.17), in each iteration using first and (approximations of) second derivatives of the objective. A complete description of the algorithm is out of scope and the interested reader may refer to, *e.g*, [Nocedal and Wright, 2006] for a comprehensive discussion of continuous optimization.

Of relevance to this thesis is that the algorithm is gradient based. The objectives $\{f_i\}_{i\in\mathcal{O}}$ are chosen to be differentiable with respect to the dose, and the robust optimization objective in (3.17) to be differentiable w.r.t. f_i. We may then use the chain rule to compute gradients. In this process, the linear dose deposition relation (3.12) allows for gradients of f_i w.r.t. the scenario dose d_s to be transformed to gradients w.r.t. the optimization variables x. Such transformations take the form

$$\nabla_x f_i(d_s) = \nabla_x f_i(D_s x) = D_s^T \nabla_d f_i(d_s) \tag{3.18}$$

where ∇_z denotes the gradient w.r.t. z. The details of the gradient computation is relevant for the implementation of the dose mapping algorithm as described further in Section 5.2.

3.4 Summary

In radiotherapy treatment planning there is a fundamental trade-off between sufficiently irradiating the tumor cells and sparing healthy tissue from radiation damage. Finding an optimal treatment is challenging. The process involves adapting the treatment to a *model* of the patient by adjusting any available treatment parameters. To this end, the treatment planner may use a TPS with mathematical optimization functionality.

Due to small errors in the patient model, the dose actually delivered to the patient might deviate somewhat from the planned dose. To compensate for this, a good treatment plan must be robust to small model errors. Without such robustness, a seemingly optimal treatment might be sub-optimal for the patient in practice.

In scenario-based robust optimization, the treatment is optimized to be as good as possible in the worst of a predefined set of scenarios. This increases the robustness of the treatment plan but comes at a computational cost. For each added scenario, a new dose deposition matrix must be calculated. In Chapter 4, different algorithms for scenario dose calculation are described, including the dose-mapping algorithm central to this thesis. The purpose of the experiments that follow is to investigate if such an algorithm can be used to speed up the scenario dose calculations, and thus also the robust optimization procedure that solves (3.17).

4

Dose calculation algorithms

4.1 Introduction

The dose deposition matrix $D \in \mathbb{R}^{|\mathcal{V}| \times N}$ is introduced in Section 3.2 where the general mathematical framework for optimization in proton PBS treatment planning is discussed. The dose deposition matrix models the linear relationship (3.2) between physical dose in the patient $d \in \mathbb{R}^{|\mathcal{V}|}$ and spot weights $x \in \mathbb{R}^N$. An element D_{ij} in the matrix can be interpreted as the dose deposited to voxel i per unit MU of the scanned proton beam while set to the position and energy corresponding to spot j.

The scenario dose deposition matrix D_s is introduced in (3.12) in the context of robust optimization, in which a predefined set $\mathcal{S}$ of scenarios is considered. A scenario is defined as a realization of some uncertainties that affect the *input* to the dose calculation algorithm. Thus, the elements of D might depend on the scenario and a separate matrix D_s is used for each scenario $s \in \mathcal{S}$.

In this chapter, accuracy of dose calculation algorithms is discussed. Accuracy in this context does not refer to the model errors (such as errors in the patient setup or in the HU to mass density conversion) reflected by the scenarios $\mathcal{S}$. These scenarios correspond to deviations from reality in the mathematical formulation of the radiation transport *problem* that the dose calculation algorithms are used to solve. Instead, accuracy in this context refers to the accuracy of the *solution*, assuming that the formulation of the radiation transport problem corresponds exactly to reality.

4.2 General algorithms

Dose calculation algorithms in PBS proton therapy can be divided into Monte Carlo (MC) and analytical algorithms, where the analytical algorithms include pencil beam (PB) and ray casting (RC) models. In this section, such general-purpose codes are described. These general algorithms can be used to calculate scenario doses D_s for any scenario $s \in \mathcal{S}$ of interest. Section 4.3 deals with approximate methods of calculating scenario doses without surrendering to a full recalculation for each scenario. These use additional approximations and may only handle specific types of scenarios. For a more comprehensive treatment of MC and PB algorithms, please refer to Chapter 8 and Chapter 14 in [Paganetti, 2018].

Monte Carlo

Monte Carlo (MC) simulation is the gold standard for accurate dose calculations in proton therapy since it stays true to our best knowledge of the relevant physical interactions at a microscopic scale [Paganetti, 2018]. MC is a statistical approach to dose calculation that involves simulating a large number of independent particle trajectories in a beam before averaging and scaling the result to some unit fluence. Accurate simulations require sufficiently accurate models of physical interactions that occur at a particle scale. Such microscopic physics processes are generally well understood, giving MC simulations a solid foundation to build upon.

Algorithm. To calculate the dose delivered by a beam, MC dose calculations use probabilistic models of the physical interactions that individual particles may experience during transport. In a run of a MC simulation, a large number of particle *histories* is generated by repeatedly initializing individual protons at the beam source, then simulating and tracking their movement through a well defined geometry while sampling physical interactions according to their probability to occur. The particle tracking is done one small step at a time, see Figure 4.1 for an illustration.

Some interactions create secondary particles. These are tracked as part of the same history as the primary particle. All particles are tracked until they leave the region of interest or until their energy

fall below some threshold value. Sometimes, secondary particles are not tracked at all to reduce the computational cost, and their energy may then be deposited locally [Paganetti, 2018].

In PBS, each spot corresponds to a unique beam direction and energy. Thus, the distribution of initialized protons at the beam source depends on the spot and a separate MC simulation is be performed for each spot, *i.e.*, for each column in D.

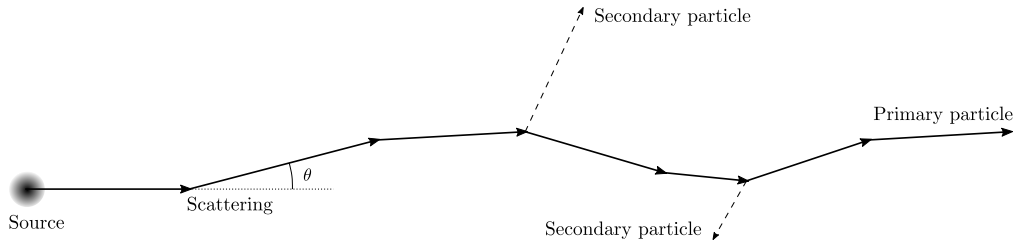


Figure 4.1 Illustration of a particle history in a MC simulation. Each primary particle originates from the source and undergoes a series of stochastic events such as scattering or creation of secondary particles. The collected history of all events for one primary particle is called a *particle history*.

Statistical motivation. A back-of-the-envelope calculation puts the number of protons per spot at around 10^5–10^9 [Kooy and Grassberger, 2015]. For the spot j, the dose deposited to voxel i by the k:th proton (and corresponding secondary particles) is a random variable X_{ij}^k. Its distribution depends on the exact geometry and material composition of anything in and between the beam source and the voxel as well as the spot-dependent distribution of primary particles at initialization. The total spot dose to voxel i for some unit MU (corresponding to K protons) is the sum of all individual contributions and may be expressed as

$$D_{ij} = \sum_{k=1}^{K} X_{ij}^k. \tag{4.1}$$

Given that the particles are reasonably independent, the central limit theorem states that D_{ij} is approximately normally distributed. The mean value of D_{ij} may then be estimated by averaging over M simulated particle histories. A nice property of such a mean estimate is that its standard deviation is approximately proportional to $1/\sqrt{M}$ for large M. This is the foundation of the Monte Carlo approach to dose calculation.

Accuracy and performance. The accuracy of a MC dose calculation depends on the accuracy of the underlying physical models used in simulation. Different implementations might use different approximations and parameters which affects the calculated dose. Such implementation choices include the choice of secondary particles to track, energy thresholds, step length choices, or cross-section data determining probabilities for nuclear interactions.

Historically, MC simulations in treatment planning has been performed by general purpose research codes not optimized specifically for treatment planning. These codes are quite slow, limiting the clinical use. Recently, however, MC codes optimized for treatment planning have become available in some commercial TPS. For a discussion on such codes and their accuracy compared to pencil beam algorithms, see [Saini et al., 2018].

An important performance aspect is that MC simulations may be run in parallel on a multi-core CPU or GPU. This significantly decreases calculation times [Paganetti, 2018].

Analytical algorithms

Analytical dose calculation algorithms, in contrast to the stochastic Monte Carlo simulations, are deterministic and does not depend on any (pseudo)random numbers. Pencil beam (PB) algorithm, constitute the most common class of analytical dose calculation algorithms. However, the class of analytical algorithms is wider and contains, *e.g.*, ray casting (RC) algorithms as well.

Neither pencil beam nor ray casting algorithms are used in the experiments in this thesis. However, both share some interesting similarities with the range-corrected proton dose approximation method [Park et al., 2012] which is of central interest.

Pencil beam algorithms. The central idea in pencil beam (PB) algorithms is to calculate the dose as a convolution between the proton fluence and a dose kernel [Hogstrom et al., 1981], [Petti, 1992].

The dose kernel represents the dose deposited by an infinitesimally thin proton beam in a homogeneous volume of water. Depending on the choice of approximations and the physical effects taken into account, the kernel may be an expression of varying complexity.

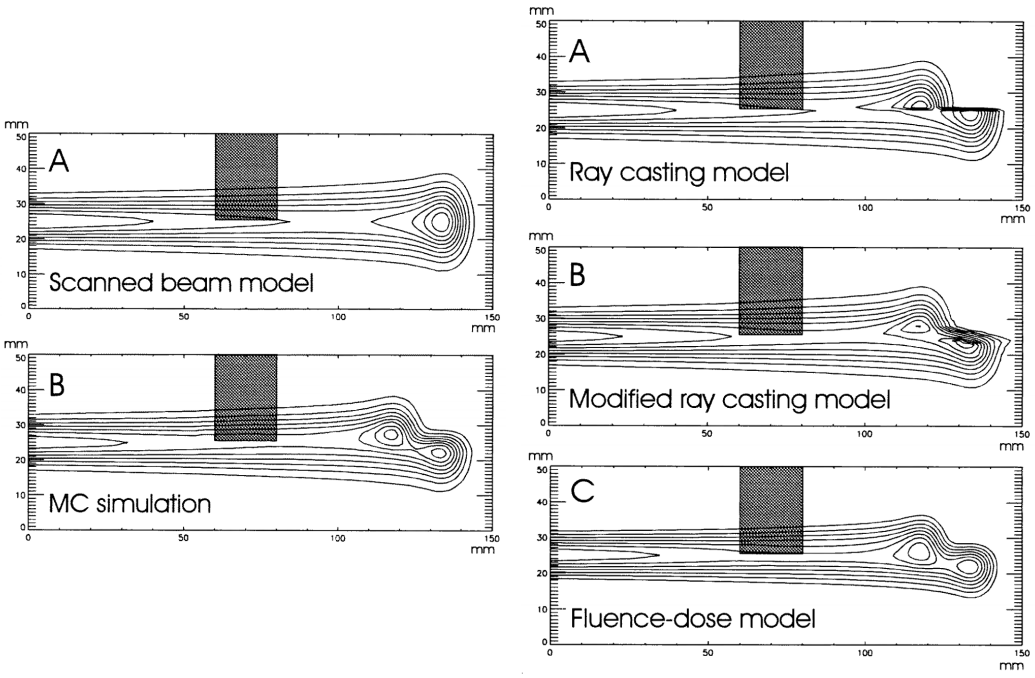


Figure 4.2 Figure 1 (left) and Figure 4 (right) reproduced with permission[5] from [Schaffner et al., 1999]. The spot doses are calculated in a simple geometry consisting of a 2 cm thick box of 'cortical bone' (stopping power 1.8 relative to water) surrounded by water. The heterogeneity introduced by the bone affects the spot dose in a way most accurately represented by the MC simulation. The 'scanned beam model' only accounts for the material at the beam central axis, completely ignoring the bone in this case. The 'ray casting model' scales the depth-dose curves along (parallel) rays from the source resulting in a sharp edge in the shadow behind the bone. A modified RC model may attempt to smooth this sharp shadow. The 'fluence-dose model' in this figure corresponds to a PB algorithm.

Ray casting algorithms. The central idea in ray casting (RC) algorithms is to transform a model of the spot dose in water to a corresponding spot dose in a heterogeneous medium. This is achieved by scaling the depth-dose distributions of the spot in water according to WEPL along rays from the source through the medium [Schaffner et al., 1999]. RC algorithms thus uses models of full spot doses, as compared to doses from infinitesimally thin pencil beams. The spot-dose model could contain a look-up table of depth-dose curves and depth-width relations in water for spots of varying nominal energy.

Figure 4.2 shows a comparison between spot doses calculated by different algorithms in a simple geometry. The figure seems to indicate that the PB 'fluence-dose model' gives results more in accordance with MC than the RC models. However, in a more detailed analysis that authors conclude that RC and PB have different strengths. RC corresponds to placing all inhomogeneities (differences to water) right in front of the dose calculation voxel, and thus best models cases with inhomogeneities close to the Bragg peak. PB, on the other hand, corresponds to placing all such inhomogeneities right at the patient surface. Thus, PB better models the situation when inhomogeneities are close to the surface [Schaffner et al., 1999].

Accuracy and performance. As previously stated, MC is the gold standard for accurate dose calculations in proton therapy [Paganetti, 2018]. The most pronounced discrepancies between MC and analytical algorithms are expected in complex geometries since analytical algorithms typically model the effect of lateral inhomogeneities incorrectly. Often, analytical algorithms disregard the depth position of such inhomogeneities and only compensate for the total WEPL [Paganetti, 2012].

[5] © Institute of Physics and Engineering in Medicine. Reproduced by permission of IOP Publishing. All rights reserved.

Several studies have indicated the importance of using MC for certain sites, see, *e.g.*, [Taylor et al., 2017], [Yepes et al., 2018], and [Saini et al., 2017], who also report similar dose calculation speed for analytical dose calculations and MC calculations at 2 % statistical uncertainty.

Claims that analytical algorithms are unsuitable are however disputed by [Winterhalter et al., 2019], who report no substantial differences in clinically relevant dosimetric indices between analytical (RC, PB) and MC algorithms but significant differences in computational time. These authors recommend using an RC algorithm for fast optimization, combined with a subsequent MC calculation for quality assurance.

4.3 Scenario dose approximations

Any general dose calculation algorithm described in Section 4.2 can be used to calculate the scenario dose deposition matrices needed for robust optimization. However, such full recalculations can be prohibitively slow if many scenarios are considered. This section covers some scenario dose approximation methods, following Section 5.5 in [Unkelbach et al., 2018], intended to speed up calculations by using approximations tailored to specific types of scenarios.

Static dose cloud

The static dose cloud approximation is used extensively in photon therapy but is unsuitable for proton therapy. The approximation entails that the dose distribution in room coordinates is unaffected by any changes in the patient positioning or anatomy. For a setup shift of the patient, this means that the dose cloud remains stationary in room coordinates, or in patient coordinates that the dose cloud is rigidly shifted in opposite direction of the setup shift. For a patient shift $(\Delta x, \Delta y, \Delta z)$ in patient coordinates, this can be expressed as

$$d^s_{x,y,z} = d^{s_0}_{x+\Delta x, y+\Delta y, z+\Delta z} \tag{4.2}$$

where d^s is the scenario dose and d^{s_0} the nominal dose.

Spot weight shifting

Spot weight shifting, referred to as *beamlet shifting* in [Unkelbach et al., 2018], is implemented in RayStation and is used in the experiment section of this thesis.

Algorithm. In spot weight shifting, patient shifts are approximated as shifts of the spot weights in the isocenter plane. A patient shift of 0.5 cm in the negative x-direction (BEV coordinates) is implemented as an equally large but positive x-shift of the spot weights in the isocenter plane. The idea is that spot doses at the new positions can be *interpolated* based on already computed spot doses instead of being fully recalculated. See Figure 4.3 for an illustration. Range error scenarios are interpreted as systematic errors in the primary proton energy. Dose calculation in such scenarios is implemented as interpolation of spot weights *between* energy layers, as compared to *within* each energy layer for setup shifts [Unkelbach et al., 2009].

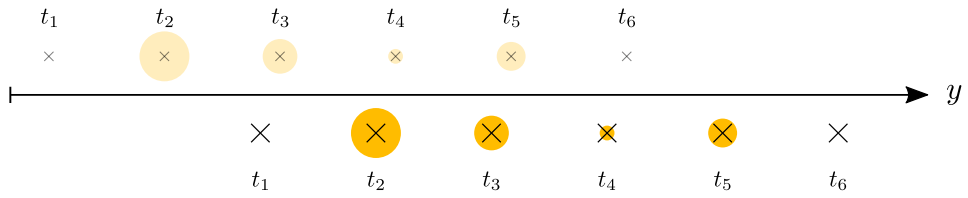


Figure 4.3 A 1D example of spot weight shifting for patient setup shifts. Assume that the treatment plan contains six spots $t_1, \ldots, t_6$ all in the same energy layer. In the nominal scenario, these spots (upper row) are lined up along some line y in the isocenter plane. In a setup shift scenario s where the patient moves in the negative y-direction, the spots move accordingly along the y-line as illustrated by the bottom row of spots. Since there are six spots, D_s has six columns. To calculate the second column in D_s, the spot weight shifting method *interpolates* the already calculated third and fourth column of the nominal dose deposition matrix D_{s_0}. The situation is more complicated for spots t_5 and t_6 which end up outside the grid of nominal spots. To also handle these, the spot weight shifting must calculate additional, *virtual* spots t_7, t_8 in the nominal scenario. These virtual spots are only used for interpolation and are not part of the actual treatment plan.

Advantages and limitations. The advantages of spot weight shifting include decreased computational times since fewer accurate spot doses must be calculated. Additionally, by reusing spot doses across scenarios, the memory footprint is decreased. However, interpolating spot doses is not equivalent to actually computing accurate spot doses at the new positions which leads to decreased accuracy. A further disadvantage is that edge spots require special attention. Both when interpolating within an energy layer and between energy layers, the edge spots might be shifted outside the initial spot grid which necessitates that additional, *virtual*, spots are added before interpolation can take place. For each virtual spot added, a new accurate spot dose computation must be performed.

An additional disadvantage is that spot weight shifting is not easily extended to other scenarios such as anatomical changes or rotations of the patient.

Range-corrected approximation method

The paper *Fast range-corrected proton dose approximation method using prior dose distribution* [Park et al., 2012] describes a scenario dose approximation method which is of central interest in this thesis. In the experiments sections, a modified version is implemented (Chapter 5) and evaluated (Chapter 6-8).

This range-corrected method is somewhat related to the static dose cloud approximation. Both methods map the entire (beam) dose distribution instead of calculating the dose spot-by-spot. However, the range-corrected method is a more refined method that also *deforms* the dose cloud, accounting for range-affecting changes along the beam path. The original algorithm deals with scenarios of setup shifts and anatomical changes.

Motivation. The algorithm is motivated by the "heuristic observation that perturbation of the WEPL along the beam path pulls the dose profile proximally or distally from the nominal setting without significantly changing its overall shape", resulting in the assumption that "the scattered dose from small changes in the patient's anatomy will remain constant and the dose effect is only caused by the change in the cumulative WEPL at each voxel along the beam path" [Park et al., 2012].

Published algorithm. The range-corrected method maps the dose beam-by-beam from the reference setting (nominal scenario) to a scenario setting parameterized by some combination of a setup shift $(\Delta x, \Delta y, \Delta z)$ and anatomical changes represented by a new patient image CT_s. The scenario patient image is aligned to the nominal image CT_{s_0} via deformable registration. The sign of the setup shift depends on whether the coordinate system is fixed relative the patient or relative the beam. Here, we reformulate the patient setup shift as a shift of the beam in a patient-fixed coordinate system where the z-axis is aligned with the beam central axis.

A central concept in the algorithm is the *water-equivalent path length* (WEPL), which for some curve l in the physical space equals the line integral of the *relative stopping power* (RSP) along the curve

$$\mathrm{WEPL} = \int_l \mathrm{RSP}(x,y,z)\, dl \tag{4.3}$$

as previously stated in (2.6). The RSP field over the patient is a function of the patient image data and is denoted $\mathrm{RSP}_{s_0}(x,y,z)$ in the nominal case, and $\mathrm{RSP}_s(x,y,z)$ in the scenario.

The central idea of the algorithm is that the nominal dose located at (x,y,z) in BEV coordinates should, under the scenario, be mapped to somewhere on the line

$$(x+\Delta x, y+\Delta y, z'), \quad z' \in \mathbb{R} \tag{4.4}$$

to account for the setup shift. The choice of z' constitutes the core part of the method and is what differentiates it from the static dose cloud method. Its value is determined by solving for z' in the equation

$$\int_{-\mathrm{VSAD}}^{z} \mathrm{RSP}_{s_0}(x,y,\hat{z})\, d\hat{z} = \int_{-\mathrm{VSAD}+\Delta z}^{z'} \mathrm{RSP}_s(x+\Delta x, y+\Delta y, \hat{z})\, d\hat{z} \tag{4.5}$$

which via (4.3) can be interpreted as finding the point z' along the new beam path (4.4) that is at equal WEPL from the source as z is in the nominal scenario. The lower integration limit corresponds to the z-component of the virtual source position in the BEV coordinate system, VSAD being short for *virtual source-axis distance*.

Under this approximation, dose is mapped along parallel beam lines. In reality, the beam is always diverging, and the authors include the inverse square factor

$$\mathrm{INV}(z', z, \Delta z) = \left(\frac{z+\mathrm{VSAD}}{z'+\mathrm{VSAD}-\Delta z} \right)^2 \tag{4.6}$$

to account for this fact. The final expression for the beam dose mapping is then

$$d^{s}_{x+\Delta x,y+\Delta y,z'} = d^{s_0}_{x,y,z} \times \mathrm{INV}(z', z, \Delta z) \tag{4.7}$$

where d^s is the beam dose in the scenario s and d^{s_0} the corresponding nominal dose.

Accuracy and performance. The range-corrected approximation method described in this section was developed with tasks in mind where accuracy of the computed dose may be sacrificed for computational efficiency. Such tasks include robust evaluation and robust optimization which requires a large number of dose calculations [Park et al., 2012].

The authors compare the dose calculation accuracy to the static dose cloud approximation and a clinically commissioned PB algorithm. They conclude that the static dose cloud approximation is inadequate for proton plans, whereas their range corrected method is sufficiently accurate for the intended purposes. The authors also conclude that the main dosimetric errors as compared to a PB-computed dose occur in regions where modelling of lateral scattering is important, and suggest further assessment of the method by comparison with an MC algorithm.

5

Implementation

5.1 Monte Carlo and spot weight shifting

Both the MC and spot weight shifting methods are implemented in the commercial TPS RayStation 11A and were used in the experiments without modification.

The spot weight shifting method in theory only handles setup shift and density shift scenarios. However, for scenarios that also include anatomical changes, it is implemented to run one accurate MC simulation for each new patient image before proceeding with the spot weight shifting.

5.2 A range-corrected dose-mapping method

The range-corrected method in [Park et al., 2012] is described in Section 4.3. The article describes the idea of the algorithm but does not provide many details about the implementation. To run the experiments in this thesis, an extended and modified version of the method was implemented in C++ within the CoreORBIT framework of RayStation. This section describes the implementation and how it differs from the published algorithm. A proper description of the implementation unfortunately requires many of the symbols introduced in Section 3.2, of which a summary is provided in Table 5.1.

Table 5.1 Summary of symbols. See Chapter 3 for details.

Symbol	Interpretation
$b \in \mathcal{B}$	Beam index
$l \in \mathcal{L}_b$	Index of spot energy layer for beam b
$t \in \mathcal{T}_{bl}$	Index of spot in energy layer l
$s \in \mathcal{S}$	Scenario index
$s_0 \in \mathcal{S}$	Nominal scenario index
$i \in \mathcal{V}$	Voxel index in the *scenario* dose vector
$j \in \mathcal{V}$	Voxel index in the *nominal* dose vector
$d \in \mathbb{R}^{\lvert\mathcal{V}\rvert}$	Dose vector
$x \in \mathbb{R}^N$	Optimisation variables
$D \in \mathbb{R}^{\lvert\mathcal{V}\rvert \times N}$	Dose deposition matrix
$A \in \mathbb{R}^{\lvert\mathcal{V}\rvert \times \lvert\mathcal{V}\rvert}$	Dose mapping matrix
f	Objective function
∇	Gradient operator

Summary of differences to the published method

The published method deals with scenarios of patient setup shifts and anatomical changes. Reusing ideas from the article, the implementation described here is extended to also handle patient *setup rotation* and *range error* scenarios, where the latter is implemented as systematic density shifts of the patient.

In the published algorithm, the dose is deformed along parallel rays. In the new implementation, the dose is instead deformed along diverging rays originating from the virtual source.

The article mentions potentially using the method for scenario dose calculation in robust optimization but does not report any such results. The implementation described here is used in continuous, gradient-based optimization, and is thus implemented such that the deformed dose remains differentiable with respect to the optimization variables.

Dose mapping matrices

Recalling the dose deposition relation (3.2), the dose is linear in the optimization variables x. Following the notation in Section 3.2, we now group the optimization variables by beam and define a column vector of optimization variables $x_b = (x_{blj})_{(l,t)\in\mathcal{L}_b\times\mathcal{T}_{bl}} \in \mathbb{R}^{N_b}$ for each beam $b \in \mathcal{B}$. The length of each such vector is given by N_b, calculated as in (3.1) without summing over the beams. After reordering, the dose deposition matrix can then be partitioned as $D = (D_1, \ldots, D_{|\mathcal{B}|})$ and the dose deposition relation rewritten as

$$d = \sum_{b\in\mathcal{B}} d_b = \sum_{b\in\mathcal{B}} D_b x_b \tag{5.1}$$

where $D_b \in \mathbb{R}^{|\mathcal{V}|\times N_b}$ are the beam-specific dose deposition matrices.

The range-corrected dose-mapping method deforms the dose beam-by-beam with a separate mapping for each scenario $s \in \mathcal{S}$. Each such dose-dose mapping is described by a matrix $A_{bs} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$. This allows the scenario dose deposition relation (3.12) to be expressed as

$$d_s = \sum_{b\in\mathcal{B}} A_{bs} d_{bs_0} = \sum_{b\in\mathcal{B}} A_{bs} D_{bs_0} x_b \tag{5.2}$$

where D_{bs_0} is the nominal beam-specific dose deposition matrix for beam b. From this follows that the full scenario dose deposition matrix can be expressed as

$$D_s = \begin{pmatrix} A_{1s}D_{1s_0} & \ldots & A_{|\mathcal{B}|s}D_{|\mathcal{B}|s_0} \end{pmatrix} \tag{5.3}$$

assuming that the optimization variables are ordered by beam as described above.

Gradient computations

Convenient gradient calculation is one reason for implementing the dose-mapping as a matrix multiplication. As discussed in Section 3.3, gradients with respect to the optimization variables can be calculated as transformed gradients with respect to the dose. Following (3.18), gradients with respect to x_b are calculated using the chain rule as

$$\nabla_{x_b} f(d_{bs}) = \nabla_{x_b} f(A_{bs}D_{bs_0}x_b) = D_{bs_0}^T A_{bs}^T \nabla_d f(d_{bs}) \tag{5.4}$$

for the objective f and beam $b \in \mathcal{B}$.

Storing the matrices

Each dose mapping matrix A_{sb} is of dimension $|\mathcal{V}| \times |\mathcal{V}|$, where $\mathcal{V}$ is the set of dose grid voxels. Due to the large number of voxels in the typical dose grid, these matrices are too large to be stored in a dense format. However, since the dose from a single voxel is generally mapped to a small number of voxels, A_{bs} is a sparse matrix.

The choice of sparse format affects how efficiently rows and columns can be accessed. From (5.2) it is apparent that dose computations include multiplying A_{bs} with a column vector from the right, implying that efficient *row access* is important. At the same time, gradient computations (5.4) include a similar matrix-vector multiplication with the transposed A_{bs}, implying that efficient *column access* is important.

In our current implementation, A_{bs} and A_{bs}^T are stored as two separate compressed row-storage (CRS) matrices for simplicity and speed. CRS matrices are designed for efficient row access and storing both matrices allows for equally efficient dose and gradient computations.

Interpreting the dose mapping

The dose-mapping matrix A_{bs} maps the nominal beam dose d_{bs_0} to the scenario beam dose d_{bs} according to

$$d_{bs} = A_{bs} d_{bs_0}. \tag{5.5}$$

Interpreting this relation, the scenario dose in voxel i is a linear combination of the nominal dose in voxels $j \in 1 \dots |\mathcal{V}|$ with the coefficients being stored on row i in A_{bs}. Alternatively, the nominal dose in voxel j is spread out over voxels $i \in 1 \dots |\mathcal{V}|$ with coefficients stored in column j. A non-zero element $a_{ij} = (A_{bs})_{ij}$ thus means that some dose is mapped from voxel j to voxel i. To define the dose-mapping matrix, we need to identify such *corresponding voxels* and find appropriate values of a_{ij}.

Ray tracing

The article [Park et al., 2012] describes an algorithm that loops over all points (x,y,z) in the patient image and for each such point solves for z' in equation (4.5). Solving for z' is an iterative procedure and the most time-consuming part of the algorithm. The integral equation (4.5) can be interpreted as finding points along the lines $(x,y,z), z \in \mathbb{R}$ and $(x+\Delta x, y+\Delta y, z'), z' \in \mathbb{R}$ that are at corresponding WEPL from the respective virtual source positions.

In our implementation, the depth integration is instead performed along a set of diverging rays, as illustrated in Figure 5.1. Each ray originates from the virtual beam source and are aimed at some bixel in an equally spaced bixel grid covering (with some margin) all spots in the isocenter plane. The bixel grid is illustrated in Figure 5.2. A sufficiently high bixel resolution is needed to sample each voxel of interest, but with increased resolution comes increased computational cost.

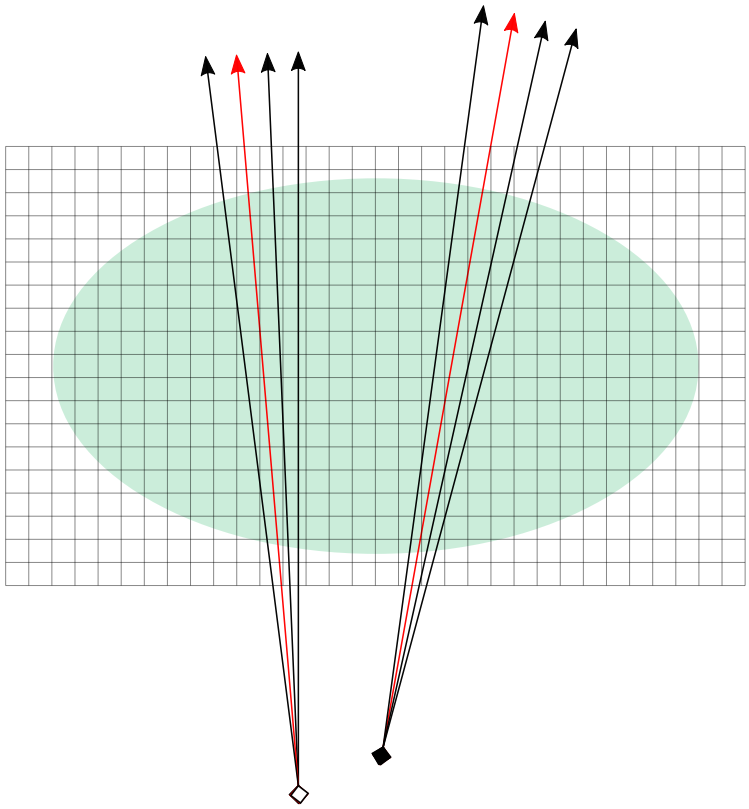


Figure 5.1 Illustration of the ray tracing. The rays originating from the white diamond follow the beam in the *nominal scenario*. In a *scenario* of patient setup shift and rotation, both the isocenter and virtual source point (black diamond) move to new positions in the patient coordinate system, causing the rays to intersect a new set of dose grid voxels. The difference between the scenarios is exaggerated for illustration. If the scenario contains a range (density) shift or anatomical changes, the material distribution over the voxels as seen by the nominal (white diamond) rays differs from the distribution seen by the scenario (black diamond) rays.

The depth integration along rays is performed once per scenario, including the nominal scenario. For each scenario, the virtual source position and beam isocenter are moved within the patient coordinate system to account for any patient setup shifts and rotations. Thus, two *corresponding rays* may intersect different voxels depending on the scenario. A nominal dose voxel j and a scenario dose voxel i are considered corresponding *in the angular sense* if they are intersected by corresponding rays. The red rays in Figure 5.1 exemplifies a pair of corresponding rays.

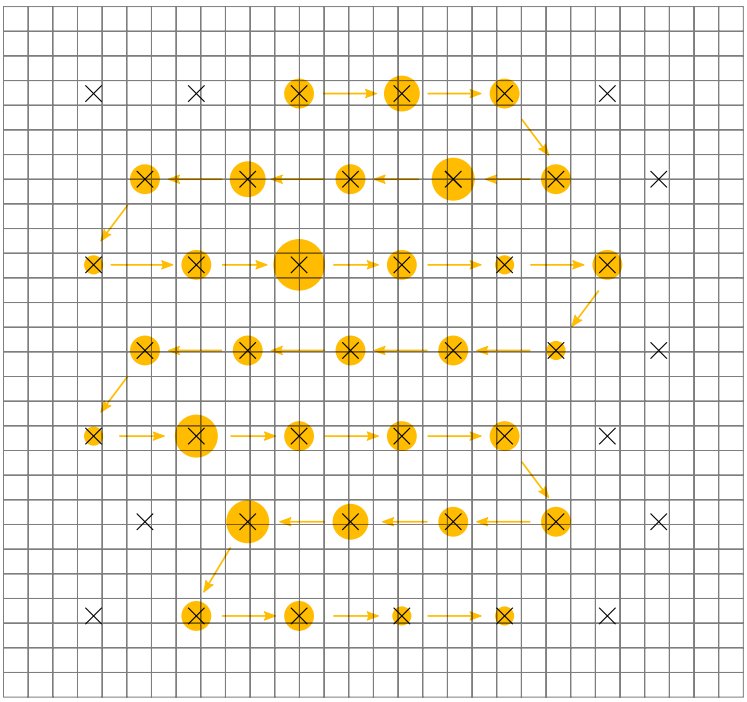


Figure 5.2 Illustration of the bixel grid covering all spots. Only one spot energy layer is included for simplicity. For each bixel in the grid, the ray tracing operation integrates WEPL along the line that originates at the virtual source and passes through the location of the bixel center in the isocenter plane.

Depth integration

The ray tracing identifies corresponding voxels across scenarios in the *angular* sense. To find corresponding voxels for the range-corrected dose mapping, we must also identify which of these voxels are corresponding in terms of WEPL from the virtual source, similar to solving for z' in (4.5).

Each voxel contains a single material. Thus, the RSP is constant in each voxel. However, the RSP depend on anatomical changes and systematic density shifts and is represented by a scenario-dependent scalar field RSP_s over the voxel grid. Its value in voxel k is denoted RSP_{sk}.

Each ray r is implemented to iteratively intersect the voxel grid one voxel at a time. If the ray hits the grid at an angle, the geometrical voxel intersection lengths L_{bsrk} might be different for each encountered voxel k. The increase in WEPL over the voxel k is denoted $\Delta\rho_{bsrk}$ and is given by the product between the geometrical intersection length and the RSP of the voxel

$$\Delta\rho_{bsrk} = L_{bsrk} \times \text{RSP}_{sk}. \tag{5.6}$$

For the n:th voxel along the ray r, counting from the virtual source, the WEPL where the ray enters the voxel is calculated by summing the quantity in (5.6) over the $n-1$ previously intersected voxels. Figure 5.3 provides an illustration.

Figure 5.3 illustrates a challenge with the chosen ray tracing approach. In the figure, it is clear that voxel i_5 is at a different WEPL from the virtual source than *e.g.* voxel j_3. Although being hit by corresponding rays, the *depth* does not correspond and thus no dose should be mapped from j_3 to i_5. This implies that element (i_5, j_3) in A_{bs} should be zero.

Where then should the dose from j_3 be mapped? Voxels i_3 and i_4 are the natural candidates, but how should the dose be distributed? To complicate things further, each voxel is expected to be hit by several rays, so Figure 5.3 does not convey the full picture. It is entirely possible that voxels j_3 and i_5 are at the same depth according to some *other* pair of corresponding rays, implying that element (i_5, j_3) in A_{bs} should not be zero after all. Which ray should then be trusted? Answering this question is of **central importance** in defining the method.

Voting among rays

In our implementation, each ray r is allowed to cast votes on ”how much” voxel j and voxel i should be considered corresponding. The vote σ_{bsrij} cast by ray r has a weight proportional to the WEPL

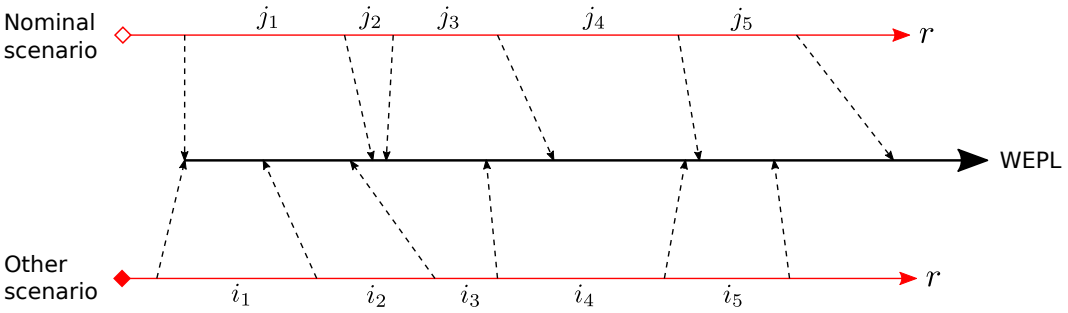


Figure 5.3 The corresponding rays illustrated in Figure 5.1 intersect different voxels in their paths. The nominal scenario ray intersects voxels $(j_1, j_2, \dots)$ and the corresponding scenario ray intersects voxels $(i_1, i_2, \dots)$. For each scenario, WEPL is calculated iteratively by summing the quantity in (5.6) for voxels intersected by the ray. The dashed arrows indicate the WEPL at each voxel boundary.

overlap between the voxels

$$\sigma_{bsrij} = \left| \left[\rho^{\text{entry}}_{bs_0rj}, \rho^{\text{exit}}_{bs_0rj} \right] \cap \left[\rho^{\text{entry}}_{bsri}, \rho^{\text{exit}}_{bsri} \right] \right| \tag{5.7}$$

where for example $\rho^{\text{entry}}_{bs_0rj}$ is the WEPL from the virtual source position to the *entry* of voxel j, integrated along ray r in the nominal scenario s_0. Here $|\cdot|$ denotes the length of an interval on $\mathbb{R}$. Figure 5.3 is useful for the interpretation. The votes are summed over the rays and, unless summing to zero, normalized over the contributing nominal dose voxels

$$\sigma_{bsij} = \frac{\sum_r \sigma_{bsrij}}{\sum_j \sum_r \sigma_{bsrij}} \tag{5.8}$$

to ensure that $\sum_j \sigma_{bsij} = 1$.

Inverse square law

The inverse square factor INV_{bsij} is calculated analogous to equation (4.6)

$$\text{INV}_{bsij} = \left(\frac{\text{VSAD}_{bs_0} + z_{bs_0j}}{\text{VSAD}_{bs} + z_{bsi}} \right)^2 \tag{5.9}$$

for all combinations (b, s, i, j) such that $\sigma_{bsij} \neq 0$. The numerator is interpreted as the physical distance from the virtual source to voxel j in the nominal scenario s_0. The denominator is the same quantity for voxel i in scenario s.

Putting everything together

Finally, the elements of the dose-mapping matrix are calculated as

$$(A_{bs})_{ij} = \sigma_{bsij} \times \text{INV}_{bsij}. \tag{5.10}$$

Going back to (5.5) the beam-dose in voxel i under scenario s is given by a linear combination of the the elements of the nominal beam-dose vector d_{bs_0}. This can be expressed as

$$(d_{bs})_i = \sum_{j \in \mathcal{V}} \sigma_{bsij} \times \text{INV}_{bsij} \times (d_{bs_0})_j \tag{5.11}$$

where most σ_{bsij} equals zero in practice. The expression can be compared to (4.7) for the published algorithm. The weights σ_{bsij} sum to one and indicate how much each voxel j in the nominal beam-dose should contribute to the scenario beam-dose in voxel i.

Hyperparameters

The implementation can be tweaked by adjusting some parameters that may affect both accuracy and performance. Mainly, this relates to the ray trace bixel grid (see Figure 5.2 on the previous page) which determines the density and number of rays. The grid should cover all spots in the isocenter

plane but may also be extended further in all directions. This is later referred to as the ray trace bixel grid *expansion*. Another parameter is the *resolution* of the grid parameterized by the bixel width.

Another parameter is the particle energy for which the RSP is evaluated. As discussed in Section 2.2 and noted in some prototyping experiments, this choice is of less importance for the method. All experiments later in the report evaluate RSP at 130 MeV. The choice of bixel grid parameters is discussed further in Section 7.1.

Summary

Scenarios of patient setup shifts and setup rotations are implemented to modify the gantry-to-patient coordinate transform. Scenarios of anatomical deformations and range errors are implemented as modifications of the RSP field derived from the patient image. The dose mapping from the nominal scenario to some other scenario is calculated based on the result of corresponding ray trace operations under the two scenarios.

As implemented, this scenario dose approximation in theory extends to any scenario that boils down to modifying the gantry-to-patient coordinate transform and/or the patient image. However, the *accuracy* of the approximation is fundamentally limited by the physical assumptions adopted from [Park et al., 2012], which might not be valid for large deviations from the nominal scenario.

6

Experiments

This chapter describes the experiments used to evaluate the performance of our implementation (Section 5.2) of the range-corrected scenario dose approximation method from [Park et al., 2012]. This implementation is referred to as 'deformed' for short since the method can be interpreted as a deformation of the nominal dose cloud. The method 'deformed' is benchmarked against the spot weight shifting method 'spotshift' which is currently offered as an approximate alternative to MC scenario dose calculations ('accurate') in RayStation.

The three methods of interest are summarized in Table 6.1. Please refer to Section 4.2 for a general description of MC dose calculation algorithms and Section 4.3 for a description of the scenario dose approximation methods 'spotshift' and 'deformed'. Implementation details are provided in Chapter 5.

Table 6.1 Summary of algorithms used in the experiments.

Algorithm	Description
'accurate'	The MC algorithm in RayStation 11A.
'spotshift'	The spot weight shifting method offered as an approximate alternative to 'accurate' in RayStation 11A.
'deformed'	Our modified implementation (Section 5.2) of the range-corrected approximation method [Park et al., 2012].

6.1 Prototyping and hyperparameter selection

During the development phase, a simple geometry similar to Figure 4.2 on page 32 was used to quickly gauge the effects of various changes in the implementation of 'deformed'. A few such results are included in the report to explain design and hyperparameter choices as well as to illustrate some similarities between 'deformed' and the ray casting method described by [Schaffner et al., 1999].

6.2 Patient test cases

After prototyping and hyperparameter selection, the scenario dose calculation methods are evaluated on the patient cases summarized in Table 6.2. All patient cases are described in detail in Appendix A.

Table 6.2 Summary of the patient cases used in experiments. The table shows the number of beams, energy layers, and spots in the optimized plans as well as the combinations of scenarios considered. Setup shift $\pm 0.5\,\text{cm}$ means isotropic setup shifts in all three axis directions. Organ motion scenarios are either simulated or inferred from a 4DCT scan.

Case	Beams	Layers	Spots	Setup	Density	Organ motion
Lung Ia	3	42	1299	$\pm 0.5\,\text{cm}$	$\pm 3\,\%$	10 phase 4DCT
Lung Ib	3	42	1299	$\pm 0.5\,\text{cm}$	$\pm 3\,\%$	–
Lung Ic	3	41	613	–	$\pm 3\,\%$	10 phase 4DCT
Lung II	2	54	2815	$\pm 0.5\,\text{cm}$	$\pm 3\,\%$	–
Breast I	2	73	51768	–	$\pm 3\,\%$	$\pm 0.5\,\text{cm}$ simul.

The two lung cases (Lung I, Lung II) are chosen since lung tissue is notoriously difficult for dose calculation algorithms. However, all considered lung plans include only a small number of spots,

meaning that the GPU-accelerated Monte Carlo dose calculation is already very quick. To study the potential of the proposed method to improve efficiency, a breast case (Breast I) with a much larger number of spots is considered as well. The breast case also features a shallow target with smaller density heterogeneities than the lung cases.

All treatment plans were set up and optimized by experienced treatment planners. The provided plans include segmented ROIs, beams, objective functions, robustness settings, and optimized spot weights. Lung Ib is the only departure from the provided plans. It is added to study the Lung I case without any organ motion scenarios which allows for fair comparison between 'spotshift' and 'deformed'.

Although the proposed method extends to scenarios of patient setup rotation, no such scenarios are included in the experiments due to time constraints.

6.3 Scenario dose calculation accuracy

Monte Carlo as ground truth

In this thesis, dose calculation by Monte Carlo (MC, 'accurate') simulation is used as the ground truth to which the two approximate methods 'spotshift' and 'deformed' are compared. This is motivated by the fact that MC is generally regarded the gold standard for accurate dose calculation since it stays true to our best knowledge of the relevant physical interactions at a microscopic scale. Furthermore, the MC implementation in RayStation is used clinically by a large number of proton therapy clinics worldwide, meaning it has passed a large number of QA tests.

Fundamentally, the accuracy of any MC algorithm depends on the number of simulated particle histories. Experiments in this thesis used 20000 ions per spot unless stated otherwise.

Comparing doses for fixed spot weights

Each scenario dose calculation algorithm boils down to an alternative way of calculating the scenario dose deposition matrices D_s (3.12) which are part of the optimization problem statement (3.17). To compare two such methods, one could calculate two versions of D_s for a given patient, beam and scenario and then compare the matrices element by element. Each element D_{ij} is physically interpretable as the dose to voxel number i from spot number j at unit spot MU. When comparing the matrices, all elements should be equal for the dose calculation algorithms to be equivalent.

The dose calculation methods in this thesis are **not** compared by direct comparison of elements in the dose-deposition matrices. Instead, the comparisons are done in the dose-domain for some fixed values x^* of the optimization variables. The benefit of this approach is that doses are easy to visualize. For technical reasons, it also proved easier to implement.

A set of realistic test cases (Section 6.2) are used in the dose calculation experiments. Each such test case includes a patient model, a treatment machine model, a set $\mathcal{S}$ of scenarios for robust optimization, and spot weights $x^* \in \mathbb{R}^N$ corresponding to an already optimized treatment plan. When comparing dose calculation algorithms A and B for such a test case, the scenario dose deposition matrices D_s^{A} and D_s^{B} are implicitly compared by comparing the doses $d_s^{\mathrm{A}} = D_s^{\mathrm{A}} x^*$ and $d_s^{\mathrm{B}} = D_s^{\mathrm{B}} x^*$ for all scenarios $s \in \mathcal{S}$. This is similar to using the algorithms for robust evaluation of an already optimized plan, then comparing the accuracy of the evaluation doses.

Evaluation metrics

There exist several metrics for comparing three-dimensional dose distributions. Such metrics are commonly applied in various QA procedures in which a measured dose distribution is compared with a calculated ditto. In this context, the gamma evaluation method [Low et al., 1998] is ubiquitous. The gamma evaluation index combines a dose-difference metric with a distance-to-agreement metric, where the latter compensates for the sensitivity of dose-difference metrics to small spatial disagreements in regions of large dose gradients. However, the gamma metric is insensitive to several systematic errors and its use in evaluation of dosimetric accuracy has been criticized [Nelms et al., 2013]. Using an insensitive metric as the sole verification might lead to over-confidence in a new method.

In the experiments in this thesis, a simple metric of root-mean-square error (RMSE) is used for quantitative comparison of doses against the ground truth. RMSE is a pure dose-difference metric. For some dose d^{A} (where the algorithm A is either 'spotshift' or 'deformed') compared to the

'accurate' MC dose d^{MC}, it is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{V}_{\text{nz}}|} \sum_{i \in \mathcal{V}_{\text{nz}}} \left(d_i^{\text{A}} - d_i^{\text{MC}}\right)^2} \tag{6.1}$$

where the sum is taken over the voxels $\mathcal{V}_{\text{nz}} = \{i \in \mathcal{V} \colon (d_{\text{A}})_i \neq 0 \text{ or } (d_{\text{MC}})_i \neq 0\}$ where at least one of the compared doses is non-zero.

Keeping in mind that a simple metric might hide interesting errors, a selection of dose distributions are compared visually to also allow for a more qualitative evaluation of the accuracy. In addition, the *scenario-averaged voxel-wise error*

$$\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} d_{si}^{\text{A}} - d_{si}^{\text{MC}} \tag{6.2}$$

and the *scenario-averaged voxel-wise absolute error*

$$\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left| d_{si}^{\text{A}} - d_{si}^{\text{MC}} \right| \tag{6.3}$$

are also investigated for voxels $i \in \mathcal{V}$. By studying (6.3) for all voxels, one can identify areas where the algorithm A shows large errors on average. When keeping the signs as in (6.2), random errors are canceled out whilst systematic errors remain.

6.4 Robust optimization

Apart from the dose calculation accuracy experiments, the performance of the three algorithms 'accurate', 'spotshift' and 'deformed' are also evaluated for use in robust optimization. The robust optimization experiments are run on the same realistic test cases (Section 6.2) as the dose calculation accuracy experiments. Each such test case includes a patient model, a treatment machine model, a set $\mathcal{S}$ of scenarios for robust optimization, and spot weights $x^* \in \mathbb{R}^N$ corresponding to an already optimized treatment plan. Before running the robust optimization experiments, each treatment plan is copied four times and for each copy x^* is discarded by resetting the optimization.

Three plan copies are used for repeating the same optimization procedure three times while using either 'accurate', 'spotshift' or 'deformed' for scenario dose calculation. This means solving an optimization problem on the form (3.17) three times with the same solver and settings but with different matrices D_s.

The fourth copy is used to run the optimization procedure with robustness turned off for all objectives. This means modifying the sets of robust ($\mathcal{O}_r$) and non-robust ($\mathcal{O}_{nr}$) objectives in the optimization problem according to

$$\mathcal{O}_{nr} \leftarrow \mathcal{O}_r \cup \mathcal{O}_{nr} \tag{6.4}$$

$$\mathcal{O}_r \leftarrow \emptyset \tag{6.5}$$

such that the second term of the full objective in (3.17) disappears. A similar modification could be applied to the dose constraints, but the considered test cases do not include any dose constraints at all. The purpose of running the non-robust optimization is to provide a baseline to which the other three optimized plans can be compared in terms of robustness.

For all runs of the robust optimization, the number of (outer) SQP iterations is a fixed number. The settings are kept unaltered from the original treatment plans, meaning 100 iterations for 'Lung I' and 'Breast I', and 80 iterations for 'Lung II'. For all test cases, a spot filtering procedure that prunes the number of optimization variables is run after 20 iterations.

A comment on the non-robust baseline. All test cases considered here use robust objectives defined on the clinical target volume (CTV). The purpose of these objectives is to ensure target coverage in all scenarios. A real treatment planner would never remove robustness of these objectives without modifying other aspects of the objective as well. For instance, the target coverage objectives could be defined on an expanded CTV volume called a planning target volume (PTV) which to some degree may increase robustness. Still, the concept of a PTV is flawed for proton therapy which is the reason for using robust optimization in the first place.

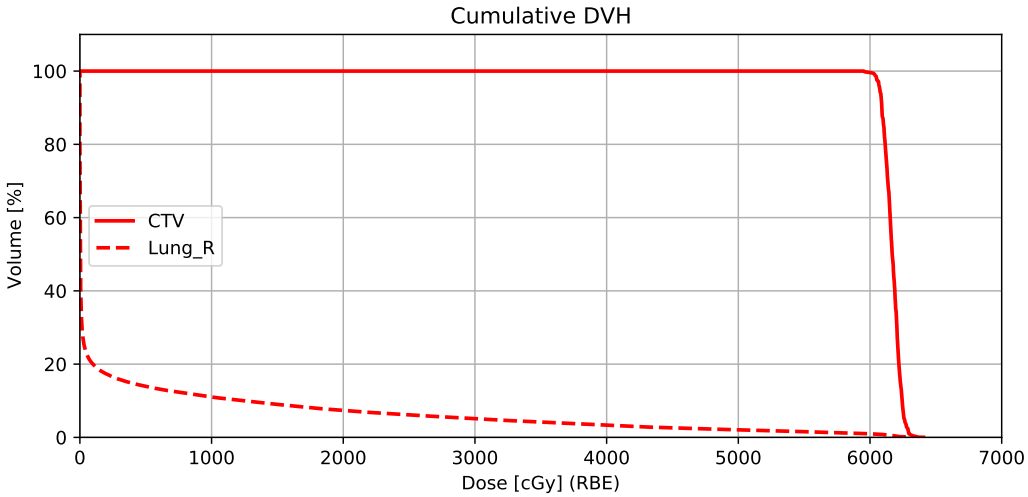


Figure 6.1 A cumulative dose-volume histogram.

The purpose of this thesis is not to investigate differences between PTV and CTV planning. Instead, the question of interest is how the robustness of CTV coverage depends on the choice of scenario dose calculation algorithm used in robust optimization. In this context, the most natural baseline is to solve the nominal optimization problem (3.11) instead of the robust optimization problem (3.17) with the same same set of objectives $\{f_i\}_{i\in\mathcal{O}}$.

Evaluation metrics

There is unfortunately no universal way to report the robustness of a treatment plan, and there are almost as many metrics as there are papers on the topic [Lomax, 2018]. In conventional (non robust) treatment plan evaluation, differential and cumulative dose-volume histograms (DVHs) and metrics derived thereof are commonly used tools. In this thesis, such histograms are used for robust evaluation in an approach inspired by, *e.g.*, [Bokrantz and Fredriksson, 2017].

Dose-volume histograms. Dose-volume histograms (DVHs) summarize three dimensional dose distributions in a single graph by disregarding all spatial information. There are two types: differential DVHs and cumulative DVHs. Differential DVHs discretize the dose into bins and display for each bin the number (or fraction) of voxels within some ROI where the dose lies within the bin interval. Cumulative DVHs use the same bins, but display instead the number (or fraction) of voxels within the ROI where the dose lies within *or below* the bin interval.

Figure 6.1 illustrates the concept. A curve in the cumulative DVH can be interpreted as a plot of the volume-at-dose function which is defined as

$$V_y(d) = \sum_{i\in\mathcal{R}} r_i \mathbb{1}_{d_i \geq y} \tag{6.6}$$

and interpreted as the volume percentage of a ROI (indexed by $\mathcal{R}$) receiving at least the specified dose level y. Here $\mathbb{1}$ is the indicator function, d the dose vector, and r_i the ROI relative volume vector (see Section 3.2).

The most common goal in treatment planning is to achieve a uniform dose in the target volume. In the example DVH in Figure 6.1, we see that the entire CTV receives at least 6000 cGy but no part of the CTV receives more than 6500 cGy. Thus, the dose distribution is fairly uniform within the CTV. For the right lung ('Lung R' in the figure), the DVH curve instead drops quickly from one, and the DVH curve for example tells us that less than 10 % of the volume of the right lung receives more than 2000 cGy. The lung should ideally receive no dose at all. However, that is not feasible in this example since the CTV is located within the lung.

Using DVHs in robust evaluation. To evaluate and compare the robustness of optimized treatment plans in this thesis, a family of DVH curves is plotted for each plan. Plotting one DVH curve for each ROI and scenario used in optimization would provide the most information, but that quickly crowds the graph. To keep graphs readable, the robust evaluation DVHs in this report show three curves per

ROI: one curve for the nominal scenario and one curve each for the scenarios giving the *least/most* amount of dose to the ROI. The DVHs are restricted to only show the CTV since all robust objectives are defined on the CTV in all considered cases.

Computational efficiency

The built-in timing functionality in RayStation is used to provide some estimates on computational efficiency of the robust optimization. To estimate the runtime of the initialization phase, including initial dose computations, optimizations are run for zero iterations and the timer log entry 'time in optimization' is noted. To estimate the SQP iteration runtime, the experiment is re-run with the full number of iterations. The average iteration runtime is then calculated as the sum of the individual iteration runtimes divided by the number of iterations. A technical comment is that the experimental `RobustnessModule.useRobustnessMask` was set to false (non-default) to allow exporting the full, non-cropped scenario doses. This has an impact on the absolute numbers of efficiency.

Table 6.3 Some information on the computer used for experiments.

Component	Information
CPU	Intel Core i9-10940X
GPU	NVIDIA Quadro RTX 5000 GPU
RAM	65536 MB

Although absolute numbers of efficiency are not of main interest in this thesis, some computer-specific information is listed in Table 6.3 for completeness.

7

Results & discussion

7.1 Prototyping and hyperparameter selection

This section presents some results to motivate design and hyperparameter choices for 'deformed'. All experiments use the geometry and nominal dose depicted in Figure 7.1, which shows the dose delivered by a single 140 MeV spot in a homogeneous body of water into which a 2 cm thick slab of bone ('Bone 1' in RayStation) has been inserted. The bone material has mass density $1.85\,\mathrm{g/cm^3}$ as compared to $1.00\,\mathrm{g/cm^3}$ for water and also differs in elemental composition. The dose is calculated with 'accurate' using 100000 simulated ions in a uniform dose grid with resolution $2 \times 2 \times 2$ mm or $3 \times 3 \times 3$ mm depending on the experiment.

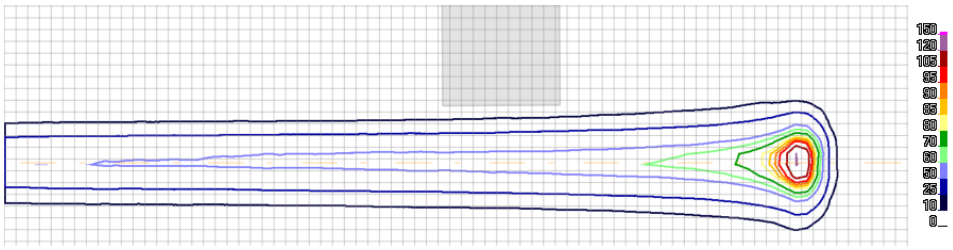


Figure 7.1 The nominal dose used in the hyperparameter experiments, here in a $2 \times 2 \times 2$ mm dose grid. The geometry is inspired by [Schaffner et al., 1999], see Figure 4.2 on page 32.

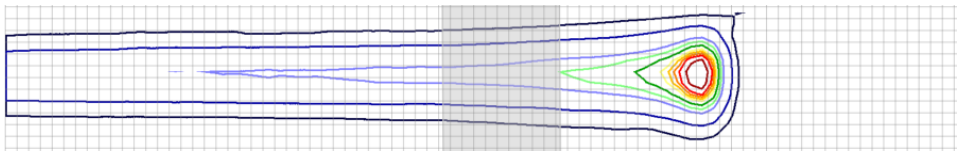
RSP or mass density for WEPL integration

An early experimental implementation (internal at RaySearch) of a method similar to [Park et al., 2012] approximated WEPL (2.6) as the path integral of (relative) *mass density* rather than (relative) *stopping power*. The stopping power of a material is proportional to the mass density, but as (2.2) shows it also depends also on other material properties which means that the two WEPL definitions are not equivalent. A simple experiment illustrates the importance of this distinction for 'deformed'.

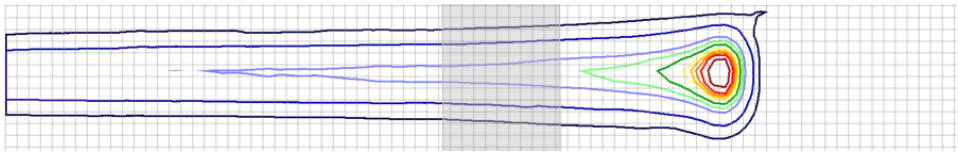
In the experiment, a patient setup shift of 2.5 cm moves the spot in Figure 7.1 upwards such that the beam path intersects the bone slab. Figure 7.2(c) shows the effect of this setup shift as calculated with 'accurate' (100000 ions) and the result as calculated with 'deformed' is shown in (b). The dose in (a) is calculated with a modified version of 'deformed' in which RSP is substituted for relative mass density in the WEPL integration (5.6).

Figure 7.2(c) reveals that the setup shift in this case leads to a decreased range. This is expected since the beam path traverses the bone in this scenario. A close comparison to Figure 7.1 of the isodose curves also reveals that the dose magnitude at the Bragg peak increases compared to the nominal scenario. The 'deformed' dose in (b) correctly mimics the range effect, but it fails to model whatever physical phenomenon causes the increased Bragg peak dose. The 'deformed' dose in (a), using relative mass density for WEPL calculations, overestimates the range effect due to the bone and underestimates the spot range by two voxel widths, *i.e.*, 4 mm.

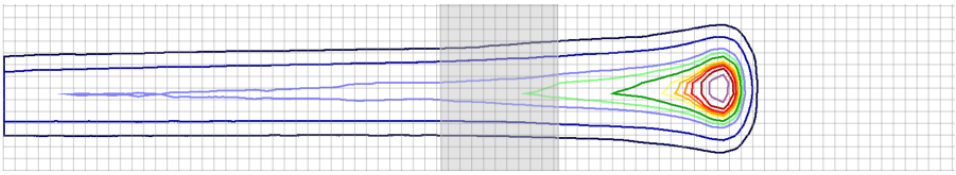
Since the implementation of 'deformed' is agnostic to the choice of scalar field used in the WEPL integration, and since both RSP and mass densities are easily retrievable in RayStation, using RSP is the better choice.



(a) 'deformed': RSP ≈ relative mass density



(b) 'deformed': exact RSP



(c) 'accurate'

Figure 7.2 A patient shift of 2.5 cm shifts the spot dose in Figure 7.1 upwards such that the beam passes through a slab of bone. The figure illustrates the range error introduced for 'deformed' when approximating RSP with the relative mass density.

Ray trace bixel grid resolution

An important parameter for both accuracy and speed of 'deformed' is the resolution of the ray trace bixel grid. The bixel grid is illustrated in Figure 5.2 on page 39. Making the resolution higher means decreasing the width of the square bixels. Here, a simple experiment on one spot demonstrates how a suitable resolution depends on the dose grid resolution.

For this experiment, the dose grid resolution is changed to $3 \times 3 \times 3$ mm in order to harmonize with the patient test cases in Section 6.2. The nominal dose for this new resolution is shown in Figure 7.3. In the figure, the isocenter is marked with a circle and a cross. The bixel grid is defined in the isocenter plane, *i.e.*, the plane that intersects the isocenter and that is perpendicular to the beam central axis.

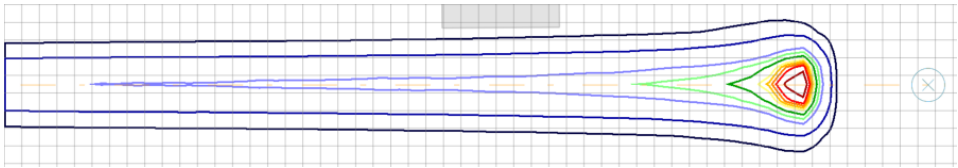


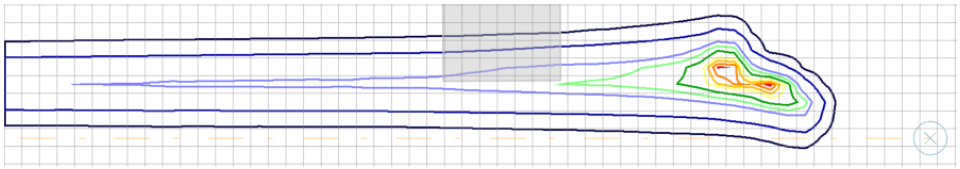
Figure 7.3 The nominal 'accurate' dose recalculated in a $3 \times 3 \times 3$ mm dose grid.

This time, the patient setup shift is set to 1 cm downwards in the figure, which shifts the spot dose the same distance upwards such that half the beam intersects the bone slab and half does not. Figure 7.4(a) on the facing page shows the effect of this setup shift as calculated with 'accurate' with 100000 ions. Using 'deformed' with a high bixel grid resolution gives the result in (b). The subsequent (d)-(f) illustrate how the quality of the 'deformed' dose degrades when the bixel grid resolution is decreased, *i.e.*, when the bixel width is increased.

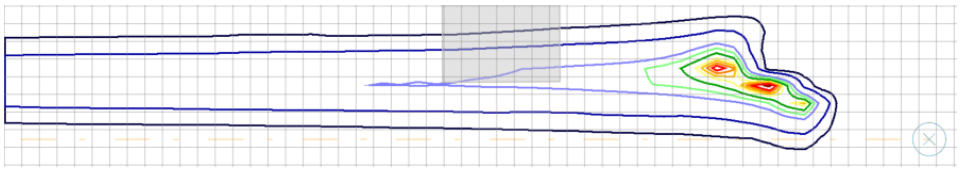
Figure 7.4 shows that the 'deformed' dose is almost unaffected by doubling the bixel width from 0.5 mm to 1 mm. However, doubling the bixel width cuts the number of rays (and thus the ray tracing computational costs) to a fourth. This has a significant effect on the efficiency since ray tracing is the most time-consuming part of the algorithm.

For bixel widths larger than half the dose grid voxel width (larger than 1.5 mm) we observe a degradation of the quality of the 'deformed' dose. In a similar experiment with dose grid resolution $2 \times 2 \times 2$ mm (data not shown), a similar degradation was observed for bixel widths larger than 1 mm. This reinforces a conclusion that the ray trace bixel width should at least be smaller than half the width of a dose grid voxel.

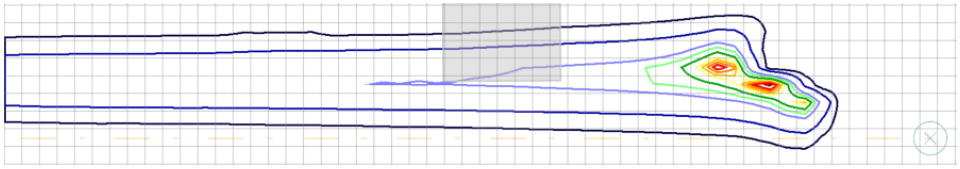
Figure 7.4 fails to reveal that the aliasing effect is more pronounced when the beam central axis is not lined up perfectly with the dose grid. Further investigation into how such aliasing effects impact the accuracy of 'deformed' is needed if the method is to be developed further. A theoretical analysis is made complicated by the diverging rays and non-uniform sampling in the WEPL depth domain. Due to time constraints, we here settle with observing the effect and proceed to use bixel widths of 0.7 mm for experiments on a $3 \times 3 \times 3$ mm dose grid.



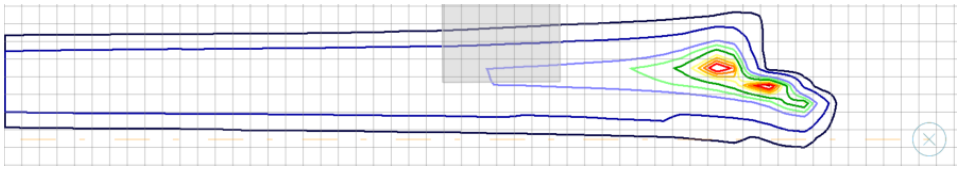
(a) 'accurate'



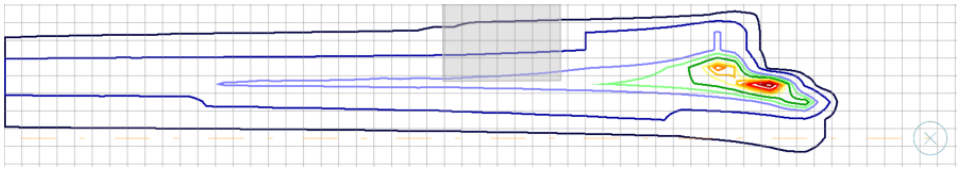
(b) 'deformed': 0.5×0.5 mm ray trace bixel grid



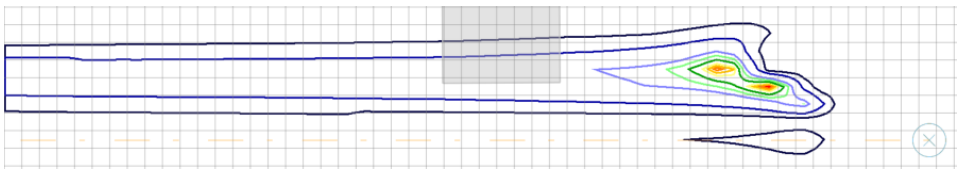
(c) 'deformed': 1.0×1.0 mm ray trace bixel grid



(d) 'deformed': 1.5×1.5 mm ray trace bixel grid



(e) 'deformed': 2.5×2.5 mm ray trace bixel grid

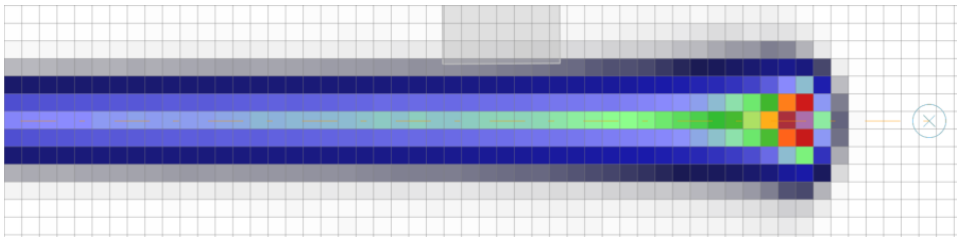


(f) 'deformed': 3.5×3.5 mm ray trace bixel grid

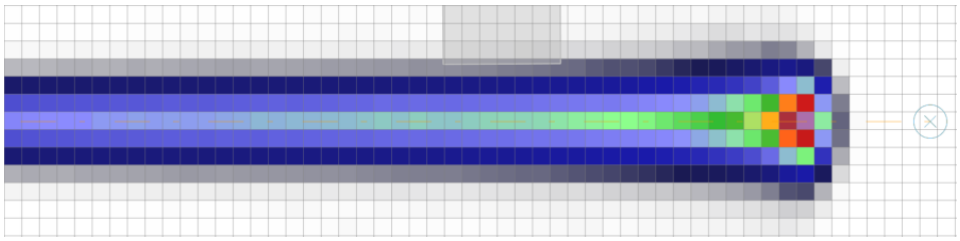
Figure 7.4 Illustration of the effect of lowering the the ray trace bixel grid resolution too much. In these images, a patient shift of 1 cm downwards has shifted the spot dose 1 cm upwards such that half the beam traverses the bone slab.

Ray trace bixel grid expansion

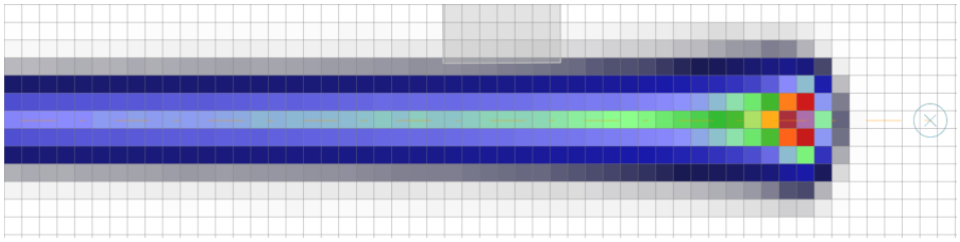
Another important parameter of 'deformed' is the *bixel grid expansion*. To ensure that the ray trace rays sample all dose grid voxels with a non-negligible amount of dose, the bixel grid cannot be tightly cropped to the convex hull of the spots in the isocenter plane. As illustrated in Figure 7.5(a), each spot dose is at least a few centimeters wide. To properly sample the dose from the edge spots, the



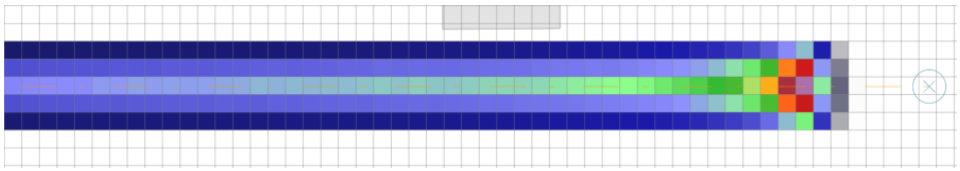
(a) 'accurate'



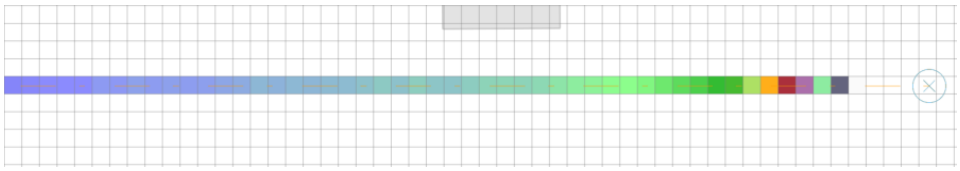
(b) 'deformed': 3.0 cm ray trace bixel grid expansion



(c) 'deformed': 1.5 cm ray trace bixel grid expansion



(d) 'deformed': 0.5 cm ray trace bixel grid expansion



(e) 'deformed': 0.0 cm ray trace bixel grid expansion

Figure 7.5 Illustration of the effect of lowering the ray trace bixel grid expansion too much. In these images, the dose is deformed to a dummy scenario which is identical to the nominal scenario.

ray trace bixel grid must extend outside the spots as illustrated in Figure 5.2 on page 39. Expanding the grid in all directions ensures that all relevant dose grid voxels are sampled. However, it also negatively affects the computational performance.

Figure 7.5 illustrates the effect of choosing various bixel grid expansions. Here 'deformed' is used to deform the 'accurate' dose from the nominal scenario to the nominal scenario which should, in theory, have no effect whatsoever. However, if the dose is only sampled along the beam central axis as in (e), a majority of the dose disappears. By gradually increasing the grid expansion in (d)–(b), the accurate dose is gradually recovered in full.

The difference between expanding the grid 1.5 cm and 3.0 cm is small in this case, but since accuracy rather than efficiency is of primary concern in this thesis, 3.0 cm is used for the experiments.

Summary of the hyperparameter experiments

The hyperparameter experiments suggests setting the ray trace bixel grid to a uniform, square grid that covers all spots in the isocenter plane plus an extra 3 cm in all directions. The resolution is

parameterized by the bixel width which is set to 0.7 mm. Furthermore, we conclude that using the exact RSP gives better results than approximating it with the relative mass density.

As a side effect of the experiments in this section, we may also do some qualitative observations regarding the accuracy of 'deformed'. Figure 7.5(b) on the facing page serves as a first sanity check. It shows that deforming the dose to and from the same scenario, as desired, has no effect.

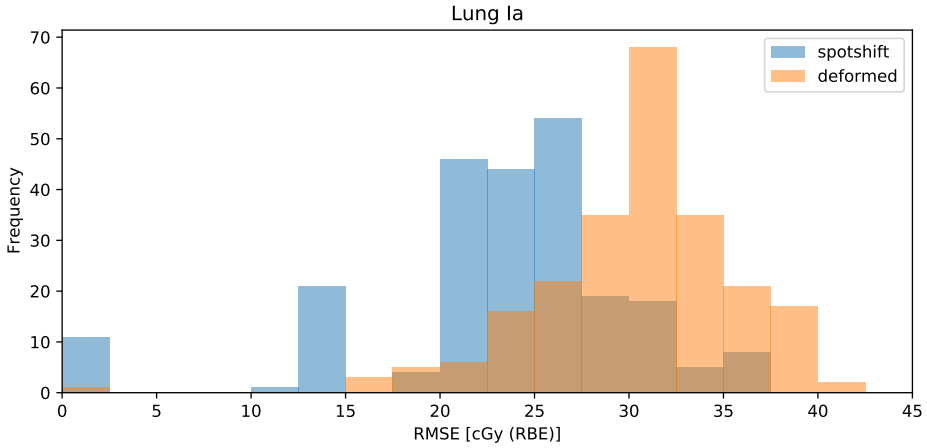
Comparing Figure 7.4(b) on page 49 with Figure 4.2 on page 32 reveals some interesting similarities between 'deformed' and the ray casting algorithm in [Schaffner et al., 1999]. The significant difference between the two methods is that 'deformed' uses the nominal, 'accurate' dose as reference whereas the ray casting method uses some generic dose in water. For both methods, the scenario dose is calculated by stretching the reference dose along rays from the source, accounting only for the WEPL along each ray. One erroneous effect of this approximation is the sharp dose gradient seen behind the bone slab. The methods only account for what happens along each ray and even ignores the depths at which deviations from water occur. The 'accurate' MC simulation, correctly accounts for MCS beyond the bone slab which as seen in Figure 7.4(a) on page 49 smooths the isodose curves behind the bone.

7.2 Scenario dose calculation accuracy

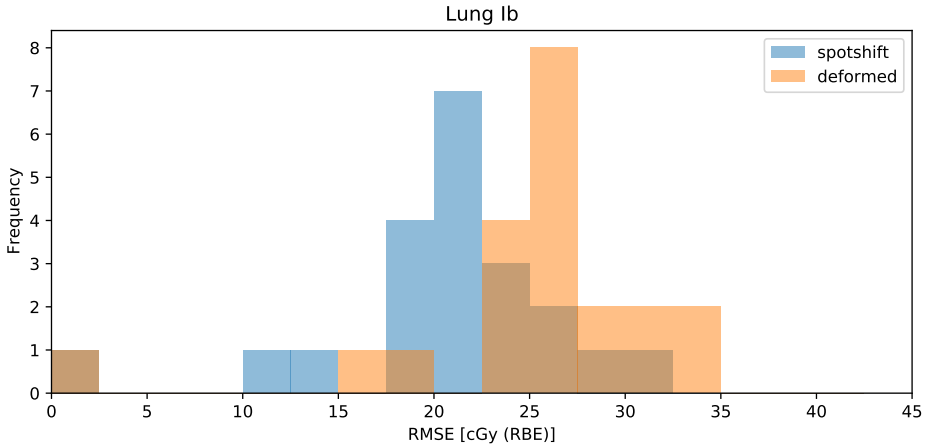
Dose difference statistics

Lung case I. The test case 'Lung I' includes seven setup shift scenarios, three density shift scenarios and 11 organ motion scenarios. Considering all combinations, this results in 231 scenarios in total. The large number of scenarios makes this case ideal for quantitative comparisons of scenario dose calculation accuracy.

Figure 7.6(b) shows a histogram of the RMSE (6.1) for the scenario doses calculated by 'deformed' and 'spotshift' for the 21 scenarios with *nominal anatomy*, *i.e.*, for 'Lung Ib'. The organ motion scenarios are excluded to allow for fair comparison between the methods. We recall that 'spotshift' uses one 'accurate' calculation for each organ motion scenario whereas 'deformed' uses a single 'accurate' calculation for the nominal scenario only. The figure shows a slightly lower accuracy for 'deformed' as compared to 'spotshift' in terms of RMSE.



(a) 'Lung Ia': all 231 scenarios, including organ motion.



(b) 'Lung Ib': all 21 scenarios with nominal anatomy.

Figure 7.6 Histogram of the RMSE (6.1) for scenario doses calculated by 'spotshift' and 'deformed' in the first lung case.

Interestingly, the accuracy of 'spotshift' decreases slightly when organ motion scenarios are introduced, as illustrated in the histogram for 'Lung Ia' in Figure 7.6(a). This result is unexpected since the method in theory does no additional approximations for organ motion scenarios. However, the difference is small and possibly not statistically significant. The same figure illustrates that the accuracy of 'deformed' also decreases (RMSE increases) when organ motion scenarios are included. This is expected since the method does additional approximations for such scenarios. However, the difference in accuracy is unexpectedly small which warrants further analysis.

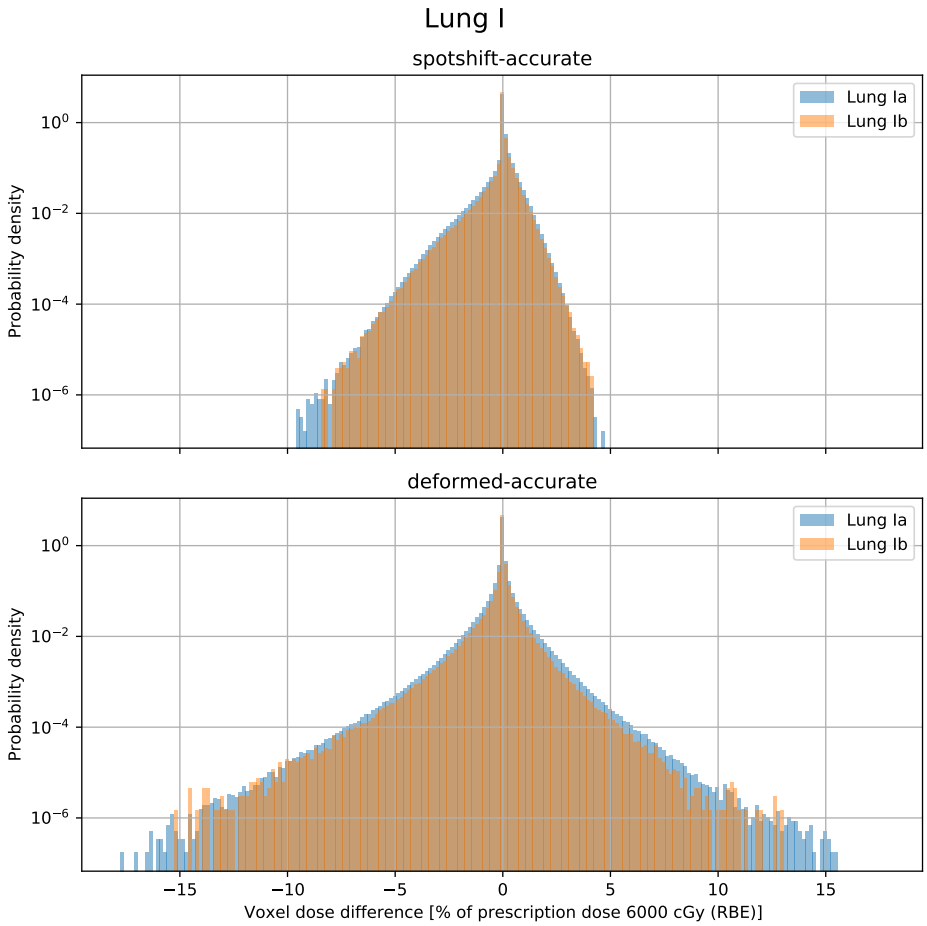


Figure 7.7 Histogram of the voxel-wise dose differences for cases 'Lung Ia' (setup shifts, density shifts, organ motion) and 'Lung Ib' (setup shifts, density shifts). Note the log scale on the y-axis. For 'Lung Ib', the RMSE standard deviations are 0.36 for 'spotshift' and 0.43 for 'deformed' expressed in % of the prescription dose.

The histograms in Figure 7.7 are calculated from the voxel-wise dose differences underlying the RMSE calculations, collected either over all scenarios ('Lung Ia') or over all scenarios with nominal anatomy ('Lung Ib'). The histograms are normalized to a probability distribution and plotted on a log scale to reveal any small differences.

The histograms in Figure 7.7 show several interesting things. First, for 'spotshift', we observe that the probability distribution is almost identical with and without organ motion. The noticeable differences are seen for the very uncommon large negative deviations. One possible statistical explanation for why these differences do not appear for 'Lung Ib' is that the voxel errors are accumulated over only 21 scenarios as compared to 231 scenarios for 'Lung Ia'. This makes it less likely for the uncommon errors to occur.

A similar trend is observed for 'deformed' where the most uncommon (and largest) errors only appear for 'Lung Ia'. Due to the square in the definition (6.1) of RMSE, a small number of large errors has a large impact on the metric. This to some degree explains the observed differences in Figure 7.6. For 'deformed', we additionally observe a small but consistent increase in the prevalence of errors of all magnitudes when organ motion scenarios are added.

Lung case II. To ensure that the above comparison can be generalized to other patient geometries, similar histograms of voxel-wise differences are plotted for the other lung case, 'Lung II', in Figure 7.8. This case only contains setup shift and density shift scenarios, making comparison between 'spotshift' and 'deformed' fair.

Based on Figure 7.8, we again note that the accuracy of 'deformed' is worse than of 'spotshift', with voxel-wise errors of larger magnitudes being more common for the latter. The 'spotshift' method is remarkably consistent, showing very similar patterns for 'Lung I' in Figure 7.7 as for

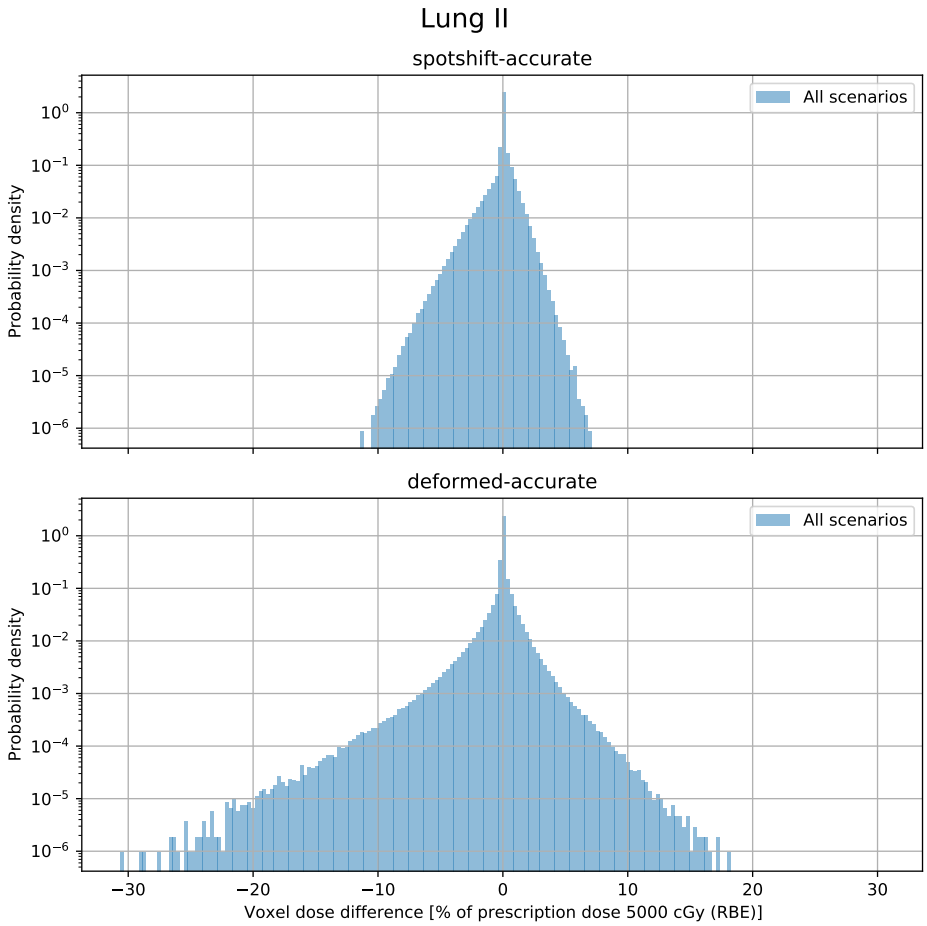


Figure 7.8 Histogram of the voxel-wise dose differences for the 'Lung II' case (setup shifts, density shifts). Note the log scale on the y-axis. The RMSE standard deviations are 0.60 for 'spotshift' and 0.91 for 'deformed' expressed in % of the prescription dose.

'Lung II' in Figure 7.8. The voxel wise errors remain within $\pm 10\,\%$ of the prescription dose with a skew towards the negative side, *i.e.*, underestimating the scenario dose compared to 'accurate' calculations. The same skew is seen for 'deformed'. However, the magnitudes of the dose errors are larger, and comparing 'Lung I' to 'Lung II', we observe that the accuracy of 'deformed' is less consistent across patient geometries. In some individual voxels, errors are as large as $20 - 30\,\%$, but the log scale visually exaggerates their prevalence.

Breast case I. For the breast case 'Breast I', Figure 7.9 shows the histogram of voxel-wise dose differences. The breast case includes combinations of three density shift scenarios and seven simulated organ motion scenarios (in which the breast moves 0.5 cm in some direction). Here, a new pattern appears where 'deformed' severely underestimates the dose in some voxels. This is not as common for the scenarios without simulated organ motion. The root cause for this issue is that the dose which is being deformed lies close to the patient surface. In some organ motion scenarios, 'deformed' tries to map dose from *outside the patient volume*. Such dose is not calculated by 'accurate' and does not exist in the dose grid. The same effect appears in some density shift scenarios. When the density increases, the dose is contracted in the beam direction, and 'deformed' tries to map dose from *behind the patient*; dose that is again not calculated by 'accurate' since it is deposited outside the dose grid. The issue is illustrated in Figures 7.10 and 7.11.

Visual dose comparisons

Figure 7.10 illustrates what happens when 'deformed' tries to fetch dose from outside the patient volume in an organ motion scenario. The scenario consists of a breast movement of 0.5 cm, as illustrated by the deformed CT image in (c). The dashed purple line corresponds to the new CTV location under the scenario. The 'accurate' dose in this scenario is depicted in (a) and can be compared to the

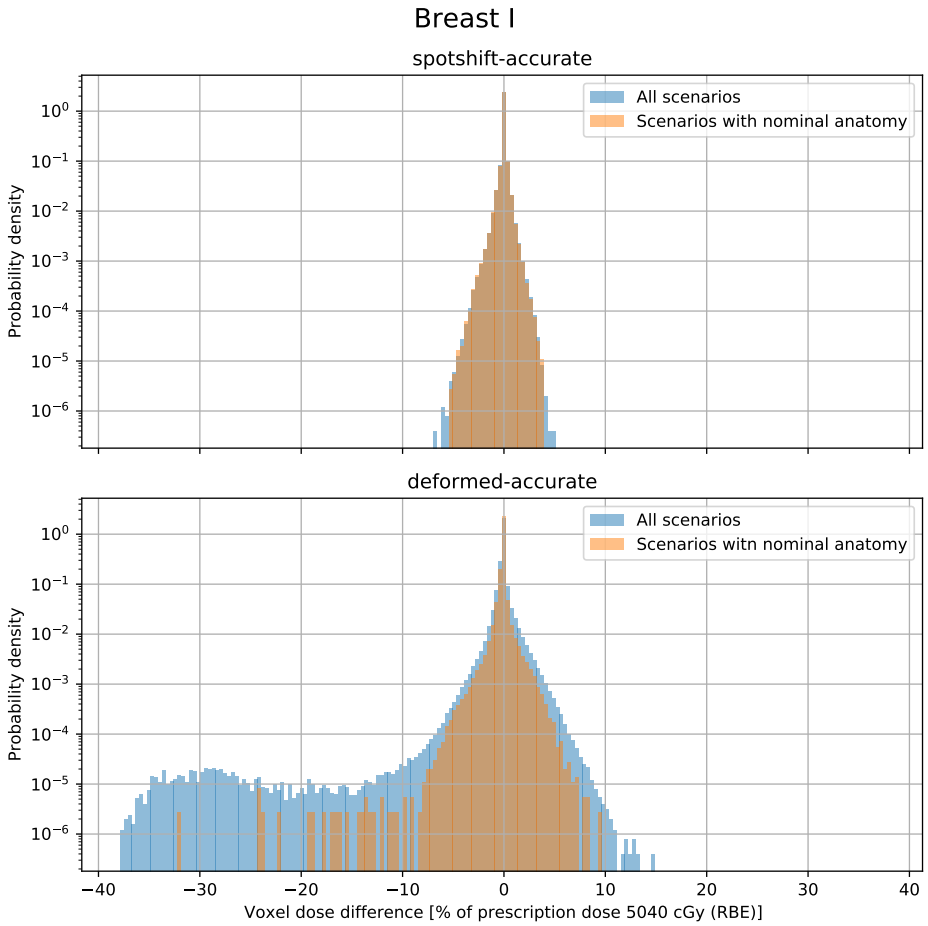


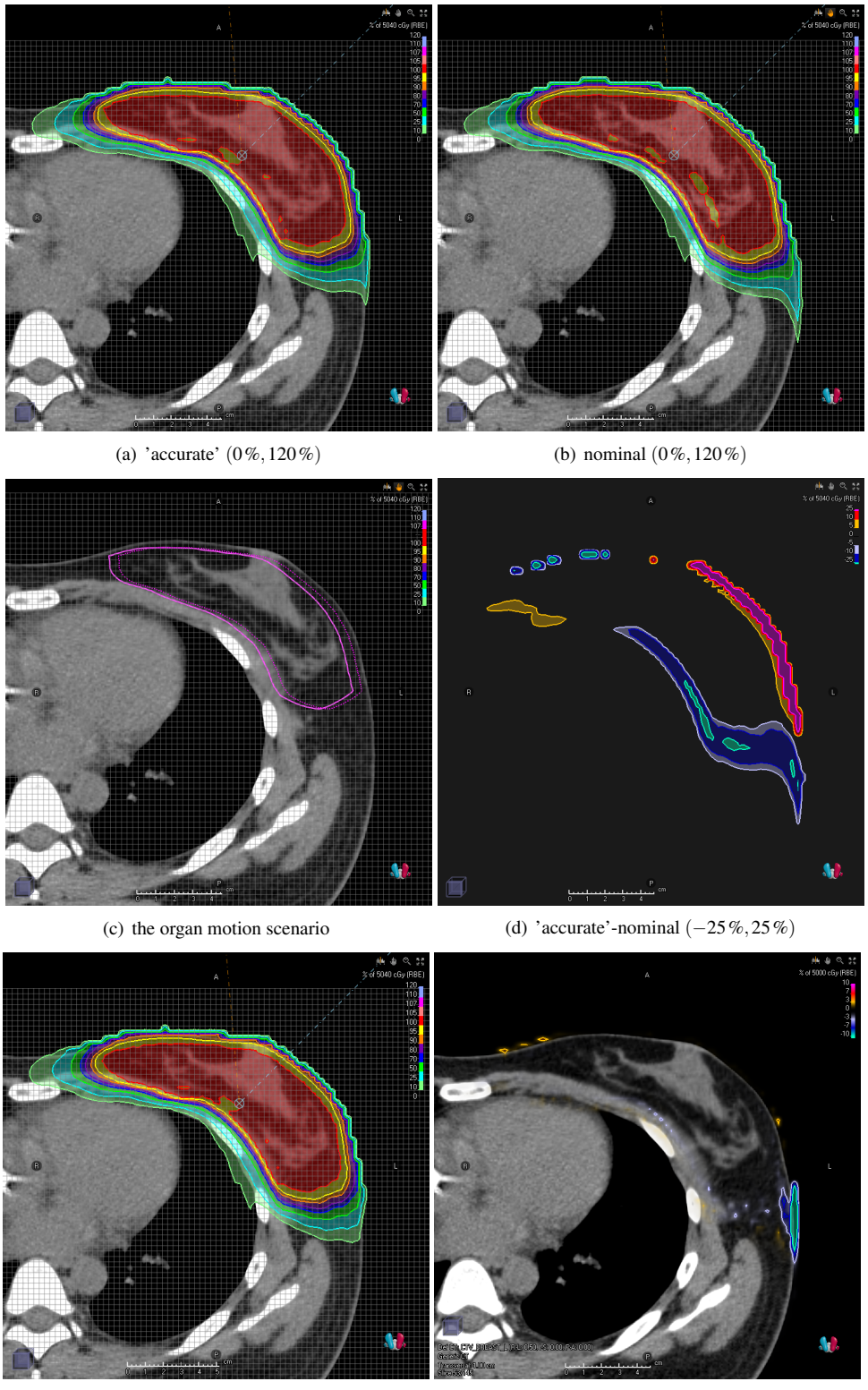
Figure 7.9 Histogram of the voxel-wise dose differences for the 'Breast I' case (density shifts, organ motion). Note the log scale on the y-axis. Considering only the scenarios with nominal anatomy, the RMSE standard deviations are 0.20 for 'spotshift' and 0.36 for 'deformed' expressed in % of the prescription dose.

nominal dose in (b). To emphasize the difference between the two doses, their voxel-wise difference is plotted in (d). The deformed dose is depicted in (e) and the voxel-wise difference to the 'accurate' dose is seen in (f).

As seen in Figure 7.10(f), 'deformed' successfully deforms the nominal dose (b) to a scenario dose which largely corresponds well to the 'accurate' scenario dose (a). However, we also observe a region in the bottom right where the dose is severely underestimated. The explanation for this phenomenon is that this region is located at an angle and at a WEPL relative to the beam source which, under the nominal scenario, corresponds to a location *outside the patient* and even outside the dose grid. Thus, 'deformed' has no prior information of the dose in these voxels.

To illustrate this problem more clearly, we also investigate a setup shift scenario (not originally part of the 'Breast I' case) which shifts the patient 1 cm to the right in the figure, or equivalently shifts the beams 1 cm to the left. The result is shown in Figure 7.11 on page 57, in which also the 'spotshift' result is included. When the beams are moved as in this scenario, the dose far to the right originates from spots which, under the nominal scenario, deposits dose outside the patient. Such spots are included in the plan to ensure irradiation of the tumor in all scenarios. As the figure clearly illustrates, this causes problems for 'deformed'.

The issue illustrated here is a major drawback of the current implementation of 'deformed' which can be expected to have largest impact on glancing-angle treatment of surface tumors. However, we also note that the accuracy of 'deformed' is high within the CTV volume, which in the end is all that matters for the robust optimization in this case since all robust objectives are defined on the CTV.

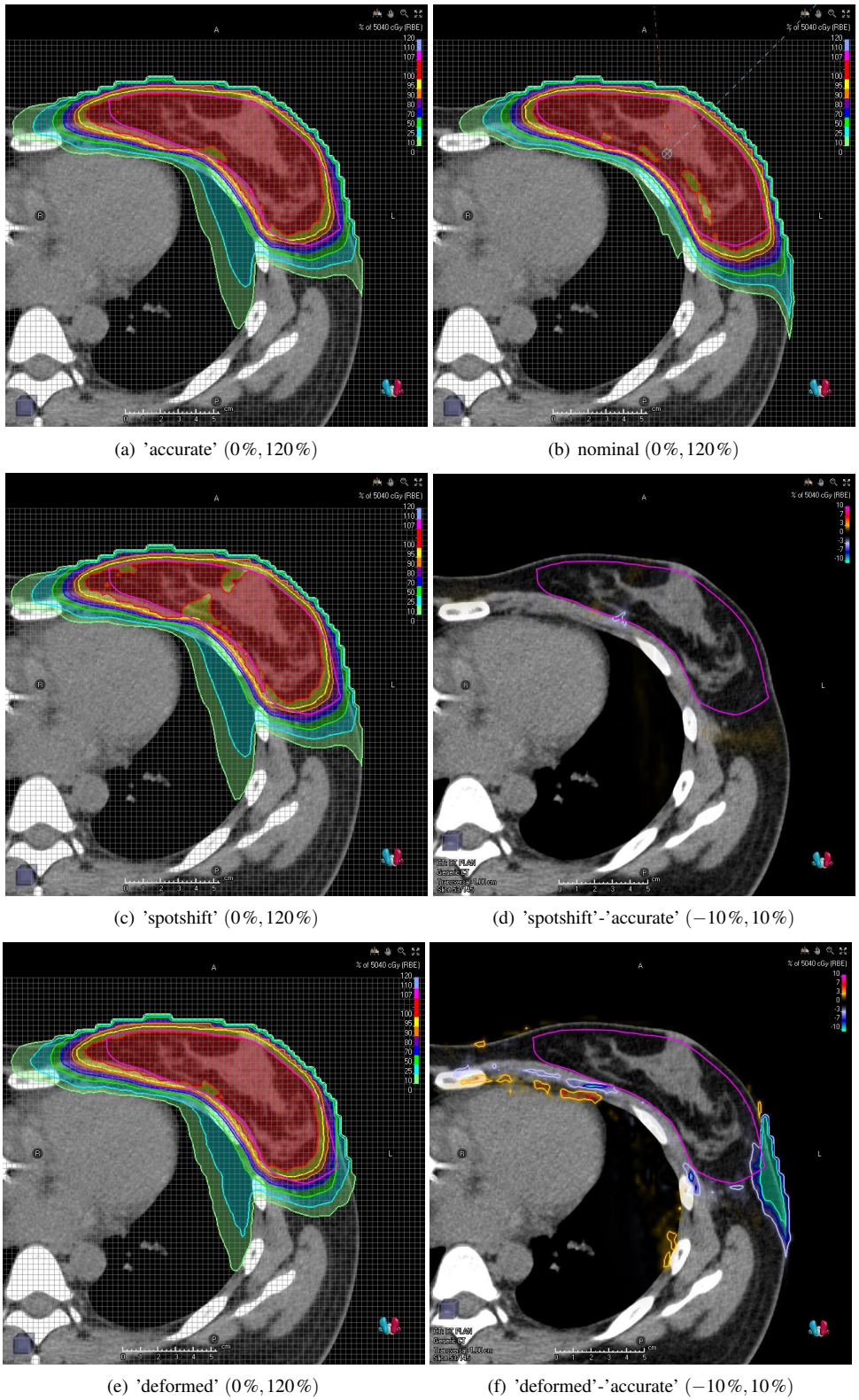


(a) 'accurate' $(0\,\%, 120\,\%)$ (b) nominal $(0\,\%, 120\,\%)$

(c) the organ motion scenario (d) 'accurate'-nominal $(-25\,\%, 25\,\%)$

(e) 'deformed' $(0\,\%, 120\,\%)$ (f) 'deformed'-'accurate' $(-10\,\%, 10\,\%)$

Figure 7.10 Illustration of the effect of dose calculation accuracy of 'deformed' under the organ motion scenario illustrated in (c). The difference between the nominal dose and accurate scenario dose is highlighted in (d). Numbers in parenthesis indicate the range of the color bar in percentages of the prescription dose. We note that the accuracy of 'deformed' is high except in the bottom right.



(a) 'accurate' $(0\,\%, 120\,\%)$ (b) nominal $(0\,\%, 120\,\%)$

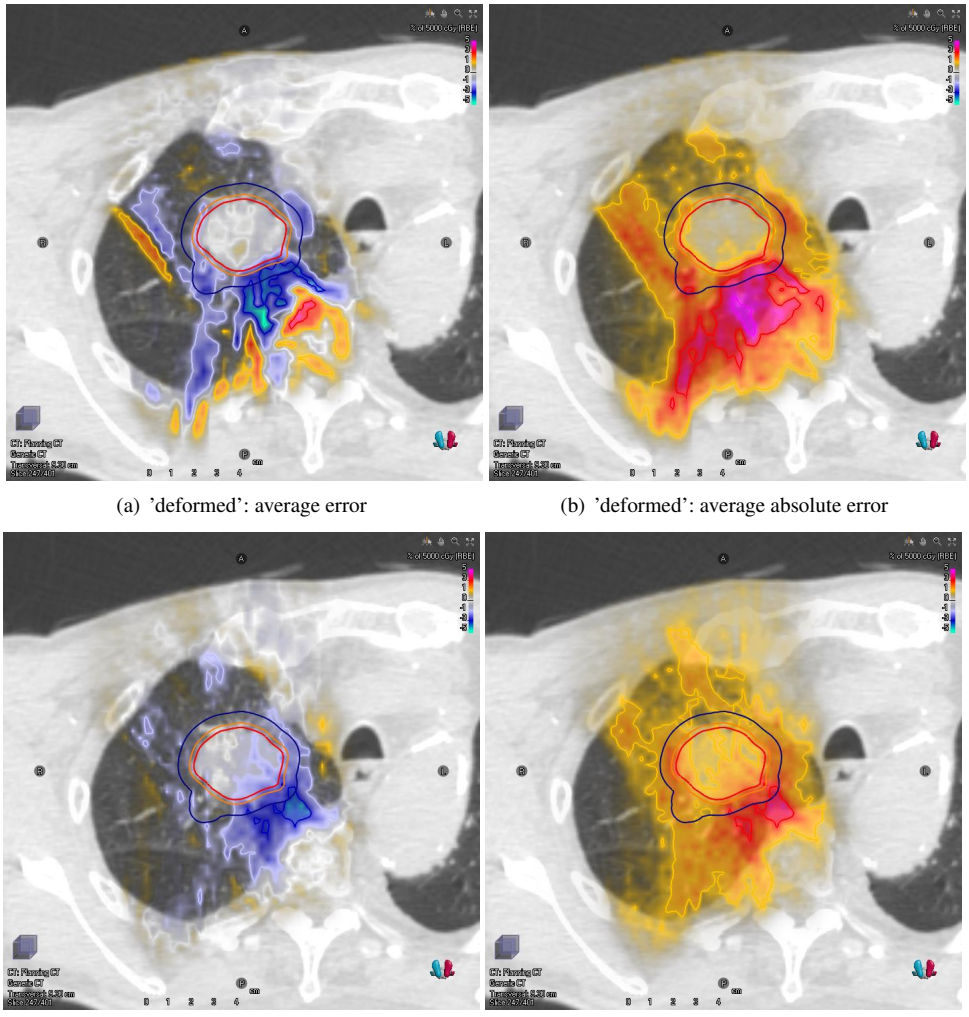
(c) 'spotshift' $(0\,\%, 120\,\%)$ (d) 'spotshift'-'accurate' $(-10\,\%, 10\,\%)$

(e) 'deformed' $(0\,\%, 120\,\%)$ (f) 'deformed'-'accurate' $(-10\,\%, 10\,\%)$

Figure 7.11 To further illustrate the edge effect in Figure 7.10, this figure shows the scenario doses under a scenario in which the patient moves 1 cm to the right in the image. Numbers in parenthesis indicate the range of the color bar in percentages of the prescription dose. We note that the edge issue is problematic in this scenario, but the accuracy is still high within the CTV.

The statistical analysis of voxel-wise differences quantifies how often and how much the approximate doses (calculated by 'spotshift' or 'deformed') deviate from the 'accurate' dose. It is also interesting to consider the spatial aspect of where these errors occur. Are errors more common in some voxels?

Figure 7.12 illustrates where the errors occur for 'Lung II'. In (b) and (d), we see the *scenario-averaged absolute error* (6.3) in each voxel for 'deformed' and 'spotshift' respectively. The absolute errors are averaged over all 21 scenarios. Figure (a) and (c) show the *scenario-averaged error* (6.2) in each voxel. For these figures, the error signs are preserved so that averaging over scenarios cancels random errors but preserves *systematic errors*.



(a) 'deformed': average error

(b) 'deformed': average absolute error

(c) 'spotshift': average error

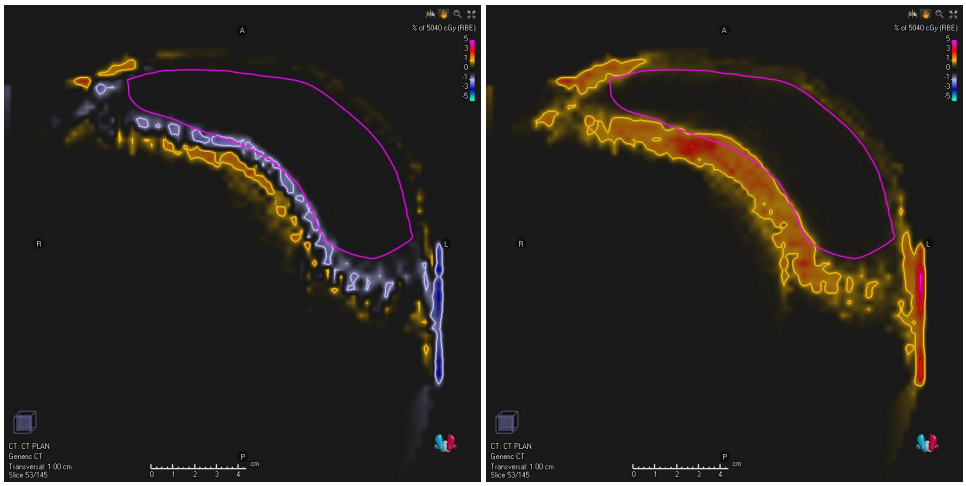
(d) 'spotshift': average absolute error

Figure 7.12 Lung II: average voxel-wise deviations from the 'accurate' ground truth. The color bar covers the range $(-5\,\%, 5\,\%)$ of the prescription dose.

Based on Figure 7.12, we observe that the average absolute error is larger for 'deformed' than for 'spotshift' as expected from the histograms in Figure 7.8. The largest errors occur right behind the tumor (the two beams enter the patient from north and north-west in the figures), where both methods systematically underestimate the dose. Furthermore, we observe that the errors of 'deformed' are concentrated at the end of the beam range where dose gradients are high, but it performs better than 'spotshift' nearer the surface. For 'deformed', there is a sharp decrease in accuracy where the beam enters the lung tissue, an effect not as apparent for 'spotshift'. Yet another observation is that the errors within the CTV are generally small for 'deformed'. This is maybe a result of the specific spot MU values x^* used in this experiment. Since the plan is optimized for uniform CTV coverage, the dose is uniform in this region which masks errors in the 'deformed' dose.

Plotting a similar plot for 'Breast I' is not as easy due to the organ motion scenarios. Averaging

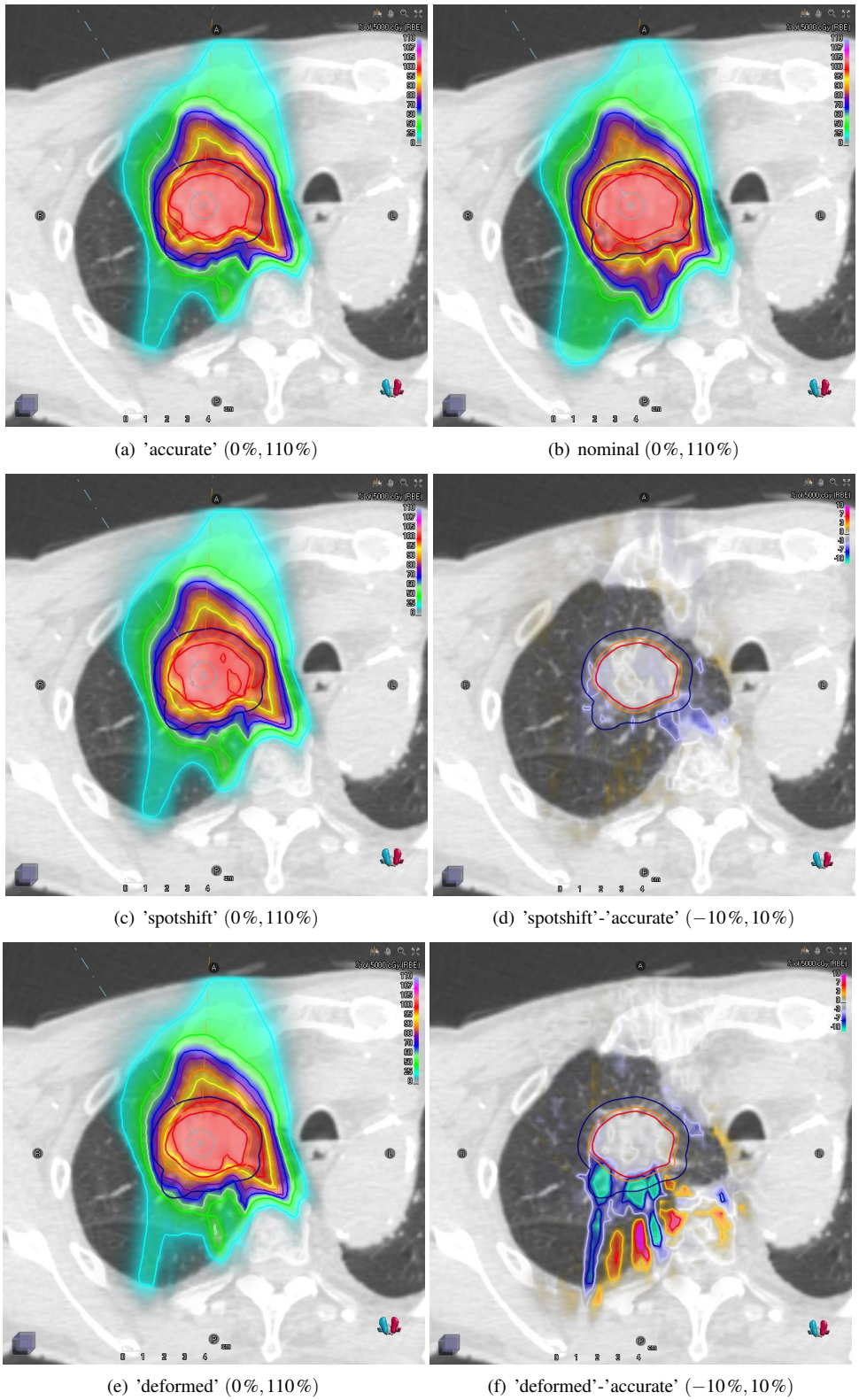
errors over the scenarios voxel-by-voxel is then not strictly meaningful. Proceeding with the calculations anyway leads to the results in Figure 7.13. We note that practically all errors occur outside the CTV, in which the dose is uniform. Again, 'deformed' systematically overestimates the dose at the end of the range and underestimates it right before. This corresponds to overestimating the depth of the Bragg peaks for the most energetic spots. Based on experiments on single spots, such as Figure 7.2 on page 48, a systematic misjudgment of the range seems unlikely. However, in that experiment it was noted that 'deformed' fails to model the magnitude of the Bragg peak dose perfectly. It should also be said that the systematic errors seen in Figure (a) are small, approximately $\pm 2\,\%$ of the prescription dose, except in the bottom right as previously discussed.



(a) 'deformed': average error (b) 'deformed': average absolute error

Figure 7.13 Breast I: average voxel-wise deviations from the 'accurate' ground truth. The color bar covers the range $(-5\,\%, 5\,\%)$ of the prescription dose.

Finally, to show how systematic errors appear in individual scenario doses, three examples are shown for 'Lung II'. The scenarios are chosen as the worst (Figure 7.14), median (Figure 7.15), and best (Figure 7.16) scenarios for 'deformed' in terms of RMSE.

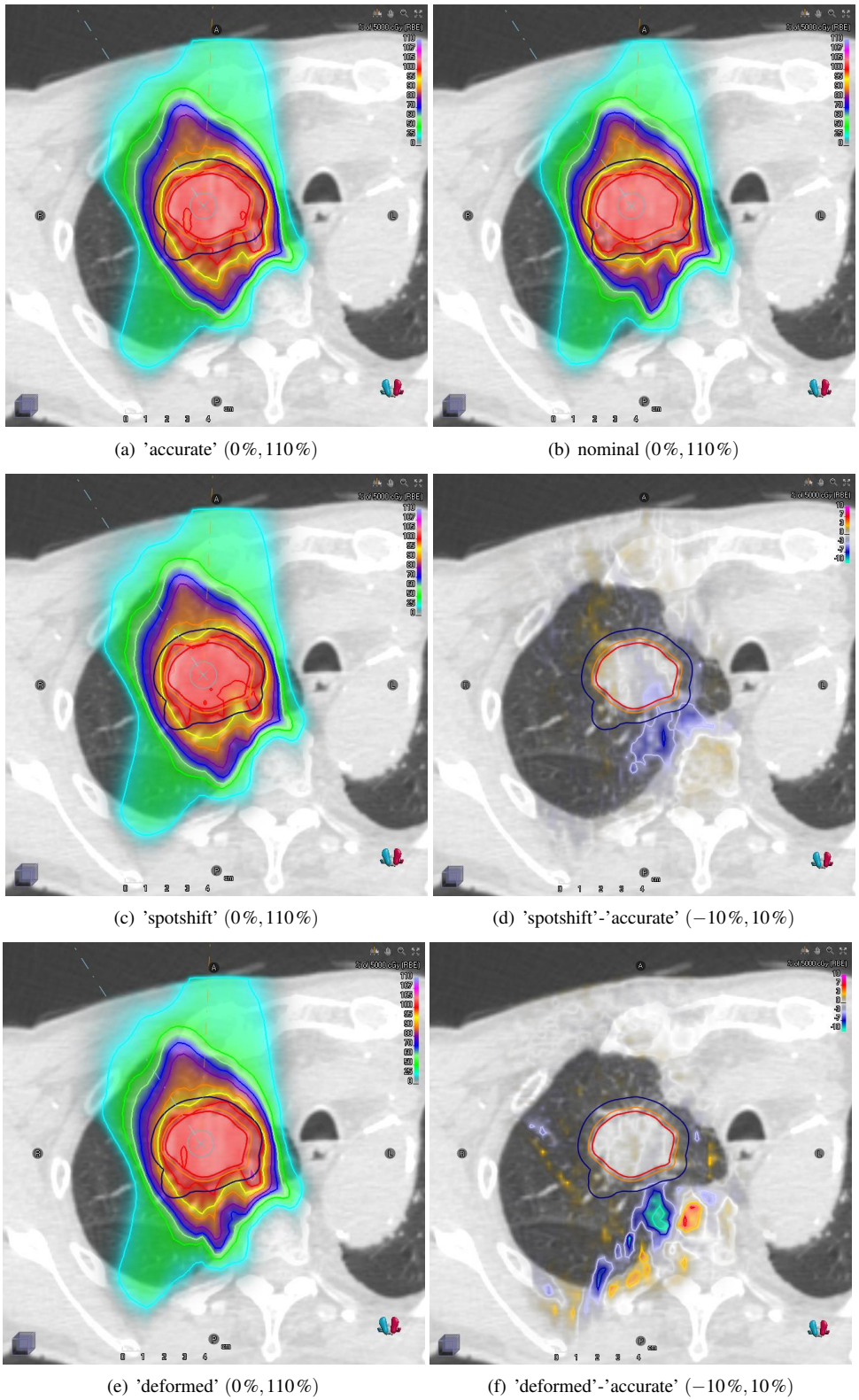


(a) 'accurate' $(0\,\%, 110\,\%)$ (b) nominal $(0\,\%, 110\,\%)$

(c) 'spotshift' $(0\,\%, 110\,\%)$ (d) 'spotshift'-'accurate' $(-10\,\%, 10\,\%)$

(e) 'deformed' $(0\,\%, 110\,\%)$ (f) 'deformed'-'accurate' $(-10\,\%, 10\,\%)$

Figure 7.14 'Lung II': the **worst** scenario for 'deformed' in terms of RMSE. The scenario consists of a density shift of $+3\,\%$ plus a patient shift of 0.5 cm to the left in the figure. Numbers in parenthesis indicate the range of the color bar in percentages of the prescription dose. Note the large errors of 'deformed' behind the tumor.

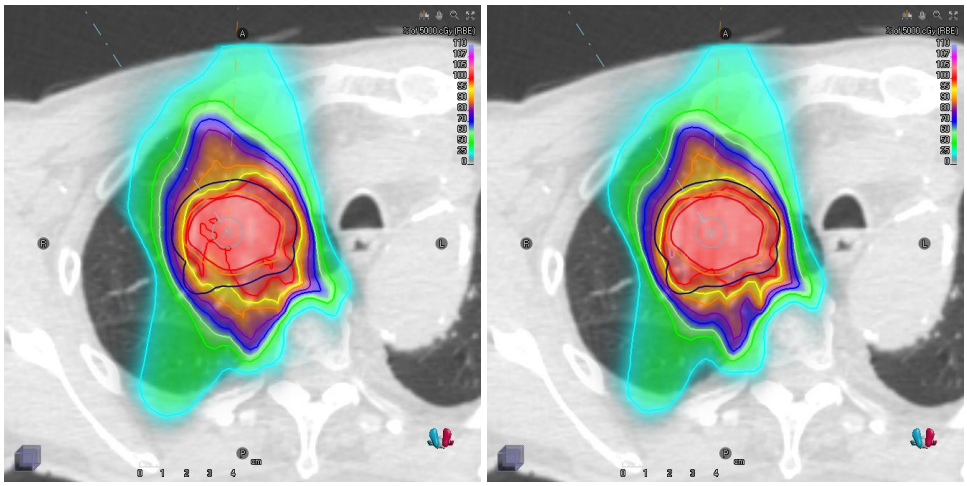


(a) 'accurate' $(0\,\%, 110\,\%)$ (b) nominal $(0\,\%, 110\,\%)$

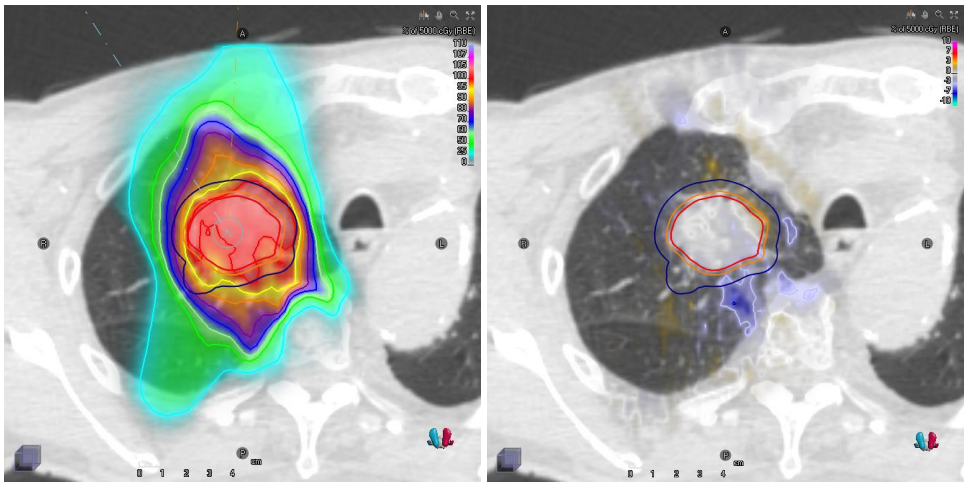
(c) 'spotshift' $(0\,\%, 110\,\%)$ (d) 'spotshift'-'accurate' $(-10\,\%, 10\,\%)$

(e) 'deformed' $(0\,\%, 110\,\%)$ (f) 'deformed'-'accurate' $(-10\,\%, 10\,\%)$

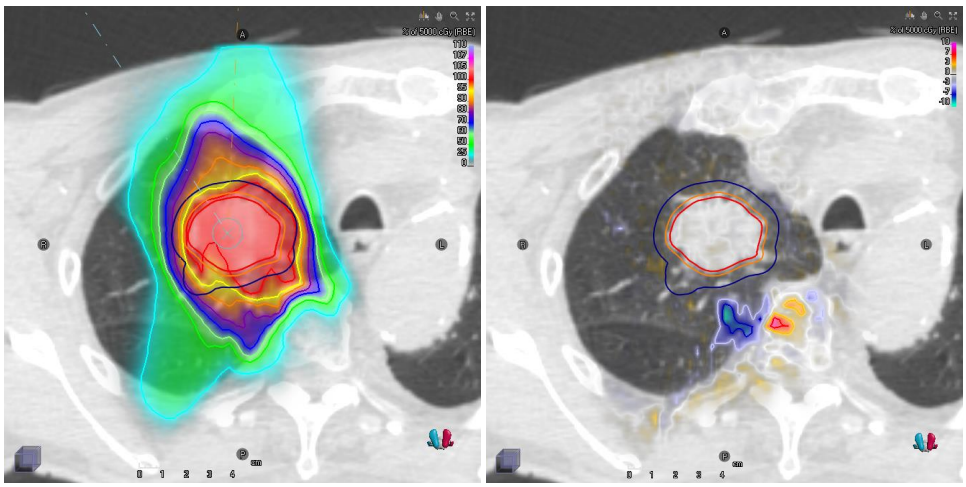
Figure 7.15 'Lung II': the **median** scenario for 'deformed' in terms of RMSE. The scenario consists of a density shift of $-3\,\%$ plus a patient shift of 0.5 cm upwards in the figure. Numbers in parenthesis indicate the range of the color bar in percentages of the prescription dose. Note the more significant errors of 'deformed' behind the tumor.



(a) 'accurate' $(0\,\%, 110\,\%)$ (b) nominal $(0\,\%, 110\,\%)$



(c) 'spotshift' $(0\,\%, 110\,\%)$ (d) 'spotshift'-'accurate' $(-10\,\%, 10\,\%)$



(e) 'deformed' $(0\,\%, 110\,\%)$ (f) 'deformed'-'accurate' $(-10\,\%, 10\,\%)$

Figure 7.16 'Lung II': the **best** scenario for 'deformed' in terms of RMSE. The scenario consists of a density shift of $-3\,\%$ plus a patient shift of 0.5 cm inwards in the figure. Numbers in parenthesis indicate the range of the color bar in percentages of the prescription dose. Here, the accuracy difference between 'spotshift' and 'deformed' is less pronounced.

7.3 Robust optimization

Clarification

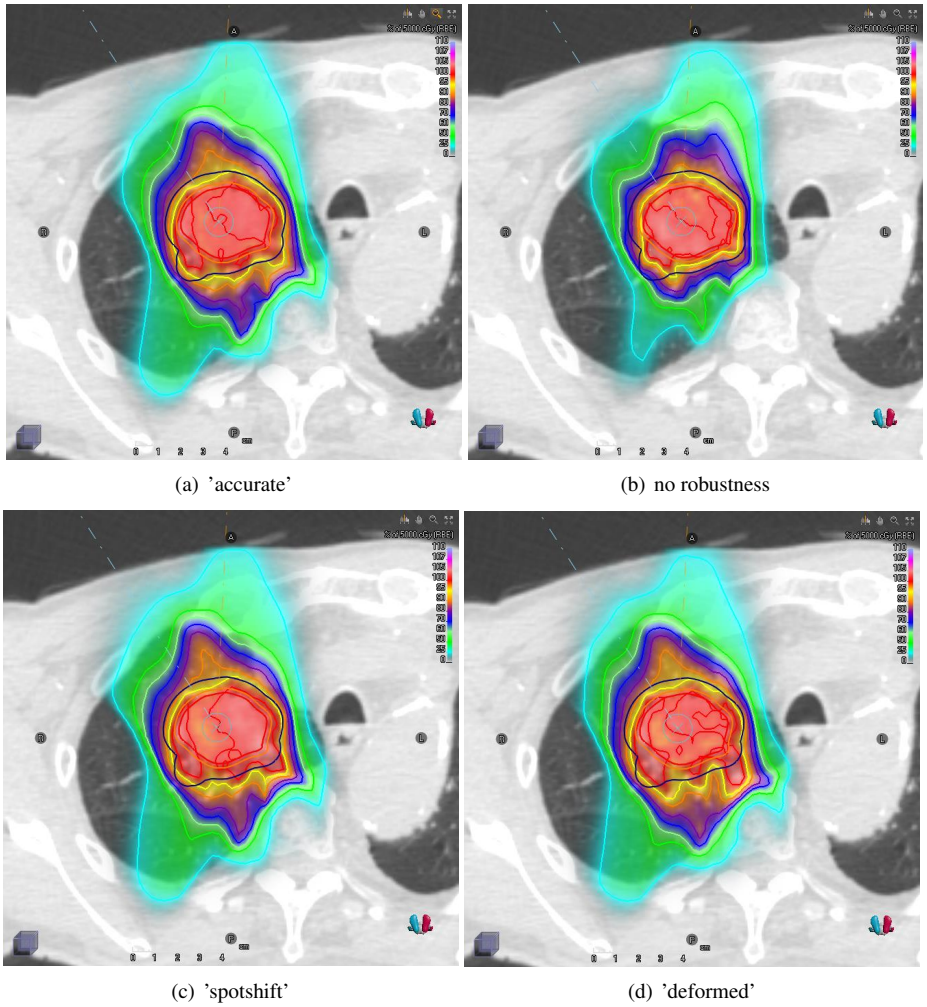
All experiments in this thesis aim to assess scenario dose deposition matrices D_s for scenarios $s \in \mathcal{S}$ calculated by various dose calculation algorithms. Before presenting the robust optimization results, it is worth to emphasize an important distinction between these results and the dose calculation accuracy results in Section 7.2.

For the accuracy results in Section 7.2, we assess the quality of the doses calculated by algorithm A by comparing $d^{\mathrm{A}} = D_s^{\mathrm{A}} x^*$ to the 'accurate' $d^{\mathrm{MC}} = D_s^{\mathrm{MC}} x^*$ for some *fixed* variables x^*. In the robust optimization experiments, on the other hand, we solve one version of the optimization problem (3.17) for each algorithm A (for each version of D_s^{A}) and arrive at *algorithm-dependent* solutions x_{A}^*. We are now interested in assessing the quality of these solutions.

Meaningfully comparing solutions x_{A}^* cannot be done directly in the domain of optimization variables. Instead, we investigate the resulting dose distributions. To achieve this, D_s^{MC} is used to transform the solutions into the dose domain according to (3.12). Thus, **all doses displayed in Section 7.3 are calculated by 'accurate'**. This includes the evaluation doses for which DVHs are plotted. The algorithms 'deformed' and 'spotshift' appear only in the alternative formulations of the optimization problem (3.17) that leads to alternative solutions x_{A}^*.

Plan dose comparison for the nominal scenario

Figure 7.17 shows the nominal dose for the four optimized 'Lung II' treatment plans.



(a) 'accurate' (b) no robustness

(c) 'spotshift' (d) 'deformed'

Figure 7.17 Lung II: Nominal dose for the optimized treatment plans which are either robustly optimized using 'accurate', 'spotshift', or 'deformed' scenario doses, or non-robustly optimized on the nominal scenario only. The color bar covers the range $(0\,\%, 110\,\%)$ of the prescription dose.

The non-robust optimization leads to the plan dose seen in Figure 7.17(b), and robust optimization with 'accurate' scenario doses to the one in (a). The non-robustly optimized plan clearly confines the high-dose region to the CTV, with isodose lines neatly following the CTV contour. The robustly optimized plan also covers the CTV, but the high-dose region is also slightly expanded in an irregular way. This can be seen by comparing the yellow 95 % isodose curves. A side effect of the robust optimization is that the healthy lung tissue in (a) receives a higher dose. However, CTV coverage in all scenarios has here been prioritized by the planner via the choice of objective functions and their relative weights.

Figures 7.17(c)-(d) show the nominal plan dose for treatment plans optimized with 'spotshift' and 'deformed' scenario doses. Although there are some differences, both these plan doses are qualitatively similar to their 'accurate' counterpart in (a) and have less in common with the non-robustly optimized plan in (b).

Robust DVH evaluation

To assess the plan quality for all scenarios $s \in \mathcal{S}$, we move on to study DVH curves for the CTV. Figure 7.18 shows such DVH curves for all four 'Lung II' plans. Note that the x-axis does not begin at zero. Each plan is summarized by three DVH curves. The thick line corresponds to the nominal scenario and the two thinner curves to the 'Lung II' scenarios in which the CTV receives the smallest/largest amount of dose.

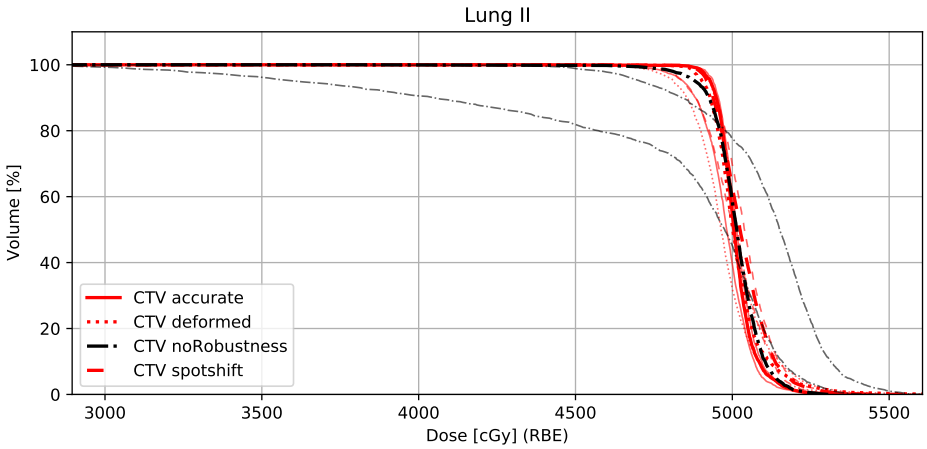


Figure 7.18 Lung II: DVH curves for the optimized treatment plans in Figure 7.17. Three curves are shown for each plan: for the nominal scenario and for the scenarios with least/most dose to the CTV.

Figure 7.18 shows that the shape of the DVH curves for the non-robustly optimized plan (black) varies heavily across scenarios. This is not surprising since the plan is optimized only for the nominal scenario. In Figure 7.17(b), we saw how this led to a high-dose region that conforms tightly to the CTV. It is easy to imagine how this high-dose region moves in and out of the CTV depending of the patient setup shift or density shift, drastically changing the dose statistics within the CTV.

The DVH curves for the three robustly optimized plans are quite similar, and a zoomed-in version is shown in Figure 7.19 for easier comparison. The 'accurate' DVH curves show the least amount of variation across scenarios, which is expected since the plan is optimized using the same dose deposition matrix and scenarios as is now used for evaluation. Evaluated on the nominal scenario, the plan optimized with 'deformed' doses seems slightly more similar to the 'accurate' plan than its 'spotshift' counterpart. However, differences are small and we simply conclude that all three methods give comparable results in this case.

The DVH curves for the 'Lung I' plans reveal similar results and can be seen in Appendix B. Interestingly, the DVH curves for 'Breast I' in Figure 7.20 seem to indicate that the robust optimization has no significant effect in this case as the curves for the four plans ('accurate', 'noRobustness', 'spotshift', 'deformed') follow each other closely. However, the differences are hidden due to the large volume of the CTV, and the robust optimization does in fact have an impact on the dose in voxels near the edge of the CTV. As follows from the square-cube law, surface voxels have less impact on volume statistics for ROIs of larger volume.

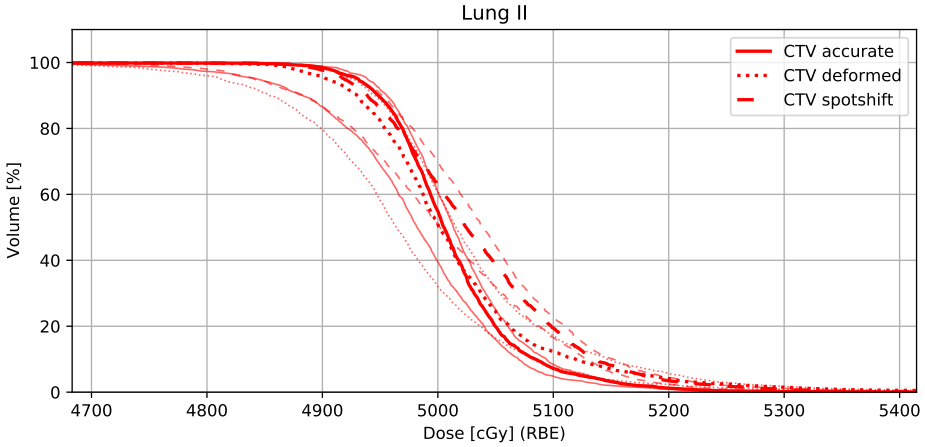


Figure 7.19 Lung II: Zoomed version of Figure 7.18, excluding 'noRobustness' for clarity. Three curves are shown for each plan: for the nominal scenario and for the scenarios with least/most dose to the CTV.

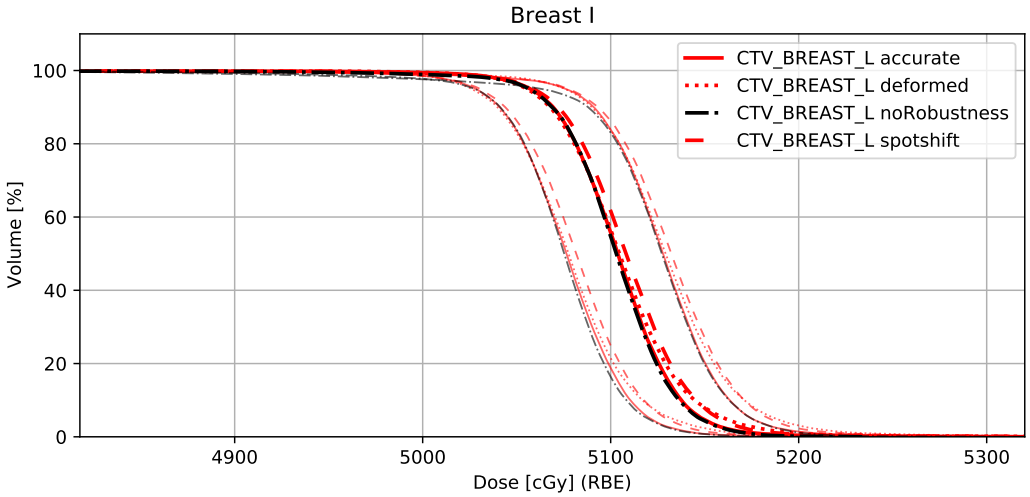


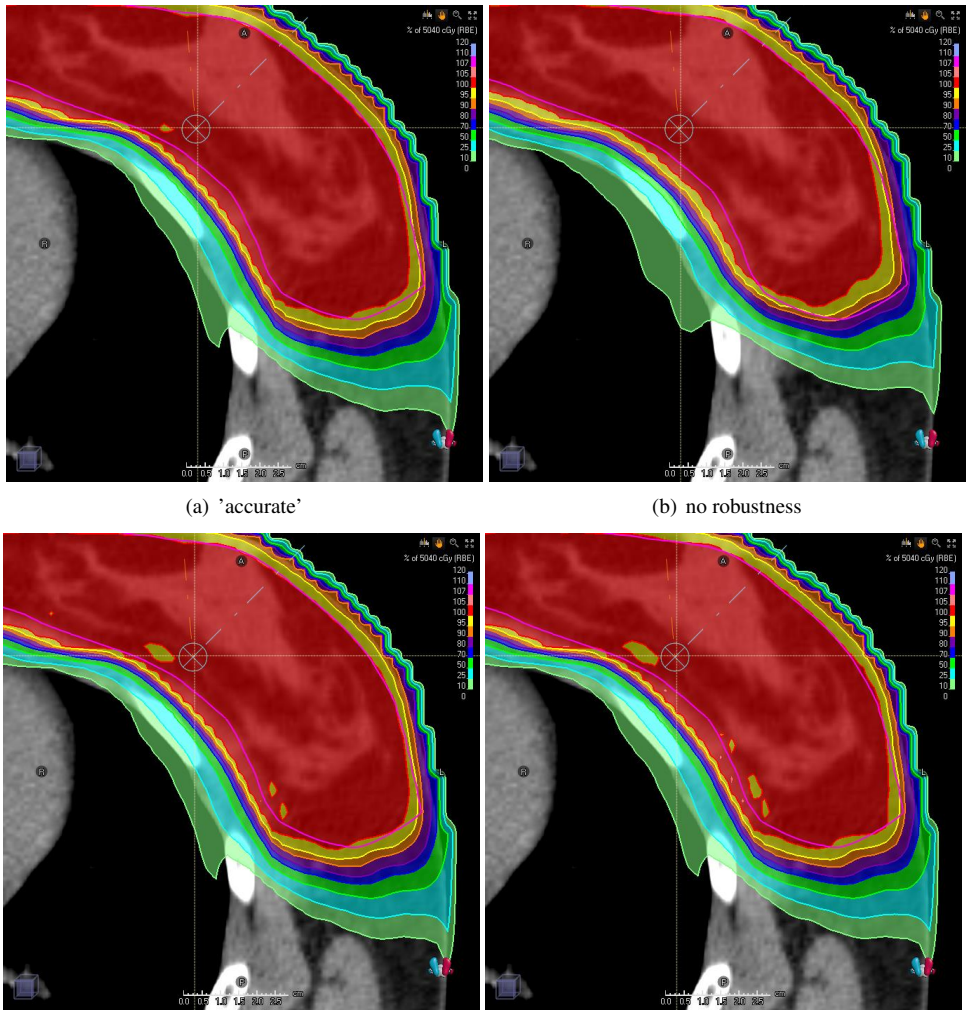
Figure 7.20 Breast I: DVH curves for the optimized treatment plans. Three curves are shown for each plan: for the nominal scenario and for the scenarios with least/most dose to the CTV.

To illustrate the effect of robust optimization for 'Breast I', we reuse the organ motion scenario illustrated in Figure 7.10(c), where the 'Breast I' CTV moves 0.5 cm, and evaluate the four plans on this scenario. The result is shown in Figure 7.21. Comparing the non-robust plan in (b) with the robustly optimized 'accurate' plan in (a), we observe a small difference in the lower right edge of the CTV (pink outline). The non-robust plan fails to deliver the prescribed dose to this corner of the ROI, whereas all three robustly optimized plans (a), (c), (d) better retain uniformity of the dose within the full volume. Again, we note that the quality and appearance of the three robustly optimized treatment plans are very similar.

Summary of optimization results

The robust optimization results indicate that the choice between 'accurate', 'spotshift', and 'deformed' for scenario dose calculation has little impact on the quality of the robustly optimized treatment plans. The three robustly optimized treatment plans resemble each other closely and show similar increases in robustness compared to the non-robust baseline.

These positive results are somewhat surprising given the results in Section 7.2 which reveal some inaccuracies of 'deformed' especially. However, we recall that high accuracy is observed *within the CTV* for both 'spotshift' and 'deformed'.



(a) 'accurate' (b) no robustness

(c) 'spotshift' (d) 'deformed'

Figure 7.21 'Breast I' Evaluation of the optimized plan under the organ motion scenario illustrated in in Figure 7.10(c). The color bar covers the range $(0\,\%, 120\,\%)$ of the prescription dose. We see that the uniform high-dose region better cover the full CTV (see bottom right of the pink outline) for all three robustly optimized plans.

For all patient test cases in this thesis, the robust objectives are defined on the CTV only. This means that, for all scenarios except the nominal, the dose outside the CTV has no effect whatsoever on the objective value in (3.17). Mathematically speaking, only the rows $i \in \mathcal{R}_{\text{CTV}}$ in D_s have an impact for scenarios $s \in \mathcal{S} \setminus \{s_0\}$ due to the design of the objective functions.

This fact gives one possible explanation for the success of 'deformed' in the optimization experiments. As seen, *e.g.*, in Figure 7.12 on page 58 for 'Lung II', neither 'deformed' nor 'spotshift' shows noteworthy errors within the CTV. However, in the volume behind the tumor, both show systematic errors of up to three percentage points and average absolute errors of up to five percentage points in some voxels. If the full objective would include robust objectives defined on ROIs in which the approximate dose calculation methods are less accurate, it is possible that we would observe larger discrepancies between the three robustly optimized treatment plans. Investigation of such cases is left for future work.

7.4 Computational efficiency

This thesis mainly investigates the *accuracy* of the scenario dose calculation algorithms being compared. Such results are presented in Sections 7.1-7.3. Sufficient accuracy is a necessary requirement for an algorithm being of practical use. However, the only reason in practice for considering 'spotshift' or 'deformed' over 'accurate' is potential improvements in terms of *computational efficiency*. We always expect 'accurate' (Monte Carlo) to provide the most accurate doses.

Computational efficiency in this context refers to the time and memory usage of robust optimization in RayStation. If the scenario doses require less memory to store, a larger number of scenarios can be considered when memory is the bottleneck. Quicker dose calculations allow for including more scenarios when runtime is the bottleneck, for iteratively running more optimizations per patient with adjusted objectives, or for planning treatments for a larger number of patients each day.

Runtime

The scenario dose approximation methods influence the robust optimization runtime both in terms of time needed to calculate scenario dose deposition matrices D_s and time needed to evaluate objective functions and gradients in each SQP iteration. Objective functions are usually defined on the dose, and gradients are transformed by the dose deposition matrix as described in (3.18).

Empirical runtime measurements are presented in Table 7.1. When interpreting the results, it is important to remember that efficiency was not of primary concern when implementing 'deformed' for the purpose of this thesis. As will soon be discussed, significant runtime improvements can be achieved by some straightforward caching. Furthermore, 'deformed' runs on CPU whereas the calculation of spot doses with 'accurate' runs on GPU. For 'accurate', changing from CPU to GPU has been quoted to reduce computational times by a factor of 10 [RaySearch Laboratories AB, 2020]. For 'deformed', ray tracing is currently the most time-consuming task, a task that lends itself perfectly for parallel computing. Thus, one may expect that 'deformed' would benefit greatly from being ported to GPU.

Table 7.1 Empirical runtimes of the optimizations. For each case and algorithm, runtime is measured in seconds and presented as 'X (Y)', where X is the runtime before the SQP iterations starts (including initialization of variables and calculation of D_s) and Y is the average runtime per outer iteration of the SQP optimization. The lacking SQP iteration runtime for the combination of 'Lung Ia' and 'deformed' is explained in the text.

Case		**Runtime [s] of initialization (SQP iteration)**					
Name	$\lvert\mathcal{S}\rvert$	'accurate'		'spotshift'		'deformed'	
Lung Ia	231	874	(9.4)	549	(7.9)	3756	(?.?)
Lung Ib	21	28	(0.7)	9	(0.8)	300	(2.3)
Lung Ic	21	121	(1.3)	102	(0.9)	404	(2.9)
Lung II	21	34	(0.9)	9	(1.2)	368	(2.7)
Breast I	21	679	(37.0)	236	(13.8)	754	(6.2)

Cases with few spots. In Table 7.1, we note that using 'accurate' for 'Lung Ib', 'Lung Ic' and 'Lung II', leads to initialization phases which takes at most two minutes. Each SQP iteration runs in a few second at most. For these cases, further speed-ups may not really be needed, but we note that 'spotshift' is even quicker. Referring back to Table 6.2 on page 42, one could expect that 'Lung Ic' would run quicker than 'Lung Ib' and 'Lung II' due to the smaller number of spots. However, the organ motion scenarios introduce some extra complexity during initialization.

In 'Lung Ia', the number of scenarios are increased, and the runtimes are scaled accordingly. For this case, with 231 scenarios, initialization and dose calculation takes almost 15 minutes for 'accurate' and a full hour for 'deformed'. It can be noted that the current implementation of 'deformed' does the ray tracing in the nominal scenario 230 times: once for each non-nominal scenario. This number could be reduced to one by simply caching the nominal ray tracing results. This fix would effectively cut ray tracing costs in half but was not implemented due to time constraints. And again, 'deformed' currently runs on CPU, not GPU. For 'Lung Ia', no SQP iteration runtimes are displayed since the program crashed after running out of memory. This is discussed further in the upcoming section.

Using the current implementation of 'deformed' slows down the robust optimization across all lung cases. This includes both dose calculation and SQP iterations. The likely explanation is the

small number of spots ($N \leq 3000$). The number of spots appears as the number of columns in the scenario dose deposition matrices $D_s \in \mathbb{R}^{|\mathcal{V}| \times N}$. Having few columns is beneficial both for matrix multiplications in dose evaluations (3.2) and for matrix multiplications in gradient transformations (3.18). When using 'deformed', both dose evaluations and gradient transformations require multiplying some vector with the dose mapping matrices $A_{bs} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ (and their transposes) as seen in (5.2) and (5.4). Although these matrices are sparse compared to D_s, it is not surprising that 'deformed' has a disadvantage when $N \ll |\mathcal{V}|$.

Cases with many spots. For the actual calculation of scenario dose deposition matrices D_s, the complexity for both 'accurate' and 'spotshift' scales with the number of spots. The same is not true for 'deformed'. For 'deformed', the D_s calculation runtime scales with the number of ray tracing rays, which in turn depends on the number of beams, the *area* spanned by the spots in the isoplane, and the resolution of the dose grid. Thus, except for influencing the 'accurate' nominal dose calculation and to some degree the sparsity of A_{bs}, **the number of energy layers and spatial density of spots have little impact on the runtime of 'deformed'**.

The SQP iteration runtime for 'deformed' depends mainly on the number of dose grid voxels which affects the size of the dose mapping matrices A_{bs}. Note that the number of rays has no influence on the SQP iteration runtime if the dose mapping matrices are cached.

The 'Breast I' runtimes are best compared with their 'Lung Ic' counterparts. Both cases include 21 scenarios, some of which include organ motion. Their most significant difference lies in the number of spots. As seen in Table 6.2 on page 42, 'Breast I' has 51768 spots after optimization whereas 'Lung Ic' has 613. For both 'accurate' and 'spotshift', the 'Breast I' initialization phase is several times slower than the 'Lung Ic' analogue. The way the runtime scales with the number of spots is masked by their common overhead related to handling dose calculation for multiple images (*cf.* the difference in initialization runtime between 'Lung Ib' and 'Lung Ic'). The SQP iteration runtimes, however, clearly reveal huge differences. For instance, the average SQP iteration takes 13.8 s in the 'Breast I' case and 0.9 s in the 'Lung Ic' case while using 'spotshift' for scenario doses.

For 'deformed', both runtime measures increase by approximately a factor of two in the same comparison. The difference is not explained by the number of spots but rather by the size of the target volumes. For 'Breast I', the irradiated volume inhibits a much larger solid angle as seen from the beam sources. The ray trace bixel grid must be extended accordingly which increases the number of rays and the time spend on ray tracing. This target-size effect on the number of rays (and on runtime) is partly counterbalanced by the fact that 'Breast I' has only two beams whereas 'Lung Ic' has three.

Summary. When the number of spots is small, the current implementation of 'deformed' is slower than 'accurate' or 'spotshift' for robust optimization. It should however be emphasized that the current implementation has not been equally optimized for performance due to thesis time constraints.

The runtime of 'deformed', unlike for 'spotshift' and 'accurate', does not scale with the number of spots, but rather their covered area in the isocenter plane. Thus, 'deformed' shows competitive performance when the number of spots is high. For the current implementation, the largest improvements are seen in terms of runtime of the outer SQP iterations. The initial dose calculation is also expected to be significantly quicker for 'deformed' if proper caching and GPU parallelized ray tracing is implemented.

Memory footprint

No empirical memory comparisons are done due to thesis time constraints. However, we may still do some theoretical analysis.

Accurate. As illustrated in Figure 7.5(a) on page 50, the dose from a single spot is delivered to quite a large number of voxels. Each column in $D_s \in \mathbb{R}^{|\mathcal{V}| \times N}$ contains the corresponding number of non-zero elements. Since most column elements are still zero, a sparse format is used which also exploits the spatial coherence of non-zero elements.

For 'accurate', a separate D_s is stored for each scenario $s \in \mathcal{S}$. As exemplified by 'Lung Ia', the number of scenarios may be a few hundred. Storing all matrices requires significant memory resources. Calculating them anew in each SQP iteration is not an option since it takes too much time.

The memory needed to store each D_s is roughly linear in the number of spots, assuming all spot doses are delivered to a roughly constant number of voxels. As exemplified by 'Breast I', the number of spots may exceed 50000. The memory footprint of D_s depends also on the dose grid resolution.

Halving the voxel size in all three dimensions means increasing the number of non-zero elements in D_s (and the memory footprint) by a factor of 8.

Spotshift. The scenario dose approximation method 'spotshift' reuses the columns in D_{s_0} to approximate the columns in D_s for combinations of *density shift* and *setup shift* scenarios by means of interpolation. To also handle organ motion, 'spotshift' relies on one 'accurate' calculation of D_k for each pure organ motion scenario $k \in \mathcal{S}$. Combination scenarios of density shifts, setup shifts and organ motion scenario k are then handled by interpolating the columns of D_k.

As discussed in Section 4.3 on page 33, extra *virtual* spots must be calculated for 'spotshift' to properly handle the edge spots. This means expanding each energy layer to handle setup shifts perpendicular to the beam and adding new energy layers to handle density shifts of the least/most energetic spots. The result of adding virtual spots is extending the number of columns in each D_k. If adding the virtual spots increases the number of spots by a factor V and there are K pure organ motion scenarios, using 'spotshift' decreases the memory footprint by a factor $|\mathcal{S}|/KV$.

If rotational scenarios would be considered, these would probably require the same special treatment as the organ motion scenarios.

Deformed. Like 'spotshift', 'deformed' relies on the 'accurate' D_{s_0} for the nominal scenario. However, handling scenarios of organ motion and rotation is part of the method and requires no additional 'accurate' calculations.

The memory footprint of 'deformed' depends instead on the storage of the dose-mapping matrices $A_{bs} \in \mathbb{R}^{|\mathcal{V}|\times N}$. The number of such matrices equals the product $|\mathcal{B}|(|\mathcal{S}|-1)$ between the number of beams and the number of scenarios (excluding the nominal). Generally, $|\mathcal{V}| \gg N$, which means that 'deformed' requires storage of roughly $|\mathcal{B}|$ times as many matrices as 'accurate', matrices which additionally are of larger dimension. Also, in the current implementation, A_{bs} and A_{bs}^T are stored separately, meaning that actually $2|\mathcal{B}|$ times as many matrices are needed.

However, A_{bs} is much sparser than D_s. As described in Chapter 5, most elements on a given row in A_{bs} are zero. It would probably even be possible to explicitly limit number of votes (5.7) per row to some fix number, say three, by only keeping the three top contributors of each row before normalizing as in (5.8). Such experiments are left for future work.

During the robust optimization experiments, the optimization of 'Lung Ia' unexpectedly crashed for 'deformed' after the first few SQP iterations. As it turned out, the very high number of scenarios combined with the current implementation of 'deformed' being less memory efficient than 'accurate' when the number of spots is small, led to the computer running out of memory.

The memory usage of 'deformed' can be halved by not storing A_{bs} and A_{bs}^T separately. The alternative should still allow for quick row and column access. Furthermore, only some rows in A_{bs} are actually needed. Specifically, rows with indices that correspond to voxels within ROIs on which robust objectives are defined. All other rows could equally well be set as zero without affecting optimization. This would have a huge impact for 'Lung I' since the CTV has volume $8\,\mathrm{cm}^3$ as compared, *e.g.*, to the full patient volume (within in the dose grid) of more than $20000\,\mathrm{cm}^3$.

Finally, in cases when storing the deformation matrices A_{bs} is less memory efficient than using 'accurate' and storing the full dose deposition matrices D_s, one could multiply the matrices for 'deformed' and store $D_{bs} = A_{bs}D_{bs_0}$ for $s \in \mathcal{S}$ instead. By doing so, 'deformed' could in theory always be at least as memory efficient as 'accurate' at the cost of the above matrix multiplication. However, it can be expected that this matrix multiplication mainly reduces the memory footprint when the number of spots is low (and D_{bs_0} is tall). As previously discussed, 'accurate' is already quite efficient when the number of spots is low. In this case there is less need to use an alternative method such as 'deformed'.

7.5 Summary

The experiments mainly investigate the *accuracy* of 'deformed', which is our implementation of a scenario dose calculation method in similar style as [Park et al., 2012]. In terms of *efficiency*, some empirical runtime measurements are provided as well as a theoretical analysis of the memory usage. The method 'deformed' is compared against the accurate Monte Carlo method ('accurate') and approximate spot weight shifting method ('spotshift') already implemented in RayStation.

Summary of accuracy results. In terms of accuracy, 'spotshift' consistently performs slightly better than 'deformed'. The absolute voxel-wise errors for 'spotshift' stay below 10 % of the prescription dose is all test cases. For 'deformed', the variance of voxel-wise errors is higher, meaning that errors of larger magnitude are more common. The difference is exaggerated by the log scale in the voxel error histograms. Considering only the scenarios where the methods are fairly compared (*i.e.*, setup shifts and density shifts), using 'deformed' over 'spotshift' increases the standard deviation of voxels errors (normalized to the prescription dose) from 0.36 % to 0.43 % for test case 'Lung Ib' and from 0.60 % to 0.91 % for 'Lung II'.

The third test case ('Breast I') reveals a fundamental limitation of 'deformed'. When the nominal dose cloud is located near the patient surface, deforming the dose may require fetching nominal dose from outside the patient volume – even from outside the dose grid. Such dose is not calculated by 'accurate'. Since the information is not present in the nominal dose, 'deformed' cannot determine the dose values in the affected voxels as illustrated in Figure 7.11 on page 57. Working around this issue is an interesting direction for further development.

Apart from the issue mentioned above, the accuracy of 'deformed' is only slightly decreased when adding organ motion scenarios to combinations of setup shifts and density shifts. Although in theory the method also extends to (small) rotations, an investigation of the accuracy is left for future work.

Comparing 'deformed' to analytical dose calculation algorithms, the ray casting model in Figure 4.2 shows interesting similarities. The theoretical advantage of 'deformed' over this other method is that 'deformed' utilizes the information in the nominal 'accurate' dose instead of deforming a generic dose kernel in water. Assuming the lateral scattering due to various density heterogeneities is roughly constant across scenarios, the accurate nominal dose might be a better predictor of the scenario dose, even if it is only deformed to the scenario by accounting for WEPL along rays from the beam source which, in itself, does not model lateral scattering. However, the nominal dose might also be a worse predictor when this assumption is not true. Verifying this is left for future work.

Summary of optimization results. Using the three different versions of scenario dose deposition matrices D_s (calculated by either 'accurate', 'spotshift' or 'deformed') in robust optimization (3.17) leads to similar results across all test cases. Quantifying robustness is not trivial, but all three versions of the optimized treatment plans show similar robust qualities when evaluated on the scenarios used in optimization. One explanation for the good results for 'deformed' could be that the robust objectives are only defined on the CTV in which the accuracy of all three methods is similar. Further research should also investigate cases where robust objectives are defined on OARs located behind the target since 'deformed' seems least accurate at the end of the beam range.

Summary of efficiency results. The efficiency of 'accurate' and 'spotshift' depends strongly on the number of spots. This is not the case for 'deformed', for which the computational cost instead scales with the number of virtual beam sources and the area covered by the spots in each isocenter plane. Further developments could reduce this area dependence by only running the ray trace depth integration for rays that intersect at least one ROI connected to a robust objective.

Given the above, we expect 'deformed' to be most useful in cases when the number of spots is high, especially if the spots are stacked densely over many energy layers. The implementation used in this thesis was not optimized for efficiency due to time constraints. To improve efficiency, the ray trace operation could be run in parallel on a GPU. Before proceeding in that direction, the format for storing the large but sparse (beam and scenario dependent) deformation matrices A_{bs} deserves some extra attention. Both efficient row and column access is important. Additional tuning of 'deformed' can be done by tuning the resolution of the ray trace bixel grid which strongly affects the number of rays.

8

Conclusion

8.1 Problem restatement

Proton therapy treatments must be robust to small uncertainties in the patient's anatomy and exact positioning in the treatment machine. In treatment planning systems, robust optimization is used to ensure sufficient plan quality in a user-defined set of scenarios, *i.e.*, sampled realizations of such uncertainties. However, since each scenario requires a separate dose calculation, the optimization procedure can be prohibitively slow if there are many scenarios. In treatment planning for pencil beam scanning this is especially true if the plan features a large number of spots since the runtime and memory footprint of dose calculations generally increase as the number of spots increases.

This thesis investigates alternative methods for scenario dose calculation, seeking efficiency improvements. Efficiency improvements may allow clinic personnel to spend more time interacting with patients. It may also allow treatment planners to account for a larger number of scenarios and to iteratively run more optimizations for each patient. Both things might improve the final plan quality. Another potential benefit is reduced time between patient imaging and treatment delivery.

8.2 Conclusion

Referring back to the research questions listed in Section 1.2 on page 12, the conclusions of this thesis can be summarized as follows:

I *Can a scenario dose calculation algorithm in the style of [Park et al., 2012] be incorporated into the robust optimization procedure of a commercial treatment planning system?*

An extended variant ('the proposed method') of the scenario dose calculation method in [Park et al., 2012] is successfully implemented in RayStation as described in Chapter 5. The method is extended to handle a wider class of scenarios which also includes density shifts and rotations.

II *Why does robust optimization efficiency depend on the choice of scenario dose calculation algorithm?*

The choice of scenario dose calculation method influences the efficiency of the robust optimization initialization phase during which the dose deposition matrices D_s are calculated. The choice of method may also influences the efficiency of each SQP iteration since the parameterization and storage of D_s affects the efficiency of dose and gradient evaluations.

III *Which scenario dose calculation algorithms are already available, and how can these be fairly evaluated and compared with the proposed method?*

RayStation currently implements two methods for scenario dose calculation for protons: one accurate Monte Carlo method and one approximate spot weight shifting method. Any method in Section 4.2 is also a competing method but the proposed method is only evaluated against the two mentioned above.

Fair comparison of competing methods require thorough qualitative and quantitative analysis. This thesis strives to provide as much qualitative information as possible about the strengths and weaknesses of the proposed method rather than relying heavily on advanced metrics.

IV *What are the strengths and weaknesses in terms of accuracy and efficiency of the proposed method compared to the algorithms that already exist?*

The proposed algorithm offers no improvements in terms of accuracy compared to the two methods already implemented in RayStation. However, the decrease in accuracy does not seem to adversely affect the robustness of the optimized treatment plans in the cases studied here. Further research is needed to verify this on additional cases, especially cases with robust objectives put on OARs behind the target volume since the proposed method appears to be least accurate at the end of the beam range. An important limitation of the current implementation of the method is that, as demonstrated in Figure 7.11 on page 57, it may fail near the surface of the patient in some scenarios.

Despite being less accurate, the proposed method may still be interesting in cases when efficiency of the alternatives is a limiting factor. The proposed method needs only one full Monte Carlo simulation regardless of the number of scenarios. The efficiency of the dose-mapping matrix calculation does not depend much on the number of energy layers or the density of spots in the isocenter plane, whereas the efficiency of the alternatives depends on these quantities. Thus, the proposed method may allow for significantly increased efficiency for plans with a large number of spots per beam.

V *How could the proposed method be improved?*

The runtime efficiency of the current implementation of the proposed method can probably be improved significantly by parallelizing the ray tracing to run on GPU. It can certainly be improved by caching the ray tracing results for the nominal scenario instead of re-doing the nominal calculation for each scenario.

The memory efficiency can be improved by storing dose-mapping matrices A_{bs} in a clever sparse format that still allows fast row and column access. The current implementation stores A_{bs} and A_{bs}^T as two separate CRS matrices. For cases with few spots, memory efficiency may be improved by evaluating and storing the matrix product $A_{bs}D_{bs_0}$ instead.

Both memory and runtime efficiency can be improved by only calculating and storing the rows of A_{bs} that correspond to voxels in ROIs with robust objectives.

We now return to the general aim of the thesis and recall the question:

> Can a scenario dose calculation algorithm in the style of [Park et al., 2012] improve robust optimization efficiency in proton therapy treatment planning?

The results in this thesis suggests that, yes, such a method can improve the efficiency for PBS treatments if the treatment plan contains a large number of spots per beam. This corresponds well to the situations when current methods are less efficient. Efficiency improvements must be weighed against the intrinsic loss of accuracy associated with an approximate dose calculation algorithm.

Quantifying efficiency improvements (where applicable) is beyond the scope of this thesis, although some numbers and discussion on the topic are still provided. Such quantification is however highly relevant to allow for proper trade-offs between methods in terms of accuracy and efficiency. The proposed method should only be considered when the benefit of improving efficiency outweighs the disadvantage of a reduced dose accuracy. The results presented here indicate that the reduced accuracy is tolerable for use in robust optimization, except in edge cases, but further research is needed to verify this claim. It should be noted that an approximate algorithm used in robust optimization can be coupled with an accurate algorithm for the final QA of the optimized treatment plan.

8.3 Outlook

This thesis only deals with proton therapy. However, the proposed method could be useful for other ion therapy modalities as well. For instance, carbon therapy treatment plans often feature a large number of spots, making the proposed method's weak efficiency-dependence on the number of spots an attractive quality.

Another possible use case is proton arc therapy, in which the gantry continuously rotates around the patient during treatment delivery [Battinelli, 2019]. For such a treatment technique, extending the approximate scenario dose calculation method to handle scenarios of small rotations is of natural interest. Also, the proposed method handles scenarios of density shifts without relying on interpolation between energy layers. This is beneficial for arcs with few energy layers per angle.

Bibliography

Abbema, J. van, M.-J. van Goethem, J. Mulder, A. Biegun, M. Greuter, A. van der Schaaf, S. Brandenburg, and E. van der Graaf (2018). "High accuracy proton relative stopping power measurement". *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **436**, pp. 99–106.

Andersson, B. (2017). *Mathematical optimization of radiation therapy goal fulfillment*. MA thesis. Uppsala universitet.

Bangert, M., P. Hennig, and U. Oelfke (2013). "Analytical probabilistic modeling for radiation therapy treatment planning". *Physics in Medicine & Biology* **58**:16, p. 5401.

Battinelli, C. (2019). *Proton Arc Therapy Optimization*. MA thesis. KTH Royal Institute of Technology.

Bokrantz, R. (2013). *Multicriteria optimization for managing tradeoffs in radiation therapy treatment planning*. PhD thesis. KTH Royal Institute of Technology.

Bokrantz, R. and A. Fredriksson (2014). *Reference manual for optimization*. RSL-D-50-08. RaySearch Laboratories AB.

Bokrantz, R. and A. Fredriksson (2017). "Scenario-based radiation therapy margins for patient setup, organ motion, and particle range uncertainty". *Physics in Medicine & Biology* **62**:4, p. 1342.

Canfield Upton, A., A. Mozumder, M. Burton, C. A. Tobias, M. Luntz, and J. Silverman (2021). *Radiation*. URL: `https://www.britannica.com/science/radiation` (visited on 2021-05-14).

Fredriksson, A. (2012). "A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization". *Medical physics* **39**:8, pp. 5169–5181.

Fredriksson, A. (2013). *Robust optimization of radiation therapy accounting for geometric uncertainty*. PhD thesis. KTH Royal Institute of Technology.

Gill, P. E., W. Murray, and M. A. Saunders (2005). "Snopt: an sqp algorithm for large-scale constrained optimization". *SIAM review* **47**:1, pp. 99–131.

Hogstrom, K. R., M. D. Mills, and P. R. Almond (1981). "Electron beam dose calculations". *Physics in Medicine & Biology* **26**:3, p. 445.

Holmes, T., C. Ma, and L. Wang (2013). "Monitor unit". In: Brady, L. W. et al. (Eds.). *Encyclopedia of Radiation Oncology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 515–515. ISBN: 978-3-540-85516-3. DOI: `10.1007/978-3-540-85516-3_351`. URL: `https://doi.org/10.1007/978-3-540-85516-3_351`.

Joiner, M. C. and A. J. van der Kogel (2018). *Basic clinical radiobiology*. CRC press.

Kooy, H. and C. Grassberger (2015). "Intensity modulated proton therapy". *The British journal of radiology* **88**:1051, p. 20150195.

Krane, K. S., D. Halliday, et al. (1988). *Introductory nuclear physics*. Vol. 465. Wiley New York.

Lomax, A. (2018). "What will the medical physics of proton therapy look like 10 yr from now? a personal view". *Medical physics* **45**:11, e984–e993.

Low, D. A., W. B. Harms, S. Mutic, and J. A. Purdy (1998). "A technique for the quantitative evaluation of dose distributions". *Medical physics* **25**:5, pp. 656–661.

Miller, K. D., L. Nogueira, A. B. Mariotto, J. H. Rowland, K. R. Yabroff, C. M. Alfano, A. Jemal, J. L. Kramer, and R. L. Siegel (2019). "Cancer treatment and survivorship statistics, 2019". *CA: a cancer journal for clinicians* **69**:5, pp. 363–385.

Mohan, R., I. J. Das, and C. C. Ling (2017). "Empowering intensity modulated proton therapy through physics and technology: an overview". *International Journal of Radiation Oncology* Biology* Physics* **99**:2, pp. 304–316.

Nelms, B. E., M. F. Chan, G. Jarry, M. Lemire, J. Lowden, C. Hampton, and V. Feygelman (2013). "Evaluating imrt and vmat dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels". *Medical physics* **40**:11, p. 111722.

Newhauser, W. D. and R. Zhang (2015). "The physics of proton therapy". *Physics in Medicine & Biology* **60**:8, R155.

Nocedal, J. and S. Wright (2006). *Numerical optimization*. Springer Science & Business Media.

Paganetti, H. (2012). "Range uncertainties in proton therapy and the role of monte carlo simulations". *Physics in Medicine & Biology* **57**:11, R99.

Paganetti, H. (2018). *Proton therapy physics*. CRC press.

Park, P. C., J. Cheung, X. R. Zhu, N. Sahoo, L. Dong, et al. (2012). "Fast range-corrected proton dose approximation method using prior dose distribution". *Physics in Medicine & Biology* **57**:11, p. 3555.

Petti, P. L. (1992). "Differential-pencil-beam dose calculations for charged particles". *Medical physics* **19**:1, pp. 137–149.

Philip, E., W. MURRAY, M. A. SAUNDERS, and M. H. Wright (2001). "User's guide for npsol 5.0: a fortran package for nonlinear programming". *Technical report, Technical report SOL 866*.

PTCOG (2021). *Particle therapy facilities in clinical operation*. URL: `https://www.ptcog.ch/index.php/facilities-in-operation` (visited on 2021-05-31).

RaySearch Laboratories AB (2020). *New release of groundbreaking treatment planning system raystation provides major advances for proton pbs monte carlo with gpu*. URL: `https://www.raysearchlabs.com/media/press-releases/2020/new-release-of-groundbreaking-treatment-planning-system-raystation-provides-major-advances-for-proton-pbs-monte-carlo-with-gpu/` (visited on 2021-06-04).

Saini, J., D. Maes, A. Egan, S. R. Bowen, S. St James, M. Janson, T. Wong, and C. Bloch (2017). "Dosimetric evaluation of a commercial proton spot scanning monte-carlo dose algorithm: comparisons against measurements and simulations". *Physics in Medicine & Biology* **62**:19, p. 7659.

Saini, J., E. Traneus, D. Maes, R. Regmi, S. R. Bowen, C. Bloch, and T. Wong (2018). "Advanced proton beam dosimetry part i: review and performance evaluation of dose calculation algorithms". *Translational lung cancer research* **7**:2, p. 171.

Schaffner, B., E. Pedroni, and A. Lomax (1999). "Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation". *Physics in Medicine & Biology* **44**:1, p. 27.

Taylor, P. A., S. F. Kry, and D. S. Followill (2017). "Pencil beam algorithms are unsuitable for proton dose calculations in lung". *International Journal of Radiation Oncology* Biology* Physics* **99**:3, pp. 750–756.

Unkelbach, J., M. Alber, M. Bangert, R. Bokrantz, T. C. Chan, J. O. Deasy, A. Fredriksson, B. L. Gorissen, M. Van Herk, W. Liu, et al. (2018). "Robust radiotherapy planning". *Physics in Medicine & Biology* **63**:22, 22TR02.

Unkelbach, J., T. Bortfeld, B. C. Martin, and M. Soukup (2009). "Reducing the sensitivity of impt treatment plans to setup errors and range uncertainties via probabilistic treatment planning". *Medical physics* **36**:1, pp. 149–163.

Wahl, N., P. Hennig, H. Wieser, and M. Bangert (2017). "Efficiency of analytical and sampling-based uncertainty propagation in intensity-modulated proton therapy". *Physics in Medicine & Biology* **62**:14, p. 5790.

Winterhalter, C., S. Zepter, S. Shim, G. Meier, A. Bolsi, A. Fredh, J. Hrbacek, D. Oxley, Y. Zhang, D. C. Weber, et al. (2019). "Evaluation of the ray-casting analytical algorithm for pencil beam scanning proton therapy". *Physics in Medicine & Biology* **64**:6, p. 065021.

Yepes, P., A. Adair, D. Grosshans, D. Mirkovic, F. Poenisch, U. Titt, Q. Wang, and R. Mohan (2018). "Comparison of monte carlo and analytical dose computations for intensity modulated proton therapy". *Physics in Medicine & Biology* **63**:4, p. 045003.

A

Patient test cases

This appendix describes the patient test cases used in the experiments. For each case, the objectives f_i used to formulate the full objective in (3.17) are summarized in a table together with the corresponding weights w_i. For each test case, different combinations of scenarios are considered as summarized in Table 6.2 on page 42.

A.1 Lung case I

Figure A.1 shows one transverse slice from the 'Lung I' case. The plan features three beams as indicated by the three dashed lines. The CTV (red) located within the right lung.

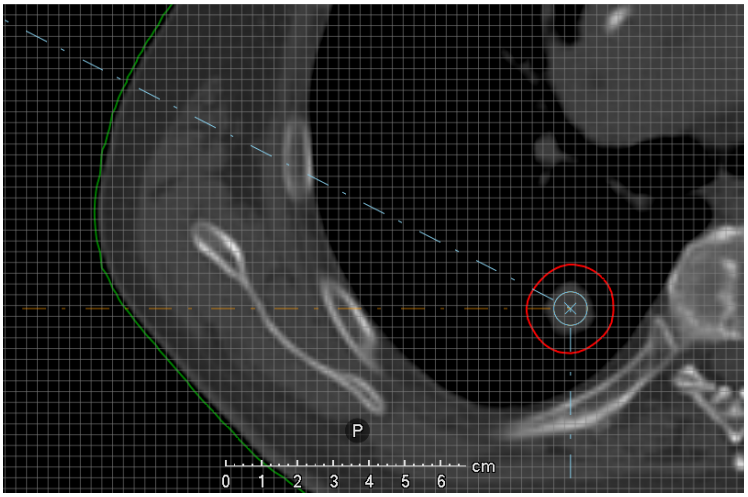


Figure A.1 One transverse slice from the 'Lung I' case illustrating the CTV location within the lung and the three (south, west, north-west) beam central axes. The colors of the ROIs can be matched to the objectives in Table A.1

Objectives defined on the ROIs in Figure A.1 are described in Table A.1. The treatment is planned for 30 fractions (fx) and the spot MUs are constrained by $0.01 \leq x \leq 15$ in units of MU/fx. The optimization runs for 100 iterations with spot filtering after 20 iterations.

Table A.1 The objectives for the 'Lung I' case. Please refer to the equations referenced under 'objective' for mathematical definitions and explanations of the parameters. Robust objectives are indicated by ($\star$) adjacent to the weight.

ROI	Objective f_i	w_i	Parameters
CTV (red)	Minimum dose (3.4)	10 ($\star$)	$d_{\text{min}} = 6100\,\text{cGy}$
CTV (red)	Maximum dose (3.5)	10 ($\star$)	$d_{\text{max}} = 6150\,\text{cGy}$
External (green)	Maximum dose (3.5)	1	$d_{\text{max}} = 6300\,\text{cGy}$
External (green)	Dose falloff (3.6)	1	$d_{\text{high}} = 6000\,\text{cGy}$
			$d_{\text{low}} = 0\,\text{cGy}$
			$\delta_{\text{low}} = 1\,\text{cm}$

A.2 Lung case II

Figure A.2 shows one transverse slice from the 'Lung II' case. The plan features two beams as indicated by the two dashed lines. The CTV (orange) is located within the right lung (turquoise).

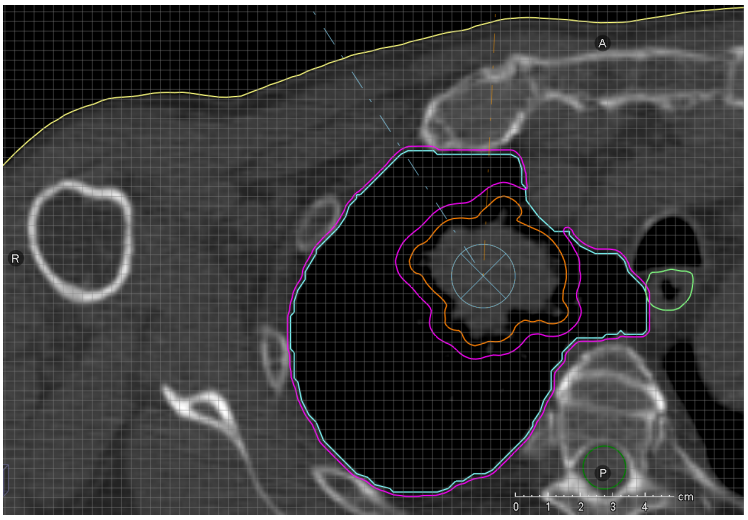


Figure A.2 One transverse slice from the 'Lung II' case illustrating the CTV location (orange) within the right lung (turquoise) and the two (north, north-west) beam central axes. The colors of the ROIs can be matched to the objectives in Table A.2.

Objectives defined on the ROIs in Figure A.2 are described in Table A.2. The treatment is planned for 10 fractions (fx) and the spot MUs are constrained by $0.01 \leq x \leq 15$ in units of MU/fx. The optimization runs for 80 iterations with spot filtering after 20 iterations.

Table A.2 The objectives for the 'Lung II' case. Please refer to the equations referenced under 'objective' for mathematical definitions and explanations of the parameters. Robust objectives are indicated by ($\star$) adjacent to the weight.

ROI	**Objective** f_i	w_i	**Parameters**
CTV (orange)	Uniform dose (3.3)	100 ($\star$)	$d_p = 5000\,\text{cGy}$
CTV (orange)	Max dose (3.5)	10 ($\star$)	$d_{\max} = 5500\,\text{cGy}$
Esophagus (light green)	Max dose (3.5)	1	$d_{\max} = 3000\,\text{cGy}$
Right lung-PTV+1 mm	Max dose (3.5)	1	$d_{\max} = 2000\,\text{cGy}$
Spinal cord (dark green)	Max dose (3.5)	1	$d_{\max} = 1000\,\text{cGy}$
External (yellow)	Dose falloff (3.6)	20	$d_{\text{high}} = 5500\,\text{cGy}$ $d_{\text{low}} = 2000\,\text{cGy}$ $\delta_{\text{low}} = 1\,\text{cm}$
Right lung (turquoise)	Dose falloff(3.6)	10	$d_{\text{high}} = 5500\,\text{cGy}$ $d_{\text{low}} = 2000\,\text{cGy}$ $\delta_{\text{low}} = 1\,\text{cm}$

A.3 Breast case I

Figure A.3 shows one transverse slice from the 'Breast I' case. The plan features two beams as indicated by the two dashed lines. The CTV (purple) is located within the left breast.

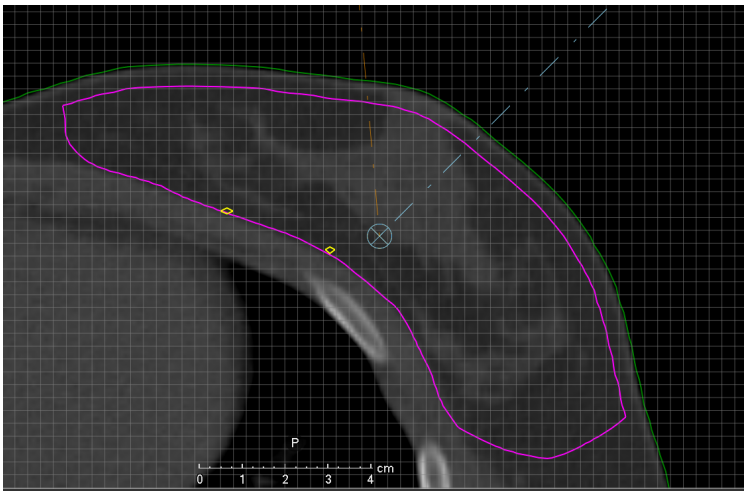


Figure A.3 One transverse slice from the 'Breast I' case illustrating the CTV location (purple) within the left breast and the two (north, north-east) beam central axes. The colors of the ROIs can be matched to the objectives in Table A.3.

Objectives defined on the ROIs in Figure A.3 are described in Table A.3. The treatment is planned for 28 fractions (fx) and the spot MUs are constrained by $0.01 \leq x \leq 15$ in units of MU/fx. The optimization runs for 100 iterations with spot filtering after 20 iterations.

Table A.3 The objectives for the 'Breast I' case. Please refer to the equations referenced under 'objective' for mathematical definitions and explanations of the parameters. Robust objectives are indicated by ($\star$) adjacent to the weight.

ROI	Objective function	Weight	Parameters
CTV (purple)	Uniform dose (3.3)	100 ($\star$)	$d_p = 5100\,\text{cGy}$
CTV (purple)	Max dose (3.5)	400	$d_{\max} = 5200\,\text{cGy}$
External (green)	Max dose (3.5)	1	$d_{\max} = 5200\,\text{cGy}$
Tune_1 (yellow)	Max dose (3.5)	100	$d_{\max} = 5200\,\text{cGy}$
External (green)	Dose falloff (3.6)	1	$d_{\text{high}} = 5040\,\text{cGy}$ $d_{\text{low}} = 2500\,\text{cGy}$ $\delta_{\text{low}} = 1\,\text{cm}$

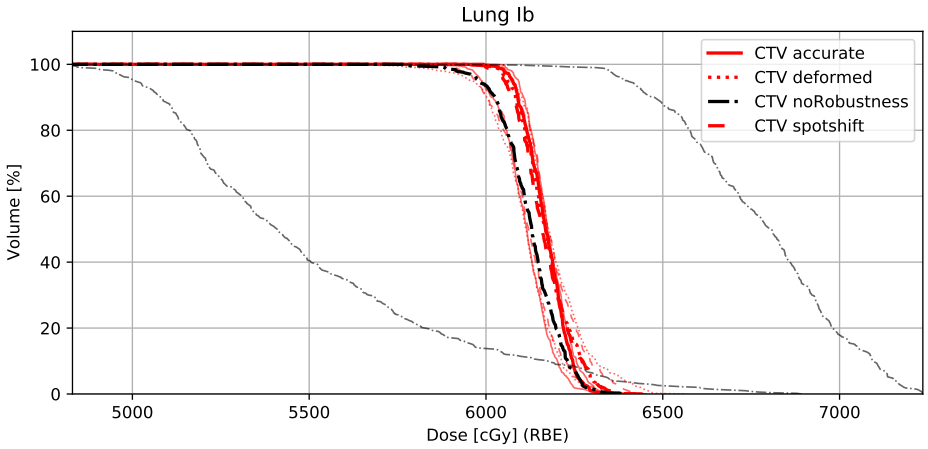
B

Additional results

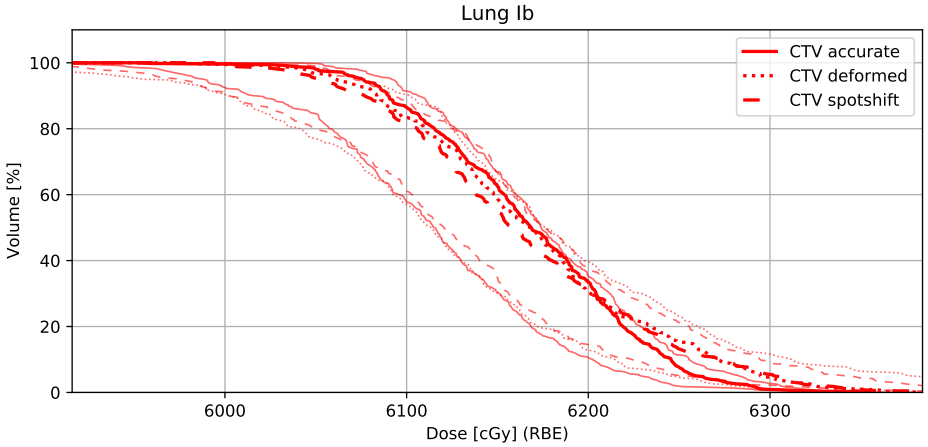
This appendix includes some results that did not fit in Section 7.3.

B.1 Robust optimization

Figure B.1 shows robust evaluation DVH curves for all four 'Lung Ib' plans. Note that the *x*-axis does not begin at zero.



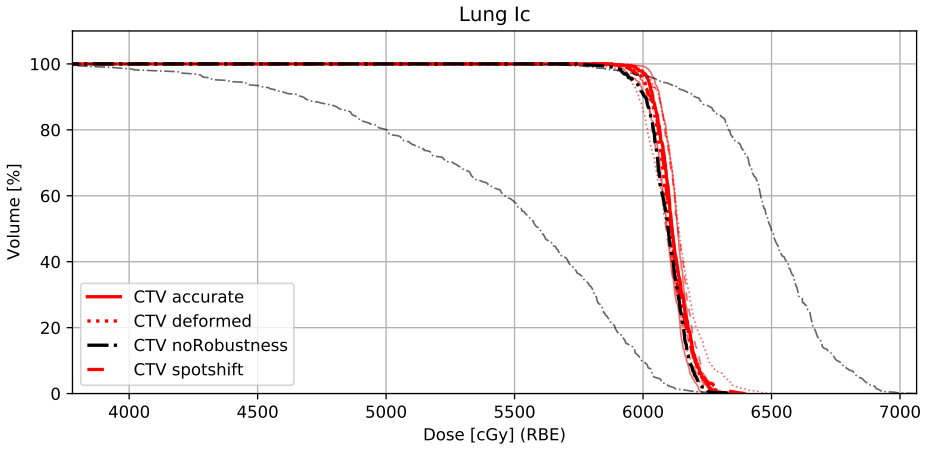
(a) Robust evaluation of all four optimized treatment plans.



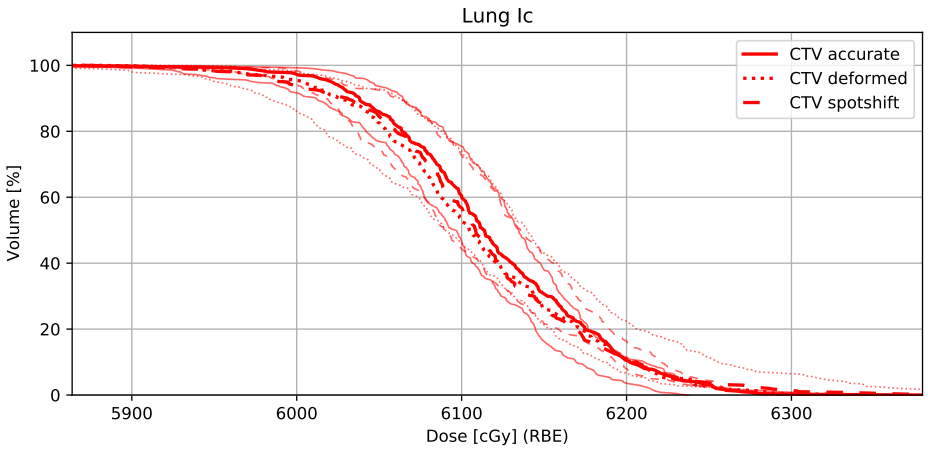
(b) Zoomed-in version of (a) excluding the non-robust baseline.

Figure B.1 Lung Ib: DVH curves for the optimized treatment plans. Three curves are shown for each plan: for the nominal scenario and for the scenarios with least/most dose to the CTV.

Figure B.1 shows robust evaluation DVH curves for all four 'Lung Ic' plans. Note that the x-axis does not begin at zero.



(a) Robust evaluation of all four optimized treatment plans.



(b) Zoomed-in version of (a) excluding the non-robust baseline.

Figure B.2 Lung Ic: DVH curves for the optimized treatment plans. Three curves are shown for each plan: for the nominal scenario and for the scenarios with least/most dose to the CTV.

Master's Theses in Mathematical Sciences 2021:E32
ISSN 1404-6342

LUTFMA-3449-2021

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/