Karlsruhe Institute of Technology

Uncertainty quantification and numerical methods in charged particle radiation therapy

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Pia Stammer

Tag der mündlichen Prüfung: 22. März 2023

1. Referent: Prof. Dr. Martin Frank
2. Referent: Prof. Dr. Mathias Trabs
3. Referent: Prof. Dr. Oliver Jäkel

Acknowledgement

First and foremost I want to thank my advisor Martin Frank for giving me the opportunity to work on this interesting research topic. Martin has created a great work atmosphere and given me the freedom to follow my research interests, while always being available for guidance and advice when needed. He has helped me balance the very interdisciplinary topic of this thesis and greatly improved this work with his feedback.

Further, I would like to thank Mathias Trabs and Oliver Jäkel for taking the time to read and improve my thesis. Your input has provided me with a different view point and has made this a better and more understandable work.

I am also grateful to Niklas Wahl, for introducing me to medical physics and welcoming me into his group at DKFZ. Our regular discussions have helped me stay motivated and focused during my time as a PhD student and his detailed feedback has been invaluable.

I also owe thanks to everyone else who has proof-read parts of my thesis, namely Achim and Pascal Stammer, Steffen Schotthöfer, Jonas Kusch and Robert Schleinhege. You have certainly changed this work for the better.

My colleagues, at KIT, DKFZ and HIDSS4Health, have shaped my time as a PhD student and made it very enjoyable. The many discussions and support I have received have been a great enrichment. In particular, I would like to thank my colleagues at HIDSS4Health for the fun retreats and for commiserating about the struggles of an interdisciplinary PhD. I have always enjoyed my visits to Heidelberg, not least due to the my colleagues at DKFZ. Even though the personal visits were less than I would have hoped initially, I feel part of this group as much as the one in Karlsruhe. I would also like to thank all of my colleagues at CSMM for the fun work atmosphere and joint activities – at work and beyond. Special thanks go to Jonas Kusch, Steffen Schotthöfer and Jannick Wolters for the regular virtual coffee breaks and chats about research among many other topics. I also want to mention my office/corridor mates at campus north and south, who have helped with problems and discussed ideas with me many times.

Last but not least, I am very grateful for the love and support I have received from my family. Thanks especially to Robert Schleinhege for his moral support and for being a constant reminder that there is more to life than work or this thesis.

If a mathematician wishes to disparage the work of one of his colleagues, say, A, the most effective method he finds for doing this is to ask where the results can be applied. The hard pressed man, with his back against the wall, finally unearths the researches of another mathematician B as the locus of the application of his own results. If next B is plagued with a similar question, he will refer to another mathematician C. After a few steps of this kind we find ourselves referred back to the researches of A, and in this way the chain closes.

- ALFRED TARSKI, The Semantic Conception of Truth (1952), p. 41.

Preface

We want to break the metaphorical chain referred to by Alfred Tarski by considering a meaningful application in addition to and as the main motivation for the presented mathematical methods.
Due to the interdisciplinarity of the topic at hand, we will fluently transfer between different fields of mathematics. The stochastic nature of particle dynamics as well as uncertainties, combined with the need to numerically solve a partial differential equation, places radiation therapy at the interface between numerical mathematics and stochastics. The core problem of particle transport can already be viewed as a stochastic or numeric problem depending on the way it is framed. This then in turn influences not only the notation and language used but also the methods which are applied and the way results are analyzed. For example, the dose deposited by particles during irradiation through a stochastic lens is the expected value of the dose deposited by particles in a random walk, where the states are defined by the phase space and transitions by interaction probabilities. In this case, a Monte Carlo method is the obvious choice and a good dose calculation method is characterized by its statistical properties, e.g. the unbiasedness of a dose estimator or its standard error. The same problem through the lens of numerical mathematics is the solution of a partial differential equation using numerical discretizations of the phase space. Here, one would analyze the stability and convergence properties of a method. This work aims to translate and draw connections between the two viewpoints and therefore draws from notations and terminology from both fields.

Motivation

Radiation therapy is applied in approximately 50% of all cancer treatments [14]. To eliminate the tumor without damaging organs in the vicinity, optimized treatment plans are determined. This requires the calculation of three-dimensional dose distributions in a heterogeneous volume with a spatial resolution of 2-3mm [219]. Current planning techniques use multiple beams with optimized directions and energies to achieve the best possible dose distribution. Each dose calculation however requires the discretization of the six-dimensional phase space of the linear Boltzmann transport equation describing complex particle dynamics. Despite the complexity of the problem, dose calculation errors of less than 2% are clinically recommended [168] and computation times cannot exceed a few minutes [219]. Additionally, the treat-

ment reality often differs from the computed plan due to various uncertainties, for example in patient positioning, the acquired CT image or the delineation of tumor and organs at risk [e.g. 217, 156, 132, 133]. Therefore, it is essential to include uncertainties in the planning process to determine a robust treatment plan. This entails a realistic mathematical model of uncertainties, quantification of their effect on the dose distribution using appropriate propagation methods as well as a robust or probabilistic optimization of treatment parameters to account for these effects.
Fast and accurate calculations of the dose distribution including predictions of uncertainties in the computed dose are thus crucial for the determination of robust treatment plans in radiation therapy. Monte Carlo methods are often used to solve transport problems, especially for applications that require high accuracy [194, 219]. In these cases, common non-intrusive uncertainty propagation strategies that involve repeated simulations of the problem at different points in the parameter space quickly become infeasible due to their long run-times [225, 219, 11]. Quicker deterministic dose calculation methods allow for better incorporation of uncertainties, but often use strong simplifications or admit non-physical solutions and therefore cannot provide the required accuracy [225, 11].

This work is concerned with finding efficient mathematical solutions for three aspects of (robust) radiation therapy planning: 1. Efficient particle transport and dose calculations, 2. uncertainty modeling and propagation for radiation therapy, and 3. robust optimization of the treatment set-up.

Outline

The thesis is accordingly structured into three main parts: The first concerned with particle transport and dose calculations (part II), the second discussing uncertainty modeling and propagation (part III) and the third touching on robust treatment plan optimization (part IV). Each part starts by discussing mathematical modeling and design decisions necessary to translate the physical problem into a form that can be tackled using mathematical methods. Subsequently, we discuss established as well as novel numerical methods and simulations to solve the resulting problem. Lastly, the results of realistic numerical experiments for an application in radiation therapy are analyzed and compared to conventional approaches.
Contrary to this structure, modeling and simulation are usually not consecutive steps. They are in fact part of a modeling cycle. In this cycle, the way real-world phenomena are represented mathematically influences the applicable methods available to solve the resulting problem. Vice versa also the available methodology dictates how these models are chosen. For example, a probabilistic representation of uncertainties in terms of a multivariate normal distribution might be convenient when aiming at an application of the efficient uncertainty propagation methods introduced in chapter 13 and chapter 14. On the other hand, the choice of viable propagation methods can be limited when modeling uncertainties in terms of a discontinuous function which is difficult to sample, even if it is a good fit for the observed data. Here we have opted for the chronological order in which the steps would be relevant in a robust planning process.

After an introduction to radiotherapy planning in part I, part II of this thesis is concerned with the problem of dose calculations, which require the solution of a particle transport problem. We initially derive a physical model in chapter 4: The linear Boltzmann transport equation describes the behavior of different

particle types used for irradiation within heterogeneous materials. A prerequisite is the knowledge of particle and material-dependent cross sections, which describe the likelihood of scattering, absorption and energy loss. We discuss the common *continuous slowing down approximation* in section 4.2, which decouples the angular deflection of particles from energy loss. Further, different models of physical interactions are introduced in sections 4.3 to 4.5.
Computational methods for the solution of the resulting equation can be grouped into deterministic and stochastic approaches. We discuss the most common approaches in chapter 5. In chapter 6, we introduce the novel software framework KiT-RT for efficient dose calculations using classic deterministic methods such as the S_N, P_N or M_N method for angular discretizations. In chapter 7, we then propose to tackle the challenges arising from high-dimensional phase spaces in radiation therapy applications by applying the dynamical low-rank approximation (DLRA), which evolves the solution on a low-rank manifold in time. To facilitate the use of boundary conditions and reduce the overall rank, the radiation transport equation is further split into collided and uncollided particles through a collision source method. Uncollided particles are then described by a directed quadrature set guaranteeing low computational costs, whereas collided particles are represented by a low-rank solution. All discussed methods are finally compared with respect to memory consumption, run-time and accuracy in chapter 8.

In part III, we add uncertainties to the transport problem. Again we start by discussing the mathematical modeling of uncertainties in radiation therapy (chapter 11). Typically, uncertainties are assumed to follow a Gaussian distribution and be invariant over time as well as for different beam parameters. More complex models are difficult to incorporate in standard uncertainty propagation and optimization approaches due to their high dimensionality. We introduce novel approximations for time-dependent and data-based uncertainty modeling, which are based on a multivariate Gaussian phase space distribution.
We then discuss common uncertainty propagation strategies in chapter 12 and propose different novel approaches to handle more complex uncertainty models in chapters 13 to 15. Specifically, we propose two efficient uncertainty propagation strategies: One aimed at Monte Carlo dose calculations and one for deterministic solutions of the transport equation. In chapter 14, we develop an importance sampling-based approach to speed up non-intrusive uncertainty quantification in Monte Carlo dose calculations. The main idea is to exploit the stochastic view of particle transport, where each phase space parameter follows a physical probability distribution. Uncertainties in the phase space parameters can then be viewed as a mere change in their probability distributions. For uncertainty propagation with deterministic dose calculation methods, we again shift to a numeric point of view. A tensor DLRA can be used to directly include (discretized) uncertain parameters. In chapter 15 we derive the time evolution equations for tensor DLRA applied to the continuous slowing down equation and show results for a time-dependent uncertainty in the tissue density.

Lastly, part IV is concerned with robust treatment plan optimization. We first discuss the design of the optimization problem in chapter 17. The objective function and constraints can be used to model the deviation of the computed dose from a prescribed treatment plan. Further, we can incorporate measures of robustness, e.g. by considering different error scenarios or penalizing a high variance in the dose. This part of the thesis thus consolidates the previous work on dose calculations and uncertainty propagation. We demonstrate how to efficiently include the approach from chapter 14 in two given robust optimization strategies.

Novelty and credit statement

This work presents a collection of novel research conducted with different coauthors put into the context of previous work. This section serves to clarify which parts of this thesis describe new research and highlight my own as well as my coauthors' contributions to this work.

Chapter 6 introduces a novel code framework for deterministic transport solvers I developed together with Jonas Kusch, Steffen Schotthöfer, Jannick Wolters and Tianbai Xiao [112]. Here, the basic code structure was developed jointly and I worked on the implementation and tuning of the continuous slowing down solvers and test cases aimed at a radiation therapy application. My coauthors on the other hand focused on the implementation of classic and novel transport solvers for the Boltzmann transport equation in non-radiation therapy applications (where the dose does not have to be computed).

The majority of chapter 7 is based on joint work I conducted with Jonas Kusch published in [113, 114]. Jonas Kusch proposed the idea of an application of the dynamical low-rank approximation to radiation therapy using the CSD equations. After jointly deriving the time evolution scheme, I implemented the initial electron CSD DLRA solver using the first collision source method and worked on designing initial conditions to correctly model the beam structure in radiation therapy. I later extended the method to solve the proton transport equation in addition to that for electrons. My coauthor was responsible for the stability proofs, as well as the extension to 3D and higher order numerical schemes. Advice on the physical models and data bases used in proton therapy was given by Niklas Wahl and Danny Lathouwers. The implementation of the recent 3D and proton extensions is openly available on GitHub but has not been published elsewhere.

The numerical test cases for the comparisons of all novel and classic methods in chapter 8 were designed and implemented by myself.

In chapter 11, the simplified model for range uncertainties is based on joint work I conducted with Niklas Wahl, Lucas Burigo, Martin Frank and Oliver Jäkel [203], published in *Physics in Medicine and Biology* and presented at PTCOG [205]. Here, Niklas Wahl pointed out the connection between density, particle range and energy and I developed as well as implemented the presented mathematical model. Further, I developed the multivariate models for autocorrelated time-dependent errors described in section 11.4.1, which are published in [206] together with the same coauthors, who were however not involved in this

part of the work. I presented an application of the multivariate Gaussian framework for time-dependent uncertainties to patient data as a proof of concept in [204] and further supervised a master's thesis conducted by Jean Radig [181] on this topic. Specifically, the implementation in the APM toolbox, validation against Monte Carlo, as well as the hyperparameter tuning for kernels based on patient data, were subjects of the master's thesis.

Chapter 14 is based on [203, 206], which I worked on together with Niklas Wahl, Lucas Burigo, Martin Frank and Oliver Jäkel. I developed the idea of reusing information accumulated during regular simulations for uncertainty quantification in a discussion together with Niklas Wahl. I then derived the mathematical framework based on importance sampling, including the required probability distributions, as well as conducted the validation for different test cases and investigation of mathematical properties such as bias, bounds and convergence. The implementation within the MC framework TOPAS was supported by Lucas Burigo and credit for advice on phase space distributions in radiation therapy goes to Niklas Wahl.

Chapter 15 describes unpublished work together with Jonas Kusch. This is an extension of our work described in chapter 7 to the tensor DLRA which enables us to directly include uncertainties.

Part IV contains novel and unpublished work on the use of chapter 14 for efficient robust optimization for MC dose calculations. Here, I worked independently on the transfer of the methods from chapter 14 to a worst-case and expected value optimization context as well as the derivation of the variance influence computation algorithm and numerical test cases.

Contents

List of Figures

List of Tables

List of Algorithms

List of Symbols

A_p	mass number of projectile
A_t	mass number of target
$BB(\boldsymbol{r};\boldsymbol{Z})$	black box function of Monte Carlo dose calculation algorithm
$D(\mathbf{r})$	dose at position $\mathbf{r}$
E	energy
$\mathbb{E}_p[\cdot]$	expected value w.r.t probability distribution p
e	elementary electric charge
$L(\mathbf{r}, E)$	stopping power
m_p	projectile mass
m_t	target mass
N_A	Avogadro constant
$\mathbb{P}(\omega)$	probability of ω
Q_{sca}	scattering operator
$(\mathbf{r}, \mathbf{\Omega}, E)$	phase space of particle density
R	particle range
$\mathbf{r} = (r_x, r_y, r_z)$	spatial position
$\mathcal{R}$	rank in dynamical low-rank approximation
$S(\mathbf{r}, \mathbf{\Omega}, E)$	source term in transport equation
$S_0(\mathbf{Z})$	physical distribution of initial phase space parameters/particle source
$S_i(\boldsymbol{Z})$	physical distribution of initial phase space parameters given realisation δ_i of uncertain parameters
$\mathcal{S}(\mathbf{Z}(\mathbf{\Delta}))$	physical distribution of initial phase space parameters/particle source under uncertainty
t	(pseudo-)time

$Var(\cdot)$	variance
Z_p	atomic number of projectile
Z_t	atomic number of target
$\mathbf{Z} = (\mathbf{r}_0, \mathbf{\Omega}_0, E_0)$	initial phase space (at time $t = 0$)
$\mathbf{\Delta}$	uncertain parameter
ϵ_0	electric field constant
$\mathbf{\Omega}$	angle/direction of flight
$\psi(\mathbf{r}, \mathbf{\Omega}, E)$	angular flux/particle density
$\psi_c(\mathbf{r}, \mathbf{\Omega}, E)$	angular flux/particle density of collided particles
$\psi_l(\mathbf{r}, \mathbf{\Omega}, E)$	angular flux/particle density of l-times collided particles
$\psi_u(\mathbf{r}, \mathbf{\Omega}, E)$	angular flux/particle density of uncollided particles
$\rho(\mathbf{r})$	tissue/material density at position $\mathbf{r}$
$\Sigma_s(\mathbf{r}, \mathbf{\Omega} \cdot \mathbf{\Omega}', E' \to E)$	differential scattering cross section
$\Sigma_t(\mathbf{r}, E)$	total cross section
θ	deflection angle in center of mass frame
ϑ	deflection angle in laboratory frame

Part I

Introduction to radiotherapy planning

In the following chapters, we give a brief overview of the process of radiotherapy planning and introduce some common domain-specific quality criteria. We will touch on the steps and mathematical problems involved in treatment planning, which will be the topic of or motivate the following chapters. Note, that this is not a comprehensive overview and further details will be discussed when deriving models or methods in the subsequent chapters. This should mainly serve as context and background knowledge for readers not familiar with radiation therapy.

CHAPTER 1

From initial diagnosis to irradiation

Once a patient has been diagnosed with cancer, the data required for treatment planning is acquired: a CT scan is taken, from which the position and extent of volumes of interest (VOI) such as the tumor and surrounding organs at risk (OAR) are determined through manual or automatic image delineation [97]. Further tissue/material parameters required for the dose calculations, such as the tissue density $\rho(\mathbf{r})$, electron density, stopping power, or scattering cross sections, are estimated at each point $\mathbf{r} \in \mathcal{X} \subset \mathbb{R}^3$ of the bounded region $\mathcal{X}$ containing the relevant patient anatomy. Typically, this is done based on the grayscale values of the CT scan [for details see e.g. 188, 189]. These conversions constitute one of the main sources of uncertainty since in addition to calibration and measurement errors, CT scans cannot capture information on all physical effects governing dose deposition [188].

Based on this data, a doctor prescribes a treatment plan and corresponding target dose distribution $D^*(\mathbf{r})$. The plan details the amount of dose the tumor should receive as well as other requirements such as the maximum dose limits for certain organs. In treatment planning, all specifics of the treatment set-up are then defined in order to best fulfill this prescription. This includes the choice of treatment modality (i.e. particle type), beam geometry, number of sittings (so-called *fractions*) and also the computation of the corresponding dose distributions. Figure 1.1 shows an overview of the complete process. In the following, we will focus mostly on treatment planning and dose calculation. The other steps will however be of interest as the sources for uncertainties in part III.

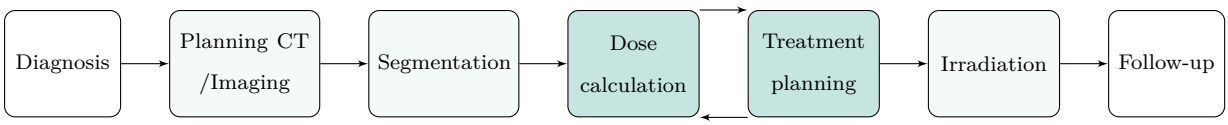


Figure 1.1: Overview of radiation therapy process from diagnosis to irradiation.

CHAPTER 2

Intensity-modulated radiation therapy

This thesis is written within the context of a variant of radiation therapy called *intensity-modulated radiation therapy (IMRT)*. IMRT uses numerous small charged particle or photon beams, called *beamlets* or *pencil beams*, with different shapes and intensities which are superimposed to achieve better coverage of the tumor and organ sparing [e.g. 229, 53, 159]. Here, *beamlet* is the smallest instance with a designated source position, energy and angle of irradiation. An ensemble of beamlets coming from the same geometrical set-up of nozzle and patient is then referred to as *beam*, a group of beamlets with the same lateral positioning within a beam as *ray* and a group of beamlets with the same energy within a beam as *energy level*. The beamlets are irradiated according to a temporal pattern dictated by machine requirements. For example, all beamlets of the same energy level are applied in sequence to minimize the waiting times during ramp-ups to higher energy levels. Note, that the term *intensity-modulated particle therapy (IMPT)* is also frequently used to denote this concept in particle-based radiation therapy [159].

Figure 2.1 shows a schematic illustration of intensity-modulated radiation therapy and fig. 2.2 illustrates how superimposing beamlets can be used to tailor the dose distribution to different target shapes.

2.1 Treatment plan optimization

A treatment plan in intensity-modulated particle therapy is given by a set of beamlets $b = 1, ..., B$, each with a defined intensity w_b, energy E_b, source position $\mathbf{r}_b$ and angle of irradiation $\mathbf{\Omega}_b$. The combined dose at position $\mathbf{r}$ can be computed from the doses $D_b(\mathbf{r}; \mathbf{r}_b, \mathbf{\Omega}_b, E_b)$ delivered by each individual beamlet as [e.g. 236]

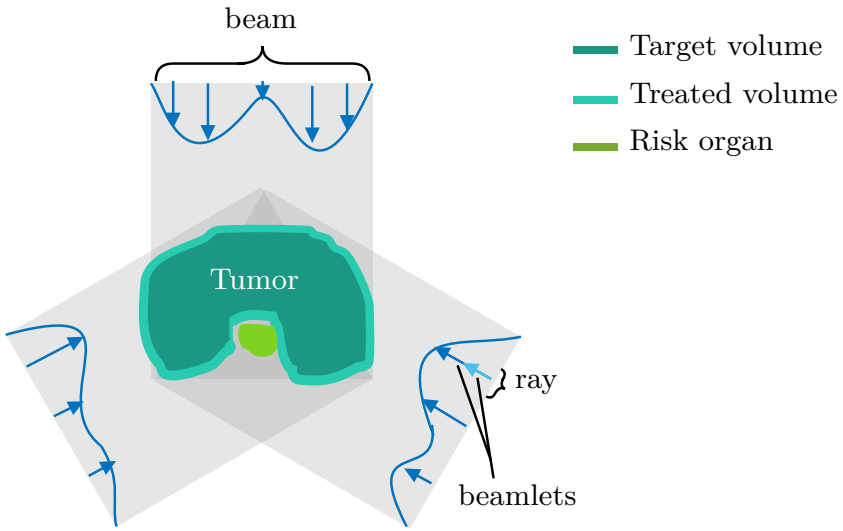


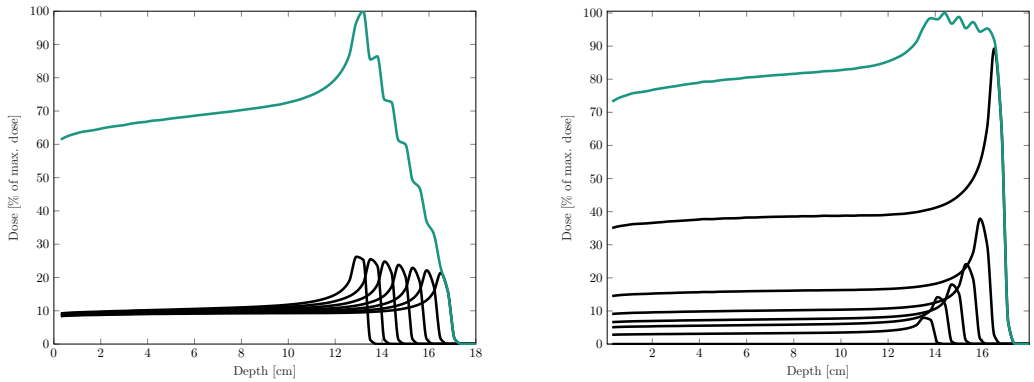
Figure 2.1: Illustration of IMRT set-up with three irradiation angles (beams) à several beamlets with different energies which are layered to spare the organ at risk partly encircled by the tumor.

$$D(\mathbf{w}, \mathbf{r}) = \sum_{b=1}^{B} w_b \cdot D_b(\mathbf{r}; \mathbf{r}_b, \mathbf{\Omega}_b, E_b). \tag{2.1}$$

For simplicity, we will omit the dependency of $D_b(\mathbf{r}; \mathbf{r}_b, \mathbf{\Omega}_b, E_b)$ on the beamlet position, angle and energy and write $D_b(\mathbf{r})$ in the following. The goal when optimizing a plan is to determine intensities w_b for a given set of viable beamlets such that the best possible conformity to a prescribed dose is achieved. Usually, this is modeled using an objective function and constraints which penalize deviation from the prescribed dose in the tumor volume and prohibit overdosing in organs at risk within the vicinity. A variety of approaches exists to implement these concepts in the design of an optimization problem. More details on this can be found in [e.g. 236, 53, 11] and two options for robust optimization will be discussed in part IV. For an examplary choice of objective and constraints [see 11], the optimization problem could take the following form:

$$\begin{aligned} w^* &= \arg\min_{w \geq 0} F(D(\mathbf{w}, \mathbf{r})) \qquad (2.2)\\ \text{where e.g. } F(D(\mathbf{w}, \mathbf{r})) &= \sum_{t=1}^{N_t} F_t D(\mathbf{w}, \mathbf{r}) + \sum_{ro=1}^{N_r} F_{ro} D(\mathbf{w}, \mathbf{r}) \\ F_t(D(\mathbf{w}, \mathbf{r})) &= \sum_{i \in \text{Tumor}} s_l^t \mathbb{1}_{D(\mathbf{w}, \mathbf{r}_i) \leq D_t^{min}} [D(\mathbf{w}, \mathbf{r}_i) - D_t^{min}]^2 + s_u^t \mathbb{1}_{D(\mathbf{w}, \mathbf{r}_i) \geq D_t^{max}} [D(\mathbf{w}, \mathbf{r}_i) - D_t^{max}]^2 \\ F_{ro}(D(\mathbf{w}, \mathbf{r})) &= \sum_{i \in \text{Risk organ}} s_u^{ro} \mathbb{1}_{D(\mathbf{w}, \mathbf{r}_i) \geq D_{ro}^{max}} [D(\mathbf{w}, \mathbf{r}_i) - D_{ro}^{max}]^2 \end{aligned}$$

Here N_t, N_r are the number of cells in the discretizations of the tumor and risk organs respectively, D_t^{min} is the minimum dose requirement in the tumor and $D_{ro/t}^{max}$ is the maximum dose limit in tumor



(a) (b)

Figure 2.2: Layered proton beams from one ray with different energies (135, 138, 142, 145, 148, 151 and 155 MeV) and intensity modulations/beamlet weights. Left: Equal weights lead to a wedge-shaped dose distribution. Right: Optimized weights lead to a plateau shape.

and risk organs. The penalties s_l^t, s_u^t, s_u^{ro} can be used to steer the importance of the different constraint violations.

Depending on the design of the optimization problem, many different optimization algorithms can be used and have been used in the context of radiation therapy [53]. In the presence of non-linear constraints, penalty or Lagrangian methods are typically used to obtain an unconstrained problem [e.g. 157]. Variations of Newton's method have been used for quadratic objective functions [236] and simulated annealing or genetic algorithms are common in IMRT [53]. In part IV, we will employ an interior point method as implemented in the open-source treatment planning software matRad [224, 234]. Here, an unconstrained problem is obtained by shifting inequality constraints to the objective function using logarithmic barrier functions. The unconstrained problem can then be solved using a Quasi-Newton algorithm and the L-BFGS approximation. More information on the optimization methods can be found in [53, 157].

Regardless of the exact optimization algorithm, the optimization of a dose-based objective function involves numerous iterations and updates of the beamlet weights/intensities [236]. In each of the iterations, the dose for updated intensities has to be computed, i.e., the linear Boltzmann transport equation (LBE) has to be solved. This process is illustrated in the left column of fig. 2.3. While the optimization can be sped up by precomputing the beamlet doses D_b, this also requires additional storage and pre-processing time.

Figure 2.3 also shows the additional steps necessary to include uncertainties in this process. From additional data on measurable uncertainties, over uncertainty models, propagation through the dose calculation to an incorporation into the treatment plan optimization, uncertainty quantification touches each step in the treatment planning process. In part III and part IV, each of these steps will be addressed, however, in the following part of this thesis we focus on the problem which is at the center and the computational bottleneck of the whole process: dose computation, i.e. the solution of the linear Boltzmann transport equation.

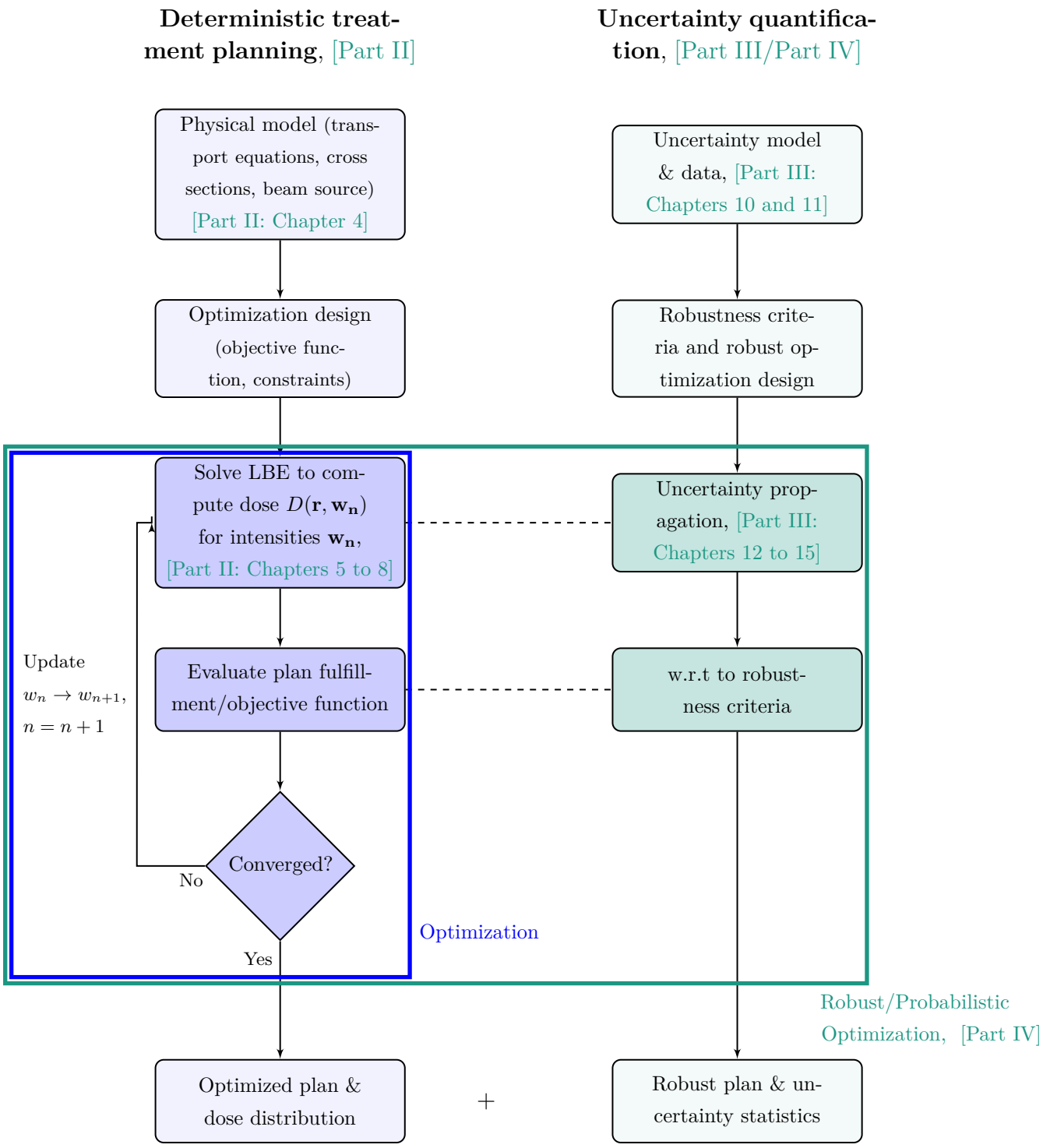


Figure 2.3: Overview of the treatment planning process with and without uncertainties, as well as which step is discussed where within this thesis.

CHAPTER 3

Quality criteria

In the following, we briefly introduce the quality criteria that will be used to evaluate the outcome of dose calculations as well as treatment plan optimizations.

3.1 Dose-volume histograms

Dose-volume histograms (DVH) summarize the amount of delivered dose associated with a treatment plan for each volume of interest (VOI) of a patient [e.g. 49]. For each VOI, the dose values are plotted against the fraction of the volume receiving at least this dose (see fig. 3.1 for an example). DVHs are frequently used to compare different treatment plans with respect to the amount of dose delivered to organs at risk as well as the uniformity of tumor coverage. Here, we will make use of this tool in part IV to compare conventional and robust treatment plans determined using different dose calculation and uncertainty propagation methods. However, the majority of positional information about dose distributions is lost by aggregating the dose for each VOI. Therefore, DVHs are typically not used as a sole evaluation criterion.

3.2 Integrated depth-dose curves

The so-called integrated depth dose (IDD) is used to evaluate the dose deposition along the depth dimension. Note, that here the coordinate system is defined in a way such that the z-axis is pointing in the beam direction. Depth-dose values are obtained by integrating the three-dimensional dose distribution in space along the lateral dimensions, i.e., dimensions orthogonal to the beam direction.

$$IDD(r_z) = \int_{\mathbb{R}} \int_{\mathbb{R}} D(r_x, r_y, r_z) \mathrm{d}r_x \mathrm{d}r_y. \tag{3.1}$$

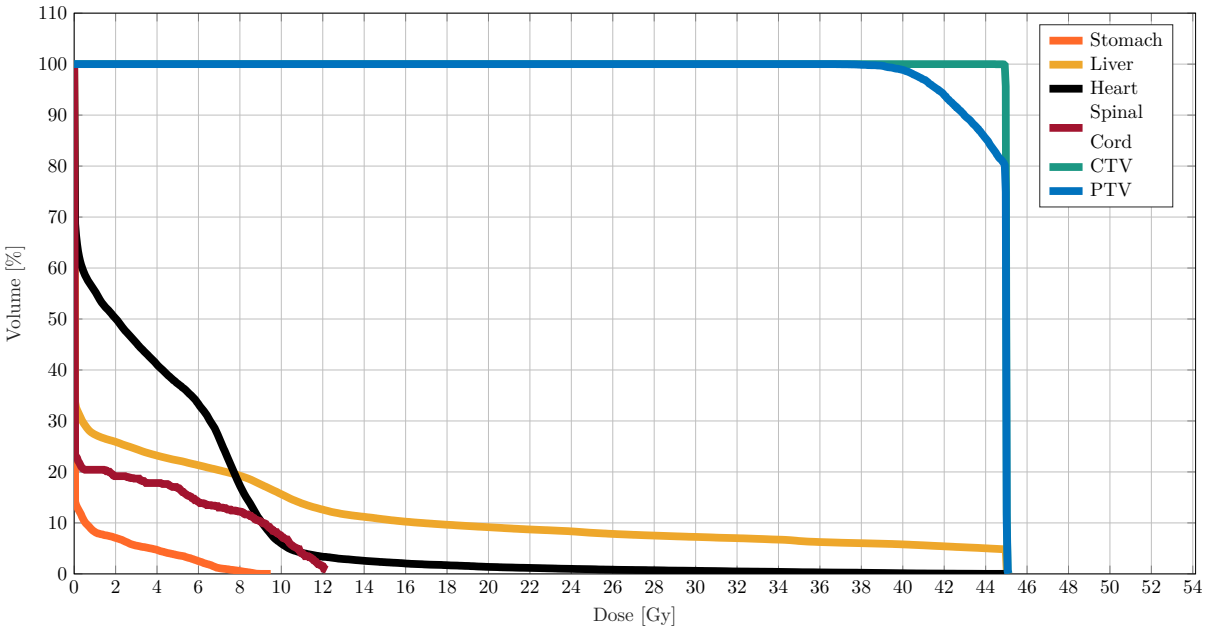


Figure 3.1: Example of a dose volume histogram for an optimized treatment plan of a liver patient. The clinical target volume (CTV) and planning target volume (PTV) describe different margins around the tumor. Here, nearly 100% of the volume receives at least 45 Gy. The remaining structures are organs at risk, where the majority of the volumes receive less than 2 Gy.

While the IDD is well-equipped to validate certain particle-specific characteristics such as the Bragg peak for protons or the particle range, it also averages out noise or oscillations and disposes of all information about lateral distributions.

3.3 Gamma index

While the previous criteria have strongly compressed the available information, the gamma index analysis aims at comparing two complete 3D dose matrices. To compare doses D and D_c, one first computes

$$\gamma(\mathbf{r}) = \min_{\mathbf{r}_c} \sqrt{\frac{|\mathbf{r}_c - \mathbf{r}|^2}{\Delta\tilde{\mathbf{r}}^2} + \frac{|D_c(\mathbf{r}_c) - D(\mathbf{r})|^2}{\Delta\tilde{D}^2}}, \tag{3.2}$$

where $\mathbf{r}$ and $D(\mathbf{r})$ are a location and the dose at that location respectively, $\mathbf{r}_c$ and $D_c(\mathbf{r}_c)$ are locations and dose in the cube to be compared and $\Delta\tilde{\mathbf{r}}$ and $\Delta\tilde{D}$ are the user-defined distance and dose tolerances [135]. The overall agreement is then measured according to the γ-passrate over all points $\mathbf{r}$ [135]. For each point

$$\begin{cases} \gamma(\mathbf{r}) \leq 1, \ \mathbf{r} \text{ passes} \\ \gamma(\mathbf{r}) > 1, \ \mathbf{r} \text{ fails} \end{cases}. \tag{3.3}$$

The criterion is clinically used and implements the idea of agreement within a predefined neighborhood

and tolerance around a single coordinate or data point. In contrast to criteria such as the difference or (root) mean square error, which compare the values at an exact location, this does not take minor fluctuations in the spatial dose distribution into account. Since Monte Carlo simulations are often used as a reference when validating novel methods, the reference itself can also be subject to statistical errors. Thus, the dose distribution is not necessarily incorrect if it deviates slightly from the reference. However, the γ-criterion is not a metric in the mathematical sense if the tolerances are larger than zero. Two dose cubes could have 100% agreement according to the γ-passrate without actually being the same. Therefore the γ-passrates always have to be interpreted together with the applied tolerances. In the following, we choose values frequently used in literature: a distance tolerance of $2\,\mathrm{mm} - 3\,\mathrm{mm}$ and a dose tolerance of $2\,\% - 3\,\%$. Furthermore, the γ-passrate can depend on the grid resolution as well as the implementation [89]. We use the matRad implementation [234], which reduces the dependency on the resolution by interpolating the comparison dose cube and comparing it at a higher resolution.

3.4 Difference maps

A far simpler criterion, which we will use in chapter 14 is the absolute difference between dose values in each voxel of the discretized matrix. For the difference maps, we compute

$$\mathrm{diff}_i(D_i^{\mathrm{ref}}, D_i^{\mathrm{est}}) = D_i^{\mathrm{ref}} - D_i^{\mathrm{est}} \,, \tag{3.4}$$

for each voxel i in the reference result D^{ref} and estimate D^{est}.

Part II

Particle transport and dose calculations

Dose calculations in radiation therapy require the solution of a particle transport problem. Particle dynamics and interactions are described by the linear Boltzmann transport equation. In the following, we discuss physical models to compute scattering cross sections and the continuous slowing down approximation for charged particles. Further, common deterministic and stochastic solution methods are introduced and compared. We present the software framework (KiT-RT) which consolidates efficient deterministic solvers for radiation transport. Lastly, the dynamical low-rank approximation is introduced as a novel method to reduce computational costs and memory associated with fine numerical discretizations in deterministic dose calculations.

CHAPTER 4

Physical modeling

4.1 The Boltzmann transport equation

We are interested in charged particle transport for an application in radiation therapy. The behavior of particles can be described by the Boltzmann transport equation. Here we focus on protons and electrons because of their dose deposition properties, which can be exploited to explicitly target tumors at certain depths. This however also makes protons especially sensitive to uncertainties due to the sharp peak in their depth-dose curve [132, 130]. Electrons are further of interest as the most common type of particle considered in mathematical research on transport equations.

In radiation therapy, we are mainly interested in the spatial distribution of deposited dose in the patient after the irradiation. Since the speed of particles is high in relation to the irradiation time and size of the irradiated region, all processes can be viewed as time-independent [e.g. 94]. For the sake of brevity, we thus consider the steady state transport equation in the following.

Let $\mathcal{X}$ be a bounded region containing the patient anatomy. The linear Boltzmann equation reads

$$\Omega \cdot \nabla\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E) + \Sigma_t(\boldsymbol{r}, E)\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E) = Q^{Sca}(\boldsymbol{r}, \boldsymbol{\Omega}, E) + S(\boldsymbol{r}, \boldsymbol{\Omega}, E) \,, \tag{4.1}$$

$$\psi(\mathbf{r}, \boldsymbol{\Omega}, E) = \psi_{\mathrm{BC}}(\mathbf{r}, \boldsymbol{\Omega}, E) \quad \text{for } \mathbf{r} \in \partial\mathcal{X} \,, \tag{4.2}$$

where Q^{Sca} is given as

$$Q^{Sca}(\boldsymbol{r}, \boldsymbol{\Omega}, E) = \int_0^\infty \int_{\boldsymbol{\Omega'} \in S^2} \Sigma_s(\boldsymbol{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega'}, E' \to E)\psi(\boldsymbol{r}, \boldsymbol{\Omega'}, E')\mathrm{d}\boldsymbol{\Omega'}\mathrm{d}E' \,. \tag{4.3}$$

The phase space of the particle density ψ consists of energy $E \in [0, E_{max}] \subset \mathbb{R}_+$, space $\mathbf{r} \in \mathcal{X} \subset \mathbb{R}^3$ and direction of flight $\mathbf{\Omega} \in \mathbb{S}^2$. Σ_t and Σ_s are material cross sections, which describe scattering and absorption interactions of particles with tissue. In particular, $\Sigma_s(\boldsymbol{r}, \mathbf{\Omega} \cdot \mathbf{\Omega'}, E' \to E)$ is the *differential cross section*, which describes the probability of scattering from direction $\mathbf{\Omega'}$ and energy E' to direction $\mathbf{\Omega}$ and energy E after a collision. Note that $\mathbf{\Omega} \cdot \mathbf{\Omega'} = \mathbf{\Omega'} \cdot \mathbf{\Omega} = \cos(\theta) := \mu$, where θ is the deflection angle, i.e., change in direction caused by the collision. $\Sigma_t(\boldsymbol{r}, E)$ on the other hand describes the *total cross section*, which can be determined by

$$\Sigma_t(\boldsymbol{r}, E) = 2\pi \int_0^\infty \int_{-1}^1 \Sigma_s(\boldsymbol{r}, \mu, E' \to E) \mathrm{d}\mu \mathrm{d}E' + \Sigma_a(\boldsymbol{r}, E), \tag{4.4}$$

where $\Sigma_a(\boldsymbol{r}, E)$ is the absorption coefficient.

The terms in eq. (4.1) can be interpreted as follows: The quantity we solve for is the angular flux or particle density $\psi(\boldsymbol{r}, \mathbf{\Omega}, E)$, which represents the density of particles moving in direction $\mathbf{\Omega}$ with kinetic energy E and at spatial position $\mathbf{r}$. The left-hand side (LHS) then describes advection plus absorption and out-scattering of particles. The first LHS term is the advection term, which describes free flight of particles through the domain. The second LHS term consolidates the total removal of particles by out-scattering to a different energy or direction and absorption. The right-hand side (RHS) then describes particles entering the state $(\boldsymbol{r}, \mathbf{\Omega}, E)$ through a source $S(\boldsymbol{r}, \mathbf{\Omega}, E)$ or in-scattering described by the Boltzmann scatter operator Q^{Sca}. Here, all particles scattered from different energies E' and different directions $\mathbf{\Omega'}$ to the currently considered energy and direction E and $\mathbf{\Omega}$ are considered. Further, $\psi_{\mathrm{BC}}(\mathbf{r}, \mathbf{\Omega}, E)$ defines the boundary condition which will later be used to model particles entering the domain, e.g., through a Gaussian beam source.

Next, we will consider the common *continuous slowing down (CSD) approximation* to the linear Boltzmann transport equation. This approximation is convenient since it decouples angular deflection and energy loss, which simplifies the dose computation and scattering terms. Further, we will show that the energy can be treated as a pseudo-time, such that methods for time-dependent partial differential equations can be applied.

4.2 Continuous slowing down approximation

When considering charged particles, i.e., electrons or protons, the interactions are dominated by Coulomb's law. The energy loss of particles is stochastic and described accurately by Bethe-Bloch's formula [18, 24, 152]. Macroscopically, energy loss may however be considered to be continuous as well as behave proportionally to the stopping power [152]. The stopping power then represents the expected energy loss. This yields the so-called continuous slowing down (CSD) approximation [120] to the linear Boltzmann equation:

$$-\partial_E \left(L(\mathbf{r}, E)\psi(\boldsymbol{r}, \mathbf{\Omega}, E)\right) + \mathbf{\Omega} \cdot \nabla_r \psi(\boldsymbol{r}, \mathbf{\Omega}, E) + \Sigma_t(\mathbf{r}, E)\psi(\boldsymbol{r}, \mathbf{\Omega}, E) \tag{4.5a}$$

$$= \int_{\mathbb{S}^2} \Sigma_s(\mathbf{r}, \mathbf{\Omega} \cdot \mathbf{\Omega'}, E)\psi(\mathbf{r}, \mathbf{\Omega'}, E)\, \mathrm{d}\mathbf{\Omega'} + S(\boldsymbol{r}, \mathbf{\Omega}, E)\,. \tag{4.5b}$$

The quantity of interest in radiation therapy is the dose absorbed by the tissue, which can be now

determined from

$$D(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})} \int_0^\infty \int_{\mathbb{S}^2} L(\mathbf{r}, E)\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E)\, \mathrm{d}\boldsymbol{\Omega}\mathrm{d}E \,. \tag{4.6}$$

Here, we use $L : \mathbb{R}_+ \times \mathbb{R}^3 \to \mathbb{R}_+$ to denote the stopping power, which describes the rate at which particles lose energy. The tissue density of the patient is $\rho : \mathcal{X} \to \mathbb{R}_+$. The stopping power L as well as material cross-sections Σ_t and Σ_s are given from physical databases or approximation formulas and discussed further in sections 4.3 to 4.5.

Since particles are assumed to lose energy continuously, the energy can be interpreted as a pseudo-time t. To simplify the evaluation of material properties, we further follow the common assumption that all materials are water-equivalent and differ only in density [e.g. 235, 161, 107], i.e.,

$$\begin{aligned} L(\mathbf{r}, E) &= L^{H_2O}(E)\rho(\mathbf{r}), \\ \Sigma_t(\mathbf{r}, E) &= \Sigma_t^{H_2O}(E)\rho(\mathbf{r}), \\ \Sigma_s(\mathbf{r}, \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}', E) &= \Sigma_s^{H_2O}(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}', E)\rho(\mathbf{r}) \,, \end{aligned} \tag{4.7}$$

where we leave out the superscript H_2O in the following. For a given maximal energy $E_{\max}$ let us define the transformed energy as

$$\widetilde{E}(E) := \int_E^{E_{\max}} \frac{1}{L(E')}\, \mathrm{d}E' \tag{4.8}$$

and the transformed particle density as

$$\widetilde{\psi}(\mathbf{r}, \boldsymbol{\Omega}, \widetilde{E}) := L(E)\rho(\mathbf{r})\psi(\mathbf{r}, \boldsymbol{\Omega}, E(\widetilde{E})) \,. \tag{4.9}$$

Then, multiplying eq. (4.5) with $L(E)$ and plugging in the defined transformation gives

$$\partial_{\widetilde{E}}\widetilde{\psi}(\mathbf{r}, \boldsymbol{\Omega}, \widetilde{E}) + \boldsymbol{\Omega}\cdot\nabla_r \frac{\widetilde{\psi}(\mathbf{r}, \boldsymbol{\Omega}, \widetilde{E})}{\rho(\mathbf{r})} + \widetilde{\Sigma}_t(\widetilde{E})\widetilde{\psi}(\mathbf{r}, \boldsymbol{\Omega}, \widetilde{E}) = \int_{\mathbb{S}^2} \widetilde{\Sigma}_s(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}', \widetilde{E})\widetilde{\psi}(\mathbf{r}, \boldsymbol{\Omega}', \widetilde{E})\, \mathrm{d}\boldsymbol{\Omega}' + S(\boldsymbol{r}, \boldsymbol{\Omega}, \widetilde{E}) \,, \tag{4.10}$$

where we define $\widetilde{\Sigma}_t(\widetilde{E}) := \Sigma_t(E(\widetilde{E}))$ and $\widetilde{\Sigma}_s(\widetilde{E}, \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}') := \Sigma_s(E(\widetilde{E}), \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}')$. Dropping the tilde notation and treating $\widetilde{E}$ as a pseudo-time t gives a slightly modified version of the classical linear Boltzmann equation

$$\begin{aligned} \partial_t\psi(t, \mathbf{r}, \boldsymbol{\Omega}) + \boldsymbol{\Omega}\cdot\nabla_r \frac{\psi(t, \mathbf{r}, \boldsymbol{\Omega})}{\rho(\mathbf{r})} + \Sigma_t(t)\psi(t, \mathbf{r}, \boldsymbol{\Omega}) &= \int_{\mathbb{S}^2} \Sigma_s(t, \boldsymbol{\Omega}\cdot\boldsymbol{\Omega}')\psi(t, \mathbf{r}, \boldsymbol{\Omega}')\, \mathrm{d}\boldsymbol{\Omega}' + S(t, \boldsymbol{r}, \boldsymbol{\Omega}) \\ \psi(t = 0, \mathbf{r}, \boldsymbol{\Omega}) &= L(E_{\max})\rho(\mathbf{r})\psi(\mathbf{r}, \boldsymbol{\Omega}, E_{\max}) \,, \end{aligned} \tag{4.11}$$

which can be treated numerically with classical closure methods and space-time discretizations.

Both the linear Boltzmann transport equation and its continuous slowing down approximation depend on material as well as particle-specific parameters. When considering different treatment modalities, i.e., types of particles used for irradiation, this affects the interactions of particles with the background medium. Besides the accurate computation of stopping power L and scattering/absorption cross sections Σ_s, Σ_t, the production of secondary particles in some cases adds to the complexity of particle dynamics. In the following sections, we will give an overview of the relevant physical interactions for electron and proton transport. For the sake of completeness, a brief insight into photon therapy will also be included.

This will however not be considered further within this thesis. We further give details on how to accurately model the charged particle interactions in the context of a system of transport equations which can be tackled with numerical solvers as described later in chapter 5.

Remark 4.1. *The transport equations as stated here are formulated in terms of the total particle energy, made up of the kinetic energy as well as the particle's rest energy (caused by its mass):* $E_{tot} = E_{kin} + E_{rest}$. *This is often neglected in the case of electrons, which have a comparably low rest energy of* $E^{e^-}_{rest} \approx 0.511$ *MeV. For the much heavier protons, the rest energy of* $E^{p}_{rest} \approx 938.272$ *MeV should however absolutely not be neglected.*

Note that physical databases often state cross section or stopping power in terms of the kinetic particle energy. Thus, it may be necessary to also transfer the tabulated values to the equivalent total energy using the simple formula above.

4.3 Electron transport

While electron therapy is not as widespread as photon or proton therapy, electrons within a range of 6 to 22 MeV are frequently applied to specific types of cancers, such as breast cancer, located near the surface of the skin [80]. This is due to their dose deposition properties which are illustrated in fig. 4.1. Most of the dose is deposited shortly after the beam entry, rendering them unsuitable for deeper seated tumors.

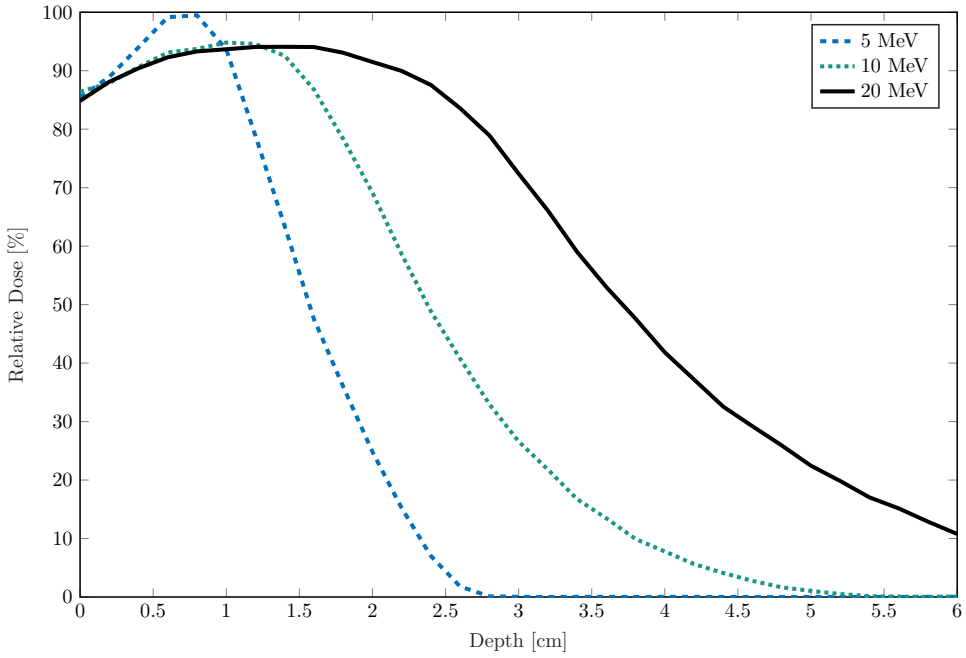


Figure 4.1: Depth dose curve, integrated along lateral dimensions, of an electron beam with 5(dashed), 10 (dotted) and 20 (solid) MeV in water. Data generated using TOPAS MC [175].

As charged particles, electrons directly deposit the majority of the ionizing radiation, i.e. dose, into the background medium. For this reason, it suffices to consider the transport equation for these primary

particles and neglect the production of secondary particles during physical interactions. Further, we assume that projectile electrons do not interact with each other, only with the background medium. Under these assumptions, we can distinguish the following types of physical interactions.

4.3.1 Elastic scattering

Elastic scattering describes interactions that merely change the direction of flight of the electron, without causing the system to lose energy. This mainly includes interactions of electrons with the atomic shell of molecules in the background material, which cause small deflections but virtually no energy loss. Electrons of higher energy can also penetrate the outer shell of an atom and scatter elastically from the nucleus. Cross sections for both of these elastic scattering events in common materials (such as water) can be found in different databases, e.g. the ICRU 77 [55] or NIST database [91]. Note, that elastic scattering is represented by $\Sigma_s(\mathbf{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}', E)$ on the right-hand side of eq. (4.5) and also included in the total cross section Σ_t, which can be obtained by integrating the differential cross sections over all angles (and energies if they are differential in energy). While a transfer of energy between particles is also in principle possible, the amount of transferred energy is negligible for the types of particles considered in this thesis [160, 152].

4.3.2 Inelastic scattering

Inelastic scattering events are responsible for the majority of a projectile particle's energy loss [160]. They are categorized into *soft* and *hard collisions* according to the amount of energy loss and deflection.

Soft e^--e^- collisions

In so-called soft collisions, electrons interact with other electrons in the outer shell, causing excitation or ionization of the target particle. Due to the low binding energies, there is however very little energy transfer and angular deflection.

Hard e^--e^- collision

During hard collisions, electrons not only transfer large amounts of energy to the target electron, but this electron is also ejected with a larger scattering angle.

In Monte Carlo algorithms, both types of interactions are frequently modeled jointly using *condensed history* methods [118, 119, 98], in which the cumulative effects of several soft collisions are approximated by a single hard collision after a certain interaction-free advection step.

In terms of eq. (4.5), energy loss and angular deflections are considered separately in the collision stopping power $L(\mathbf{r}, E)$ and inelastic scattering cross sections, respectively. Values for the collision stopping power in different materials can be found in [17] and are further discussed in section 4.3.4. Deflection due to inelastic scattering does not play a significant role in the clinically relevant energy ranges [160] and is therefore often neglected. For this reason, it will not be further discussed here, the interested reader may however find detailed information on the derivation and formulas governing these cross sections in [e.g. 185, 160, 187].

4.3.3 Bremsstrahlung

When electrons enter the electric field of a nucleus, they are accelerated and produce radiation called *bremsstrahlung*. The energy lost during this process is taken into account through the radiative component of the total electron stopping power. These quantities can be retrieved using the EStar interface of the NIST database [17].

4.3.4 Stopping power

Stopping power describes the material dependent slowing down of particles due to energy loss. For electrons, this energy loss is caused mainly by inelastic collisions (section 4.3.2) as well as bremsstrahlung (section 4.3.3). The contributions of these effects to the total stopping power are illustrated in fig. 4.2. Here, the total stopping power is determined as the sum of the collision and radiative stopping power and can be retrieved from [17]. Clearly, the collision stopping power is the dominating force in the lower to medium energy range typically used for irradiation, while the radiative stopping power only has a major impact for large energies.

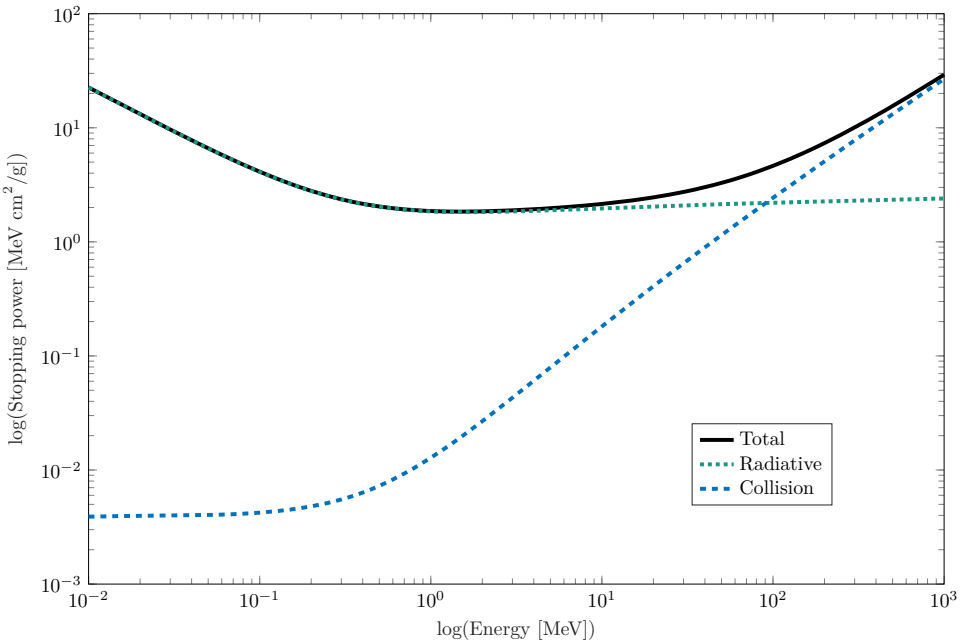


Figure 4.2: Total, radiative and collision stopping power of electrons in water. Data retrieved from EStar (based on NIST database) [17].

4.4 Proton transport

Proton therapy can offer large potential advantages over electron or photon therapy due to the characteristic way protons deposit dose along the tissue depth. Figure 4.3 shows the depth dose curves of proton beams in water. The position of the so-called *Bragg peak* visible in this graph is energy-dependent and can be exploited to achieve a high dose within the tumor while sparing preceding and surrounding tissue

or organs. In clinical application, energies of 70 to 250 MeV [148] are applied, which corresponds to a range, i.e., peak position, of 3 to 40cm depth in water.

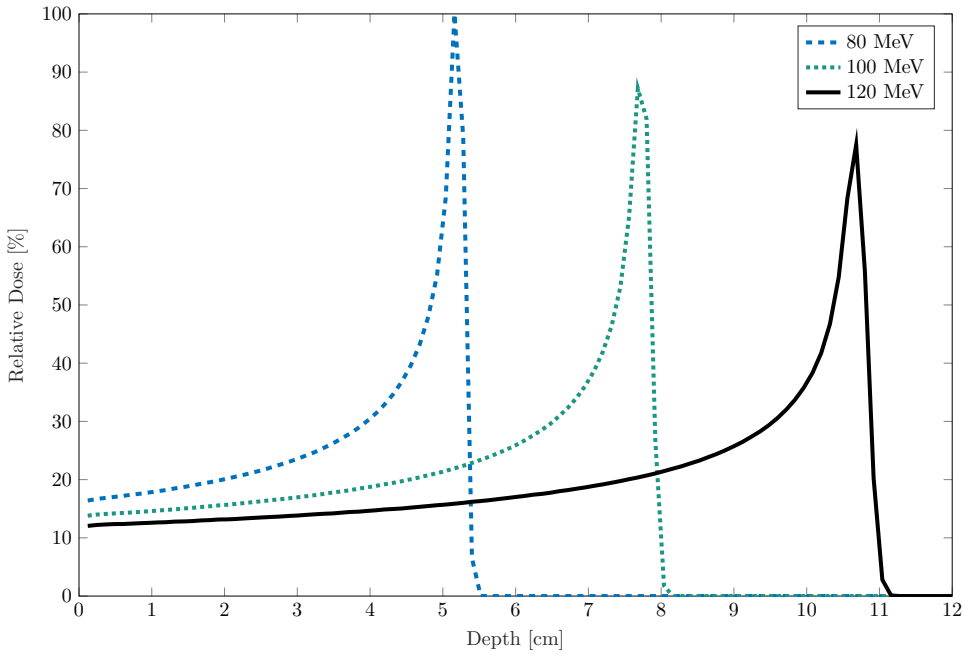


Figure 4.3: Depth dose curve, integrated along lateral dimensions, of a proton beam with 80 (dashed), 100 (dotted) and 120 (solid) MeV in water. Data generated using TOPAS MC [175].

The characteristic shape of the proton depth dose curve can be attributed to the interplay of different physical interactions. The most relevant of these interactions for a therapeutic energy range and how to obtain the corresponding cross sections will be discussed in the following sections.

4.4.1 Elastic scattering

(Multiple) Coulomb scattering

Similar to the electron, a significant part of proton interactions are governed by Coulomb forces. Recoiled by the positive charge of the nucleus, incident protons deviate from their previous direction of flight. This is referred to as *Coulomb scattering*. A single such scattering event can be described by the Rutherford formula [184].

$$\sigma_s = \left(\frac{1}{4\pi\epsilon_0} \frac{Z_t Z_p e^2}{4E_0} \right)^2 \cdot \frac{1}{\sin^4(\frac{\vartheta}{2})}, \tag{4.12}$$

where Z_t, Z_p are the atomic numbers of the target and projectile, respectively, ϵ_0 is the electric field constant, E_0 is the initial projectile energy (in Joule), e is the elementary electric charge and ϑ the deflection angle.

Remark 4.2. *The Rutherford cross sections determined by eq. (4.12) are stated in terms of a center-of-mass frame of reference, i.e. the deflection angles are defined by the relative difference between the vectors*

connecting the two involved particles (projectile and target) before and after the collision. We are however interested in knowing the deflection angle of the projectile particle in relation to its previous direction of flight defined in a static laboratory coordinate system. While the two are equivalent if the target particle remains stationary, this is not generally true in practice. According to [67], the angle ϑ in laboratory frame can be determined from the angle θ in a center of mass frame using the following relation

$$\cos(\vartheta) = \frac{\cos(\theta) + \frac{m_p}{m_t}}{\sqrt{1 + 2\frac{m_p}{m_t}\cos(\theta) + \left(\frac{m_p}{m_t}\right)^2}},$$

where m_p is the projectile (here proton) mass and m_t is the target mass. [215, 160] further demonstrate how the cross section can be directly transformed from one frame of reference to another.

With remark 4.2, the Rutherford cross section in laboratory frame takes the form:

$$\sigma_{s,lab} = \frac{\left(1 + 2\frac{A_p}{A_t}\cos(\theta) + \left(\frac{A_p}{A_t}\right)^2\right)^{\frac{3}{2}}}{1 + \cos(\theta)\frac{A_p}{A_t}} \cdot \left(\frac{1}{4\pi\epsilon_0}\frac{Z_t Z_p e^2}{4E_0}\right)^2 \cdot \frac{1}{\sin^4(\frac{\theta}{2})}, \tag{4.13}$$

where A_p, A_t are the mass numbers of projectile and target particles and θ is the deflection angle in center of mass frame.

Numerous such Coulomb interactions with small angular deflections occur along the trajectory of a particle. Therefore, often the aggregated deflection after a predefined path length is considered. For a large enough number of individual collisions, the so-called *multiple Coulomb scattering* angular distribution can be assumed to be Gaussian at small angles [140]. This is captured accurately within Molière's theory [149], but can also be modeled sufficiently well for the purpose of a Monte Carlo transport solver using simpler approximations such as (variations of) the Highland formula [77, 140].

Since most models for multiple Coulomb scattering are designed for Monte Carlo solvers, they often depend on the path length or traversed thickness, which is not explicitly defined within deterministic Boltzmann solvers. Molière's theory can also be used to derive differential scattering cross sections [see 192, 57, 27, 90]. However, applications within different radiative transport solvers [e.g. 154, 161] suggest, that within the significant parameter ranges, Rutherford's formula is sufficient to capture the relevant effects of elastic proton scattering due to Coulomb forces.

Nuclear scattering

In the case of high-energy protons, the incident particles can penetrate the outer atomic shell and scatter directly from the nucleus. This is referred to as nuclear elastic scattering. Cross sections for these interactions are available for example in the ICRU Report 63 [13]. Further, [58] introduces formulas derived from an analytical fit to the SAID data [8] for elastic nuclear proton scattering in oxygen and hydrogen. Nuclear scattering is however often neglected since it merely acts like a correction to the electromagnetic interactions [58].

Remark 4.3. *Note, that the discussed formulas for nuclear and Coulomb scattering typically compute microscopic cross sections σ, i.e. represent the interaction probability with a single particle of the background material. In water, the macroscopic cross section Σ for a specific target material T (hydrogen or*

oxygen) with density ρ_T can be determined using the following relation

$$\Sigma_s(\mathbf{\Omega} \cdot \mathbf{\Omega'}, E) = \rho_T \cdot N_A \frac{w_T}{A_T} \sigma_s(\mathbf{\Omega} \cdot \mathbf{\Omega'}, E),$$

where N_A is the Avogadro constant, w_T is the weight proportion of the target in water and A_T is the atomic weight of the target.

4.4.2 Stopping power & inelastic scattering

Inelastic scattering of protons can occur due to interactions with the electrons in the atomic shell or the nucleus itself and is the predominant cause of energy loss [152]. While bremsstrahlung is also emitted as a result of the latter interactions, it is negligible for the therapeutic energy ranges used with protons [215]. The angular deflection is also small and therefore often omitted [160, 154]. The energy loss caused by inelastic scattering is again modeled using the stopping power, divided into an electronic and a nuclear component according to the two different types of interactions. Their contributions to the total stopping power can be observed in fig. 4.4: Nuclear stopping only has a discernible contribution in very low energy ranges, whereas electronic stopping is responsible for the majority of energy loss. Values for the stopping power in different materials such as water can be retrieved using PStar, based on the NIST database [17].

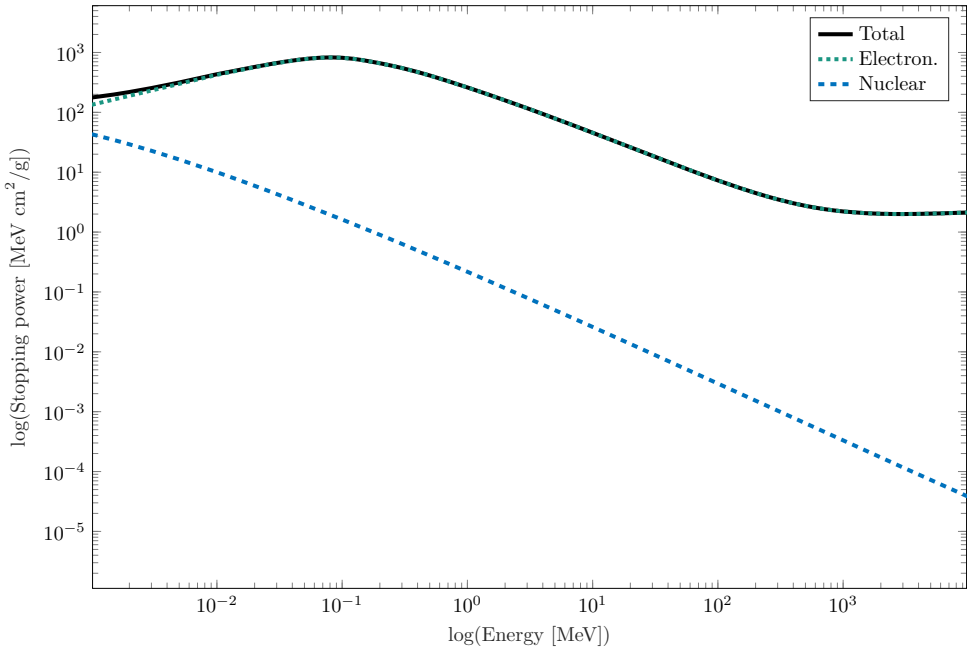


Figure 4.4: Total, nuclear and electronic stopping power of protons in water. Data retrieved from PStar (based on NIST database) [17].

4.4.3 Range/Energy straggling

In the continuous slowing down approximation, all particles are assumed to lose energy at the same rate defined by the stopping power. However, interactions with the material are actually stochastic and thus the amount of deposited energy through such interactions differs [18, 24, 152].

For protons, this largely manifests in a varying particle range, which is why the terms range and energy straggling are used interchangeably. While range or energy straggling can be explicitly included in the transport equation, the energy can then no longer be treated as a pseudo-time as derived for the CSD approximation (see section 4.2). This affects the numerical solution methods which can be applied and significantly increases the computation time [215]. Therefore, cheaper approximations have been developed [e.g. 25], where the number of particles that have lost all of their energy is assumed to be normally distributed [19]. The dose $\hat{D}(z)$ with range straggling at depth z is then approximated by folding the computed dose $D_1(z)$ without range straggling with the Gaussian distribution:

$$\hat{D}(z) = \int_0^\infty D_1(\bar{z}) \cdot \frac{e^{\frac{-(z-\bar{z})^2}{2\sigma_z^2(\bar{z})}}}{\sqrt{2\pi}\sigma_z(\bar{z})} \, \mathrm{d}\bar{z} \tag{4.14}$$

It has been shown that the depth-dependent standard deviation $\sigma_z(\bar{z})$ can be approximated well through a constant value $\sigma_z(\bar{z}) \equiv \sigma$. In [25] this value is chosen as:

$$\sigma \approx 0.012 \cdot R_0^{0.935},$$

where R_0 is the particle range, which can be computed from the kinetic energy using the Bragg-Kleemann rule (see eq. (11.5)). Note that in reality, the energy loss is described by a Landau distribution and is thus slightly asymmetric [20, 152]. Nevertheless, the approximation has been shown to produce good results in combination with an analytical pencil beam method [25]. However, inaccuracies have to be expected especially in heterogeneous materials.

4.5 Photon transport

Photon therapy is the most widespread modality of radiation therapy, mostly due to its early development and the relatively small space and costs required for building a treatment facility [74, 160]. Here, in contrast to the previously discussed electrons and protons, the production of secondary particles cannot be neglected. Secondary electrons are responsible for a majority of dose deposited in the patient. We therefore have to consider a more complex system of several coupled equations governing the transport process [74].

The depth dose curve for photons is illustrated in fig. 4.5. The influence of secondary electrons can mostly be seen in the dose build-up. After the dose maximum, charged particles are in equilibrium and the fall-off is characterized by exponential photon attenuation [9]. Though dose deposition by photons lacks the convenient peak exploited for proton therapy, it has a longer range than electrons, making photons suitable also for deeper seated tumors.

4.5.1 Scattering and production of secondary particles

The main physical interactions relevant for photon therapy are

1. *Compton scattering*: Here the photon collides with a free electron. Energy is transferred and both particles are scattered. This effect acts as a source for electrons and is therefore very relevant for dose deposition.

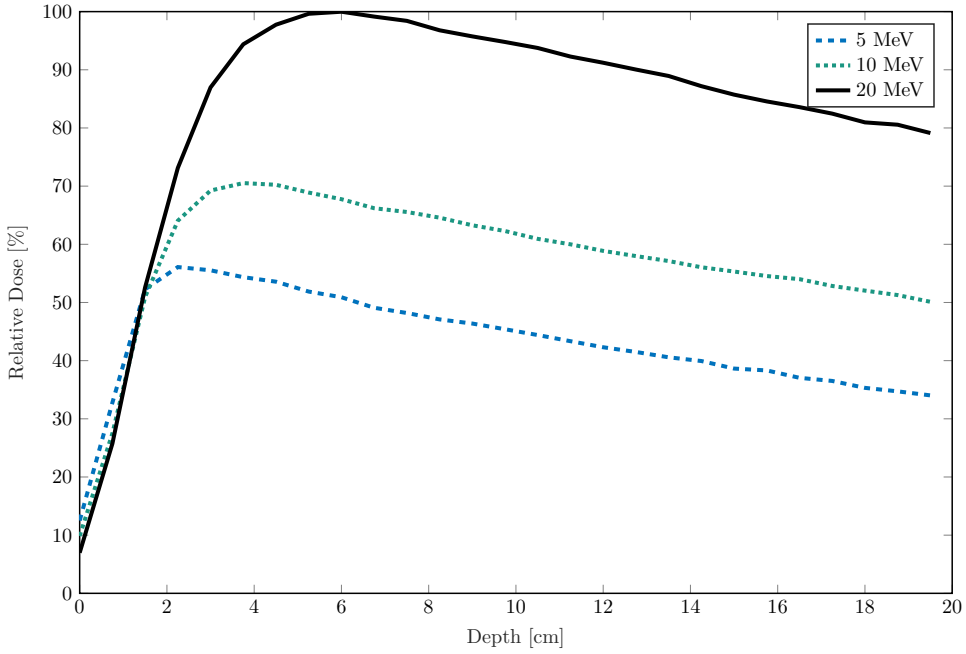


Figure 4.5: Depth dose curve, integrated along lateral dimensions, of a photon beam with 5 (dashed), 10 (dotted) and 20 (solid) MeV in water. Data generated using TOPAS MC [175].

2. *Elastic Coulomb/Mott scattering*: Here an incoming electron is deflected by the nucleus without losing energy.

3. *Inelastic Møller scattering*: Incoming electrons interact with outer shell electrons, both are deflected and energy is transferred from the incoming electron. Thus, this process acts as a further electron source.

4. *Pair production/annhilation*: During pair production a photon interacts with the atomic nucleus, emitting an electron-positron pair. This is only relevant for high-energy photons and never a dominating effect. Pair annihilation describes the opposite process of an electron-positron pair creating a photon.

5. *Bremsstrahlung*: see section 4.3.3, is negligible.

4.5.2 Coupled system of equations

The photon transport equation considering only Compton scattering is given by

$$\begin{aligned}\Omega_\gamma \cdot \nabla\psi_\gamma(\boldsymbol{r}, \boldsymbol{\Omega}_\gamma, E_\gamma) = &\int_0^\infty \int_{\boldsymbol{\Omega'}_\gamma \in \mathbb{S}^2} \Sigma_{C,\gamma}(\boldsymbol{r}, \boldsymbol{\Omega}_\gamma \cdot \boldsymbol{\Omega'}_\gamma, E'_\gamma \to E_\gamma)\psi_\gamma(\boldsymbol{r}, \boldsymbol{\Omega'}_\gamma, E'_\gamma)\mathrm{d}\boldsymbol{\Omega'}_\gamma \mathrm{d}E'_\gamma \\ &- \Sigma^t_{C,\gamma}(\boldsymbol{r}, E_\gamma)\psi_\gamma(\boldsymbol{r}, \boldsymbol{\Omega}_\gamma, E_\gamma) \ , \end{aligned} \tag{4.15}$$

where $\Sigma_{C,\gamma}$ and $\Sigma^t_{C,\gamma}$ are the differential and total cross sections for Compton scattering of photons, respectively. However, since photons do not deposit energy themselves, but through electrons and positrons

produced in their interactions with matter, we have to further consider the coupled electron transport equation

$$\begin{aligned}\Omega_e \cdot \nabla\psi_e(\boldsymbol{r},\boldsymbol{\Omega}_e,E_e) = &\int_0^\infty \int_{\boldsymbol{\Omega'}_\gamma\in\mathbb{S}^2_{1/2}} \Sigma_{C,e}(\boldsymbol{r},\boldsymbol{\Omega'}_\gamma\cdot\boldsymbol{\Omega}_e,E'_\gamma\to E_e)\psi_e(\boldsymbol{r},\boldsymbol{\Omega'}_\gamma,E'_\gamma)\mathrm{d}\boldsymbol{\Omega'}_e\mathrm{d}E'_e\\
&+\int_{E_s}^\infty \int_{\boldsymbol{\Omega'}_e\in\mathbb{S}^2_{1/4}} \Sigma_{M,e}(\boldsymbol{r},\boldsymbol{\Omega}_e\cdot\boldsymbol{\Omega'}_e,E'_e\to E_e)\psi_e(\boldsymbol{r},\boldsymbol{\Omega'}_e,E'_e)\mathrm{d}\boldsymbol{\Omega'}_e\mathrm{d}E'_e\\
&+\int_{E_s}^\infty \int_{\boldsymbol{\Omega'}_e\in\mathbb{S}^2_{2/4}} \Sigma_{M,\delta}(\boldsymbol{r},\boldsymbol{\Omega}_e\cdot\boldsymbol{\Omega'}_e,E'_e\to E_e)\psi_e(\boldsymbol{r},\boldsymbol{\Omega'}_e,E'_e)\mathrm{d}\boldsymbol{\Omega'}_e\mathrm{d}E'_e\\
&+\int_{\boldsymbol{\Omega'}_e\in\mathbb{S}^2} \Sigma_{Mott}(\boldsymbol{r},\boldsymbol{\Omega}_e\cdot\boldsymbol{\Omega'}_e,E_e)\psi_e(\boldsymbol{r},\boldsymbol{\Omega'}_e,E_e)\mathrm{d}\boldsymbol{\Omega'}_e\\
&-\Sigma^t_M(\boldsymbol{r},E_e)\psi_e(\boldsymbol{r},\boldsymbol{\Omega}_e,E_e)-\Sigma^t_{Mott}(\boldsymbol{r},E_e)\psi_e(\boldsymbol{r},\boldsymbol{\Omega}_e,E_e),\end{aligned} \tag{4.16}$$

where $\Sigma_{C,e}$ and $\Sigma^t_{C,e}$ are the differential and total cross sections for Compton scattering of electrons. Further, $\Sigma_{M,e}$ and $\Sigma_{M,\delta}$ are the cross sections for primary and secondary ("delta-rays") electrons for Møller-scattering and Σ^t_M is the corresponding total cross section. Lastly, Σ_{Mott} (Σ^t_{Mott}) is the (total) cross section for elastic Mott-scattering. Further information on how to obtain these cross sections can be found for example in [45, 74].

Note that here interactions from the previous section which produce electrons appear as in-scattering sources. If one wanted to also consider pair production and annihilation, this would introduce a third, coupled positron equation.

Despite the complexity of its governing equations, dose computations for photons are typically less time-consuming than in the case of protons. This is because the transport of primary photons is computationally inexpensive and secondary electrons deposit dose locally and virtually positionally invariant. This allows for the use of computationally cheap approximations such as the pencil beam method (section 5.2). Due to the structure of their dose deposition curve, photon therapy is also less sensitive to uncertainties compared to proton therapy [132, 130]. This robustness however reduces the necessity for further research on both better numerical solutions of the transport equation and uncertainty quantification. Further, due to their lack of charge, the Coulomb force is not as deciding for the behavior of photons and thus the continuous slowing down approximation does not hold. For these reasons, photons will not be further discussed in this thesis and we will focus on charged particle transport in the following.

CHAPTER 5

Background: Numerical solution methods

In this chapter, we will recall the most common numerical solution methods to the previously discussed transport equations. Due to the widespread and especially interdisciplinary interest in transport problems, they have been addressed using methods from different fields of mathematics. As discussed in the preface to this work, we will thus have to switch between a stochastics and numerics perspective depending on the approach in each section.

The Boltzmann- or grid-based methods are most widely used in the mathematical field of transport theory. Here, the focus lies on a numerical discretization of all phase space variables in eq. (4.5). The Monte Carlo method on the other hand focuses on a probabilistic view of particle dynamics, where the properties of particles at their source follow a probability distribution function and their behaviour over time is a stochastic process. The pencil beam method uses a macroscopic approximation, also based partly on probabilistic assumptions combined with knowledge about dose deposition in a reference material. This point of view offers a more intuitive description of particle behavior. Therefore, the Monte Carlo and pencil beam method are especially popular in (medical) physics applications.

5.1 Monte Carlo method

The Monte Carlo method [147, 223] is a numerical integration method, which is based on evaluating the integrand at random positions sampled from the probability distribution of the integration variable. For realizations $\boldsymbol{x}_i\,, i = 1, ..., N$ of the random variable $\boldsymbol{X} \in \mathbb{R}^d$ with probability density function $p : \mathbb{R}^d \to \mathbb{R}_+$, the estimator is defined as follows:

$$I(f) = \mathbb{E}_p[f(\boldsymbol{X})] = \int_{\mathbb{R}^d} f(\boldsymbol{X})p(\boldsymbol{X})\mathrm{d}\boldsymbol{X} \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i) := I^N(f)\,. \tag{5.1}$$

Note, that with the *Strong Law of Large Numbers*, one can show

$$\mathbb{P}\left(\lim_{N\to\infty} I^N(f) = I(f)\right) = 1\,, \tag{5.2}$$

i.e., the Monte Carlo estimator converges to the true value in probability. Further, the independence of the random variables $\boldsymbol{x}_i$ and the *Central Limit Theorem* provide a formula for the variance of the estimator:

$$Var(I^N(f)) = \frac{v_f^2}{N}\,, \tag{5.3}$$

which depends solely upon the variance v_f^2 of the integrand and the number of samples N. The estimator thus converges with $O(\frac{1}{\sqrt{N}})$ and its accuracy can be controlled through the number of computed realizations.

5.1.1 The Boltzmann equation in integral form

In order to solve eq. (4.1) using the Monte Carlo method, it needs to be given in its integral form. Given the integro-differential equation stated in eq. (4.1), this integral equation can be formally derived by integrating along rays as outlined in [206]. Here, we want to use the description of radiation transport as a stochastic process to directly and more intuitively obtain the required form.

Integral equations for random walks

Let $\{\mathbf{X}_i, i \geq 0\}$ be a random walk in $\mathbb{R}^d$ with the probability distribution function (referred to as *source*) $S(\mathbf{X})$ with

$$S(\mathbf{X}) \geq 0,\ \int S(\mathbf{X})\mathrm{d}\mathbf{X} = 1$$

and the transition distribution $T(\mathbf{X} \mid \mathbf{X}')$ with

$$T(\mathbf{X} \mid \mathbf{X}') \geq 0,\ \int T(\mathbf{X} \mid \mathbf{X}')\mathrm{d}\mathbf{X} \leq 1.$$

Here, the fact that the transition distribution does not integrate to one reflects to possibility for the walk to terminate at $\mathbf{X}'$.

The average of the distribution of arrivals at $\mathbf{X}$, $\mathcal{A}(\mathbf{X})$, is given by [230]

$$\mathcal{A}(\mathbf{X}) = S(\mathbf{X}) + \int T(\mathbf{X} \mid \mathbf{X}')\mathcal{A}(\mathbf{X}')\mathrm{d}\mathbf{X}'.$$

Radiation transport as a stochastic process

In order to derive the integral Boltzmann equation, we need to identify the functions governing the corresponding stochastic process. After their initialization according to a source or boundary conditions, particles can be viewed to follow a random walk $\{(\mathbf{r}, \boldsymbol{\Omega}, E)_i, i \geq 0\}$. Here, the state or position of the random walk is the particle phase space $(\mathbf{r}, \boldsymbol{\Omega}, E)$. The quantity of interest corresponding to the angular flux, i.e. probable number of particles, at state $(\mathbf{r}, \boldsymbol{\Omega}, E)$ is $\psi(\mathbf{r}, \boldsymbol{\Omega}, E)$. When just considering primary particles which are created at the boundary, the source term S_0 in the integral equation is induced only

by the boundary condition $\psi_{BC}(\boldsymbol{r}, \boldsymbol{\Omega}, E)$, $\mathbf{r} \in \partial\mathcal{X}$. In terms of a random walk this source determines the distribution of the initial state/phase space parameters, i.e., $(\boldsymbol{r_0}, \boldsymbol{\Omega}_0, E_0) \sim S_0$. In order to include particle interactions which produce secondary particles, other source terms would have to be considered. For simplicity, we however omit these in the following, as they are not relevant for the electron and proton physics models used in this work.

What remains to be derived is the transition distribution $T((\mathbf{r}, \boldsymbol{\Omega}, E) \mid (\mathbf{r}', \boldsymbol{\Omega}', E'))$, i.e. the distribution of next steps $(\mathbf{r}, \boldsymbol{\Omega}, E)$ in the random walk, given the current state is $(\mathbf{r}', \boldsymbol{\Omega}', E')$. The transition probability from one phase space state to another can be seen as the probability for a particle to

1. move from $\mathbf{r}'$ to $\mathbf{r}$ without collision
2. and then undergo an interaction at $\mathbf{r}$, which changes its direction of flight to $\boldsymbol{\Omega}$ and energy to E.

Let's collect the probabilities for the terms above: The probability for a particle to fly from $\mathbf{r}'$ to $\mathbf{r}$ without collision is proportional to [44, 230]

$$\exp\left(-\int_{r'}^{r} \Sigma_t(\mathbf{r}'', E)\mathrm{d}\mathbf{r}''\right). \tag{5.4}$$

The cross section $\Sigma_s(\boldsymbol{r}, \boldsymbol{\Omega}' \cdot \boldsymbol{\Omega}, E' \to E)$ precisely describes the probability for a particle to undergo a scattering event at $\mathbf{r}$, which changes its direction from $\boldsymbol{\Omega}'$ to $\boldsymbol{\Omega}$ and energy from E' to E. We can put the terms together to determine the transition probability per unit length [230]

$$T((\mathbf{r}, \boldsymbol{\Omega}, E) \mid (\mathbf{r}', \boldsymbol{\Omega}', E')) = \exp\left(-\int_{r'}^{r} \Sigma_t(\mathbf{r}'', E)\mathrm{d}\mathbf{r}''\right) \cdot \Sigma_s(\boldsymbol{r}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}', E' \to E).$$

Alternatively, the source could directly generate particles at state $(\mathbf{r}, \boldsymbol{\Omega}, E)$. In our application case, this would only happen with positive probability at $\mathbf{r} \in \partial\mathcal{X}$. We now want to trace particles starting from their creation at the boundary along their trajectories, which consist of straight lines between points of interaction. A particle with $(\mathbf{r}_0, \boldsymbol{\Omega}_0)$ can be traced along the line $\mathbf{r}_0 + s\boldsymbol{\Omega}_0$ until its next collision at distance s. For simplicity, we rename $\mathbf{r}_0 + s\boldsymbol{\Omega}_0 \to \mathbf{s}$.

Plugging everything into the integral formula for random walks, we get

$$\begin{aligned}\psi(\boldsymbol{s}, \boldsymbol{\Omega}, E) &= S_0(\boldsymbol{s}, \boldsymbol{\Omega}, E) \\ &+ \int_0^s \int_0^\infty \int_{\boldsymbol{\Omega}' \in S^2} \left(e^{-\int_{s'}^{s} \Sigma_t(\mathbf{s}'', E)\mathrm{d}\mathbf{s}''} \cdot \Sigma_s(\boldsymbol{s}, \boldsymbol{\Omega} \cdot \boldsymbol{\Omega}', E' \to E)\right) \psi(\boldsymbol{s}', \boldsymbol{\Omega}', E')\mathrm{d}\boldsymbol{\Omega}'\mathrm{d}E'\mathrm{d}\boldsymbol{s}'. \end{aligned} \tag{5.5}$$

Equation (5.5) is a Fredholm equation of the second kind [221], i.e. has the form

$$\psi(\mathbf{z}) = \psi_0(\mathbf{z}) + \int \psi(\mathbf{z}')k(\mathbf{z}' \to \mathbf{z})\mathrm{d}\mathbf{z}'.$$

It can be solved by sampling the associated random walk or Markov chain [221, 230].

First the initial state, here phase space coordinates $(\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)$, are sampled from their probability distribution S_0, which can for example be modeled as a multivariate Gaussian probability distribution. The subsequent states are reached by simulating particle trajectories and random interactions according to

the respective probability distributions. Here, the interaction probabilities are defined by the scattering/absorption cross sections and the length of paths of free flight is exponentially distributed (comp. 5.4). For the most accurate solution, this simulation is done event by event for each individual particle, always sampling the distance to the next collision, the type of collision, energy loss and scattering angle [221]. However, in order to increase computation speed, different variations of this method exist, such as condensed history algorithms which only explicitly simulate "hard" collision events within a certain step size distance. We will not elaborate this further here, the interested reader might take a look at [221] or [179] for a more concise derivation of the probability measures, the resulting solution algorithm and different Monte Carlo variations.

The Monte Carlo dose calculation algorithm can then be seen as a black box $BB(\boldsymbol{r};\boldsymbol{Z})$, returning the dose $D(\boldsymbol{r})$ at position $\boldsymbol{r}$ for the initial particle phase space $\boldsymbol{Z} = (\boldsymbol{r_0}, \boldsymbol{\Omega}_0, E_0) \sim S_0$ on $\mathcal{D}_{S_0}$.

$$D(\boldsymbol{r}) = \mathbb{E}_{S_0}[BB(\boldsymbol{r};\boldsymbol{Z})] = \int_{\mathcal{D}_{S_0}} BB(\boldsymbol{r};\boldsymbol{Z}) S_0(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} \approx \frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p) \ , \tag{5.6}$$

where the realisations $\boldsymbol{z}_1, ..., \boldsymbol{z}_H$ are sampled from $S_0(\boldsymbol{Z})$. Note, that the output of the black box for fixed realizations of only the input parameters is still a random variable, due to the stochasticity of the path lengths and interactions of the particles. Each response $BB(\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)$ therefore also implicitly includes realizations of the probability distributions governing these interactions. We omit this in the following for the sake of brevity.

In this work, S_0 is assumed to be Gaussian for each beamlet $b = 1, ..., B$:

$$S_0^b = \mathcal{N}(\boldsymbol{\mu}_Z^b, \boldsymbol{\Lambda}^b). \tag{5.7}$$

Then in IMRT, the source distribution for particles from all layered beamlets is a Gaussian mixture

$$S_0 = \sum_{b=1}^{B} w_b S_0^b = \sum_{b=1}^{B} w_b \cdot \mathcal{N}(\boldsymbol{\mu}_Z^b, \boldsymbol{\Lambda}^b) \ . \tag{5.8}$$

Note, that while the assumption of normal distributions is a common one [233, 217, 174, 10, 61, 226], point sources (mainly for radioactive sources), uniform distributions or the use of phase space files is also widespread [e.g. 220].

5.2 Pencil beam algorithm

One of the simplest and also most frequently used dose calculation methods is the *pencil beam algorithm* [81, 177, 3, 84]. Here the dose deposited by infinitesimally thin, so-called *pencil beams* is considered to be known in a reference medium. This *pencil beam kernel* is then convolved with the incident fluence of an incoming beam to determine the dose distribution for a given beam set-up.

The algorithm itself consists of ray-tracing along the pencil beam's central axis and modeling the dose deposition at each point along the rays through the simple kernel function. The dose can then be defined as

$$
\begin{aligned}
D(\boldsymbol{r}) &= \sum_j w_j D_j(\boldsymbol{r}) \\
&= \sum_j w_j L_j^x(\boldsymbol{r}) L_j^y(\boldsymbol{r}) R_j(\boldsymbol{r}) \\
&= \sum_j w_j \mathcal{N}(r_x; \mu_x^{(j)}, \Sigma_x^{(j)}) \cdot \mathcal{N}(r_y; \mu_y^{(j)}, \Sigma_y^{(j)}) \cdot R_j(\boldsymbol{r})
\end{aligned} \tag{5.9}
$$

For protons, the pencil beam kernel and spread along the two lateral dimensions can be modeled by Gaussians [68, 200], whereas the depth dependent scaling $R_j(\boldsymbol{r})$ is usually done according to measured or MC simulated depth dose curves [25]. Here, $\Sigma_{x/y}^{(j)}$ is computed as

$$
\Sigma_{x/y}^{(j)}(z_{rad}) = {\sigma_{x/y;init}^{(j)}}^2 + \sigma_{x/y;MCS}^{(j)}(z_{rad})^2, \tag{5.10}
$$

where ${\sigma_{x/y;init}^{(j)}}^2$ describes the initial width of the pencil beam and $\sigma_{x/y;MCS}^{(j)}(z_{rad})^2$ the depth dependent spread according to a Gaussian approximation to multiple Coulomb scattering (see section 4.4.1). Further, z_{rad} is the radiological depth on the central beam axis, which is computed using ray-tracing and the measured depth dose adjusted by the sum of CT densities along the ray up to the current point [197].

The pencil beam algorithm is very computationally efficient due to its simplified deterministic form. Since densities are only considered along the central axis, the standard pencil beam algorithm is however not able to incorporate lateral inhomogeneities aside from the central axis, well. Several approaches exist to alleviate this problem [186, 209, 201], e.g. *fine sampling* decomposes the lateral Gaussians into several Gaussian components to allow for density-related changes of the kernel along the lateral dimensions [201].

5.3 Boltzmann/Grid-based methods

Deterministic, numerical methods based on discretizations of eq. (4.1) are often referred to as *Boltzmann-solvers*. These are opposed to stochastic methods such as the Monte Carlo method and deterministic methods based on extremely simplified point-spread functions referred to as *kernel-based* methods, such as the *pencil beam algorithm* introduced in section 5.2.

To allow for numerical approximations of solutions to the Boltzmann equation, (pseudo-)time/energy, space and angle need to be discretized. Artifacts that numerical solutions exhibit depend crucially on the angular discretization, for this reason deterministic Boltzmann-solvers are classified by the type of angular discretization into nodal and modal methods. Note, that in the following we will use the energy as a pseudo-time t as derived in section 4.2.

5.3.1 Modal methods

For the modal discretization of eq. (4.5), we span the solution in terms of spherical harmonics basis functions.

For degree ℓ and order k, the spherical harmonics for an angle $\mathbf{\Omega} = (\sqrt{1-\mu^2}\cos\varphi, \sqrt{1-\mu^2}\sin\varphi, \mu)^T$, with $\mu \in [-1,1]$ and $\varphi \in [0, 2\pi]$, are defined as

$$Y_\ell^k(\mathbf{\Omega}) = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-k)!}{(\ell+k)!}}\, e^{ik\varphi} P_\ell^k(\mu) \,,$$

where P_ℓ^k are the associated Legendre polynomials. In this work, we use the real spherical harmonics

$$m_\ell^k = \begin{cases} \frac{(-1)^k}{\sqrt{2}}\left(Y_\ell^k + (-1)^k Y_\ell^{-k}\right), & k > 0 \,, \\ Y_\ell^0 & k = 0 \,, \\ -\frac{(-1)^k i}{\sqrt{2}}\left(Y_\ell^{-k} - (-1)^k Y_\ell^k\right), & k < 0 \,. \end{cases}$$

Let us collect all basis functions up to degree N in a vector

$$\mathbf{m} = (m_0^0, m_1^{-1}, m_1^0, m_1^1, \cdots, m_N^N)^T \in \mathbb{R}^{(N+1)^2}$$

and define the so called moments

$$u_l^k(\boldsymbol{r}, t) := \int_{\mathbb{S}^2} \psi(\boldsymbol{r}, \mathbf{\Omega}, t) m_l^k(\mathbf{\Omega}) \, \mathrm{d}\mathbf{\Omega}. \tag{5.11}$$

Evolution equations for these moments are derived by testing the transport equation against $\mathbf{m}$ and yield an unclosed system. The different modal methods can then be interpreted as different closures [125, 124].

P_N method

The most common modal discretization technique is the *spherical harmonics* (P_N) method [32]. Here the model is truncated at order N, setting moments of order $N+1$ to zero or choosing the modal approximation $\psi(\mathbf{r}, \mathbf{\Omega}, t) \approx \mathbf{u}(\mathbf{r}, t)^T \mathbf{m}(\mathbf{\Omega})$. Then, the P_N equations for $\mathbf{r} = (r_x, r_y, r_z)^T$ read

$$\partial_t \mathbf{u}(\mathbf{r}, t) = -\mathbf{A} \cdot \nabla_{\mathbf{r}} \frac{\mathbf{u}(\mathbf{r}, t)}{\rho(\mathbf{r})} - \Sigma_t(t)\mathbf{u}(\mathbf{r}, t) + \mathbf{\Sigma}\mathbf{u}(\mathbf{r}, t), \tag{5.12}$$

where $\mathbf{A} \cdot \nabla_{\mathbf{r}} := \mathbf{A}_x \partial_{r_x} + \mathbf{A}_y \partial_{r_y} + \mathbf{A}_z \partial_{r_z}$ with $\mathbf{A}_i := \int_{\mathbb{S}^2} \mathbf{m}\mathbf{m}^T \Omega_i \, \mathrm{d}\mathbf{\Omega}$, $i \in \{x, y, z\}$. The diagonal in-scattering matrix $\mathbf{\Sigma}$ has entries $\Sigma_{kk}(t) = 2\pi \int_{[-1,1]} P_k(\mu)\Sigma_s(\mu, t) \, \mathrm{d}\mu$. Note that $\Sigma_t(t) = \Sigma_{11}(t) > 0$ and

$$|\Sigma_{kk}(t)| \leq 2\pi \int_{[-1,1]} |P_k(\mu)| \cdot |\Sigma_s(\mu, t)| \, \mathrm{d}\mu \leq 2\pi \int_{[-1,1]} |\Sigma_s(\mu, t)| \, \mathrm{d}\mu = \Sigma_{11}(t) \,. \tag{5.13}$$

While the P_N method is computationally efficient, it has some drawbacks. It can produce oscillations and further does not preserve the positivity of the solution. There are however approaches to mitigate these effects [144].

M_N method

The M_N method [124, 125] closes the system by solving an optimization problem, which minimizes the mathematical entropy $\mathcal{H}(g) := \int_{\mathbb{S}^2} \eta(g)\, \mathrm{d}\mathbf{\Omega}$, where $\eta : \mathbb{R}_+ \to \mathbb{R}$ is the so-called *kinetic entropy density* [see e.g. 6].

Thus ψ is chosen from the set $\mathcal{F} = \left\{ g(\boldsymbol{r}, \mathbf{\Omega}, t) > 0 : \mathbf{u} = \int_{\mathbb{S}^2} \mathbf{m} g \, \mathrm{d}\mathbf{\Omega} < \infty \right\}$ by solving

$$\begin{aligned} &\min_{g \in \mathcal{F}} \int_{\mathbb{S}^2} \eta(g) \, \mathrm{d}\mathbf{\Omega} \\ &\text{s.t.} \quad \mathbf{u} = \int_{\mathbb{S}^2} \mathbf{m} g \, \mathrm{d}\mathbf{\Omega} \, . \end{aligned} \tag{5.14}$$

The M_N method has the advantage of preserving important properties of the underlying equation, including positivity of the solution [125, 6]. However, the need to repeatedly solve eq. (5.14) makes it relatively computationally expensive, especially for realistic high-dimensional transport problems [190]. For this reason, it is not as widespread as the P_N discretization for radiation therapy applications, with the exception of occasional work on photon [12] and recently also proton transport [154].

5.3.2 Nodal methods

A conventional nodal method is the *discrete ordinates* (S_N) [126] method, which evolves the solution on a chosen angular quadrature set. In radiative transport, the directional domain is usually assumed to be the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$, a suitable parametrization is therefore given by spherical coordinates

$$\mathbb{S}^2 = \left\{ \begin{pmatrix} \sqrt{1-\mu^2}\sin(\varphi) \\ \sqrt{1-\mu^2}\cos(\varphi) \\ \mu \end{pmatrix} : \mu \in [-1,1] \, , \, \varphi \in [0, 2\pi) \right\} . \tag{5.15}$$

Thus, the numerical flux is evaluated at N_Q discrete nodes $\left\{ \mathbf{\Omega}_1, ..., \mathbf{\Omega}_{N_Q} \right\} \subset \mathbb{S}^2$. Note, that different particle velocities can be incorporated by scaling the unit sphere with a given maximum velocity. The nodes and corresponding weights are determined according to the chosen quadrature rule, common choices include for example a Gaussian quadrature, spherical Monte Carlo, Levelsymmetric [134], LEBEDEV [141] and LDFESA [93]. A comparison of different quadrature sets and their approximation behaviour for S_N methods can be found in [31].

The evolution equations for nodes $\mathbf{\Omega}_q$, weights w_q and $\psi_q(\mathbf{r}, t) := \psi(\mathbf{r}, \mathbf{\Omega}_q, t), \; q = 1, ..., N_Q$ are then given by

$$\partial_E \psi_q(\mathbf{r}, t) + \mathbf{\Omega}_q \cdot \nabla_r \frac{\psi_q(\mathbf{r}, t)}{\rho(\mathbf{r})} + \Sigma_t(t) \psi_q(\mathbf{r}, t) = \sum_{p=1}^{Q} w_p \Sigma_s(\mathbf{\Omega}_q \cdot \mathbf{\Omega}_p, t) \psi_p(\mathbf{r}, t) \, . \tag{5.16}$$

A main disadvantage of S_N methods are so called ray-effects [121, 150, 142], which are spurious artifacts that stem from the limited number of directions in which particles can travel. Moreover, radiation therapy applications exhibit forward-peaked scattering, which cannot be captured well by classical quadrature rules.

To allow for moderate computational costs when computing scattering terms and to efficiently treat forward-peaked scattering, it is also possible to combine nodal and modal methods. This will be discussed further in chapter 7.

5.3.3 Spatial and time/energy discretization

Having discretized the angular variable $\mathbf{\Omega}$, the spatial as well as time/energy dimensions remain to be discretized. Both nodal and modal methods result in a system of transport equations, that can be solved using a finite volume scheme. In the following, we demonstrate the implementation of an exemplary finite volume method with an explicit Euler discretization and first order upwind flux.

In chapter 6, the spatial domain will be divided into triangular, unstructured grids in two spatial dimensions, while chapter 7 will use 2D and 3D quadrilateral structured grids. Thus, to keep it as general as possible, we consider a not necessarily structured grid $\tilde{\mathbf{X}} = \{\mathbf{X}_i\}_{i=1,\dots,N_x}$, which is a partition of a bounded spatial domain $\mathcal{X} \subset \mathbb{R}^d$. In the following, we refer to individual grid cells as $\mathbf{X}_i$, with centroid $\mathbf{c}_i$ and measure V_i. $N(i)$ denotes the indices of neighboring cells of X_i and $\mathbf{n}_{i,j}$ the unit-normal vector of the interface dividing cell i and its neighbor $j \in N(i)$.

Consider a system of transport equations

$$\partial_t \psi(\mathbf{r}, t) + \nabla_{\mathbf{r}} \cdot \mathbf{F}(\psi(\mathbf{r}, t)) = \mathbf{R}(\mathbf{r}, t, \psi(\mathbf{r}, t)), \qquad \mathbf{r} \in \mathcal{X},\ t \in [0, t_{end}) \tag{5.17}$$

where $\mathbf{F}$ describes advection and the right hand side $\mathbf{R}$ contains velocity discretizations of collision terms, sources and absorption terms. As we have already eliminated the angular dependence in the previous sections, we omit the dependence on $\mathbf{\Omega}$ here. Finite volume methods consider the solution as an average over one space-energy cell. First, we handle the time discretization and divide the domain into N_t discrete values $0 = t_0 < \cdots < t_{N_t - 1}$. Let $\psi(\mathbf{r}, t_n) := \psi^n(\mathbf{r})$ be the solution at time t_n and define the averaged solution over one time step $[t_n, t_{n+1}]$ as

$$\bar{\psi}^n(\mathbf{r}) := \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \psi(\mathbf{r}, t) \mathrm{d}t.$$

Then eq. (5.17) integrated over one time step $[t_n, t_{n+1}]$ is given by

$$\frac{\psi^{n+1}(\mathbf{r}) - \psi^n(\mathbf{r})}{\Delta t} + \nabla_{\mathbf{r}} \cdot F(\bar{\psi}^n(\mathbf{r})) = \mathbf{R}(\mathbf{r}, t, \bar{\psi}^n(\mathbf{r})), \tag{5.18}$$

where $\Delta t = t_{n+1} - t_n$. Assuming we solve the equation iteratively from t_0 to t_{N_t}, the quantities $\bar{\psi}^n(\mathbf{r})$ and $\psi^{n+1}(\mathbf{r})$ are unknown. The system can, for example, be closed by approximating $\bar{\psi}^n$ by the values at the upper (implicit Euler scheme) or lower bound (explicit Euler scheme) of the time interval. Here, we use an explicit Euler scheme, which yields

$$\frac{\psi^{n+1}(\mathbf{r}) - \psi^n(\mathbf{r})}{\Delta t} + \nabla_{\mathbf{r}} \cdot F(\psi^n(\mathbf{r})) = \mathbf{R}(\mathbf{r}, t, \psi^n(\mathbf{r})). \tag{5.19}$$

With the spatial discretization into N_x cells, the average over one cell is defined as

$$\psi_i^n = \frac{1}{V_i} \int_{\mathbf{X}_i} \psi^n(\mathbf{r}) \mathrm{d}\mathbf{r}.$$

We then integrate eq. (5.17) over a time step (see eq. (5.19)) and one spatial cell and solve the integrals in the advection term using Gauss' theorem. This yields

$$\frac{1}{\Delta t}\left(\psi_i^{n+1} - \psi_i^n\right) + \frac{1}{\Delta t V_i}\int_{t_n}^{t_{n+1}} \sum_{j \in N(i)} \mathbf{F}(\psi(\mathbf{x}_{i,j}, t)) \cdot \mathbf{n}_{i,j}\, \mathrm{d}t = \mathbf{R}(\mathbf{r}, t, \psi_i^n)), \tag{5.20}$$

where $\psi(\mathbf{x}_{i,j}, t)$ is the conserved variable evaluated at the interface between cell i and its neighbor j. In order to compute ψ_i^{n+1} we choose the common ansatz

$$F(\psi_j^n, \psi_i^n) \approx \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} \mathbf{F}(\psi(\mathbf{x}_{i,j}, t)) \cdot \mathbf{n}_{i,j}\, \mathrm{d}t, \tag{5.21}$$

where the numerical flux $F(\psi_j^n, \psi_i^n)$ at face (i, j) is approximated using the cell averaged values at cell i and j. For transport equations, a well known numerical flux is given by the Upwind scheme [123]

$$F(\psi_j^n, \psi_i^n)_{up} = \mathbf{F}(\psi_i^n) \cdot \mathbf{n}_{i,j} H(\mathbf{n}_{i,j} \cdot \mathbf{\Omega}) + \mathbf{F}(\psi_j^n) \cdot \mathbf{n}_{i,j} \left(1 - H(\mathbf{n}_{i,j} \cdot \mathbf{\Omega})\right), \tag{5.22}$$

where H is the heaviside step function. The explicit solution iteration of a first order scheme can now be obtained by approximating the source, absorption and collision terms using current cell averages:

$$\psi_i^{n+1} = \psi_i^n - \frac{\Delta t}{V_i} \sum_{j \in N(i)} F(\psi_j^n, \psi_i^n)_{up} + \Delta t\, \mathbf{R}(\mathbf{x}_i, t, \psi_i^n). \tag{5.23}$$

CHAPTER 6

KiT-RT: A software framework for kinetic transport in radiation therapy

In [112], we introduce a software package, which consolidates several variations of the three deterministic Boltzmann CSD solvers introduced in the previous chapter 5 for applications in electron radiation therapy. This makes it possible for the first time, to directly compare the classical numerical Boltzmann solvers in a realistic setting, while also being easily extendable for research and validation of new approaches. Note, that KiT-RT further contains solvers for the original linear Boltzmann equation for flux computations in general radiative transport. These will however not be further discussed here, as we focus on the specific application to radiation therapy using the continuous slowing down approximation.

6.1 Implementation and parallelization

All continuous slowing down solvers included in KiT-RT follow the scheme presented in algorithm 6.1. Here, the flux update, finite volume update and subsequent dose calculation are parallelized using OpenMP and parallel for-loops when iterating over the grid cells.

The different numerical methods S_N, P_N and M_N differ in their numerical flux functions, which are required for the finite volume method. The numerical flux for the S_N method is given by

$$\mathbf{F}(\psi_{\mathbf{i}}^{\mathbf{n}}) = \mathbf{\Omega} \otimes \frac{\psi_i^n}{\rho(\mathbf{x_i})}, \tag{6.1}$$

while that for the P_N method reads

$$\mathbf{F}(\mathbf{u}_{\mathbf{i}}^{\mathbf{n}}) = \left[\mathbf{A}_1\mathbf{u}_i^n, \mathbf{A}_2\mathbf{u}_i^n, \mathbf{A}_3\mathbf{u}_i^n\right]^T, \tag{6.2}$$

where A_i are the flux Jacobians emerging from the spherical harmonics recursion scheme. In contrast to

Algorithm 6.1 KiT-RT CSD solver execution

```
Read settings from config file
Create solver
Create/Load mesh and densities from CT file
Preprocessing: Set BC, Transform energies, compute stopping powers
for Energies E_0 to E_max do
    Compute scattering coefficients
    Flux update
    Finite volume update
    Dose update
    Write Output
end for
Save Output
```

the S_N and P_N method, the flux function of the M_N method takes a more complex form

$$\mathbf{F}(\mathbf{u}_\mathbf{i}^\mathbf{n}) = \int_{\mathbb{S}^2} \mathbf{\Omega} \otimes \mathbf{m}(\mathbf{\Omega}) \psi_{\mathbf{u}_n^i}(\mathbf{\Omega}) \, \mathrm{d}\mathbf{\Omega}. \tag{6.3}$$

where $\psi_{\mathbf{u}_n^i}$ is the reconstructed radiation flux density of the minimal entropy closure at the cell averaged moment $\mathbf{u}_n^i$. To evaluate the numerical fluxes with an upwind scheme, the formulas above are plugged into eq. (5.22). Note, that for the M_N method we need to discretize the velocity integral. This can for example be achieved using a quadrature rule.

Further, it has to be ensured that the updated solution of the M_N method remains a feasible moment for the minimal entropy closure. This can be achieved for example by a flux-limiter [106], a realizability reconstruction [109] or by using a regularized entropy closure formulation [6].

6.2 Software design and architecture

The design principle of the KiT-RT software package is focused on efficient implementation, high reusability of its components and ease of extension. It contains a set of efficient numerical solvers for radiation transport, which are constructed of basic, reusable building blocks. These building blocks can be freely arranged to implement new solvers or tools for completely different applications.

On the other hand, KiT-RT is equipped with an easy to use command line interface based on readable configuration files, which allow easy manipulation of the solvers. Thus the software is attractive for developers, who want to experiment with the framework and build their own numerical solvers as well as users and application engineers, who want to gain experimental insights without directly interfering with the codebase.

KiT-RT is implemented in modern C++ and uses mainly polymorphism for its construction. In fig. 6.1, we present the class structures used to build the numerical CSD solvers. Note that the diagram merely presents the subsection of the total classes, which is relevant to this thesis and should serve more as an insight into the software structure than a comprehensive overview. Most building blocks consist of a

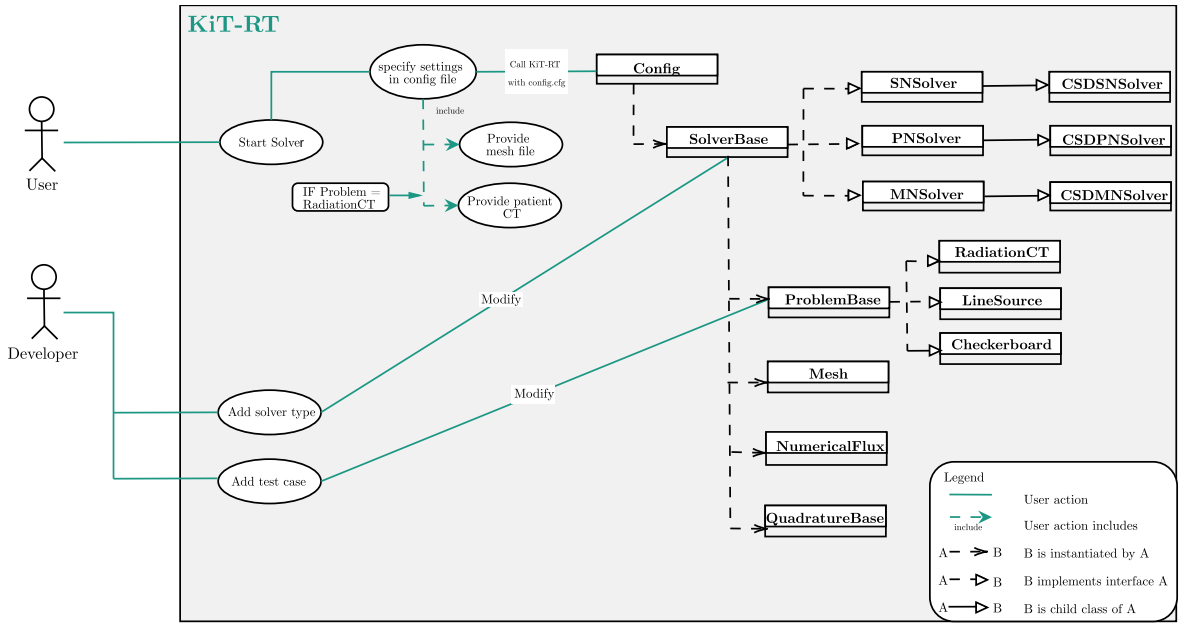


Figure 6.1: Schematic illustration of KiT-RT user interface and class structure for a subset of classes most relevant for dose computations.

virtual base class, which contains a static factory method to build an instance of the concrete derived class, defined by the given configuration details. Furthermore, the virtual base class defines the interface of this building block with other parts of the KiT-Framework.

Results generated with KiT-RT will be presented in the following section and compared to the other methods discussed in part II in chapter 8.

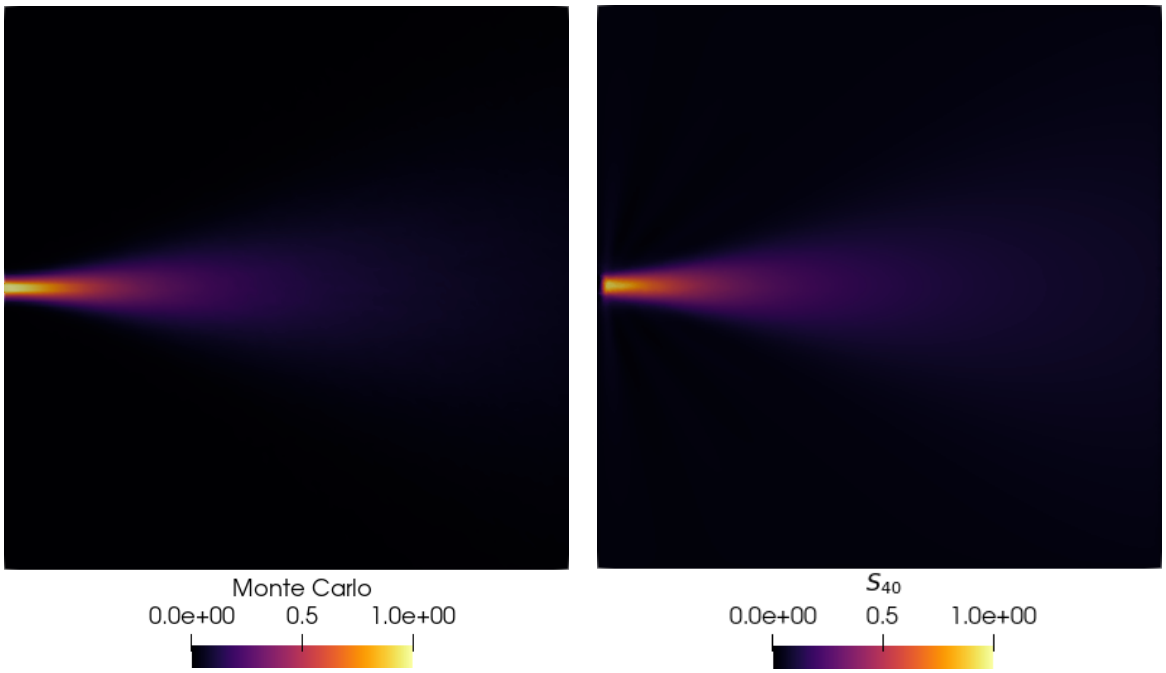
6.3 Results

The KiT-RT framework has been validated for standard line-source and checkerboard testcases using StarMAP [195] and the TOPAS Monte Carlo framework [175] in [112]. In order to validate our continuous slowing down solvers, we first consider a mono-energetic electron beam in a homogeneous 2D water box with the initial condition

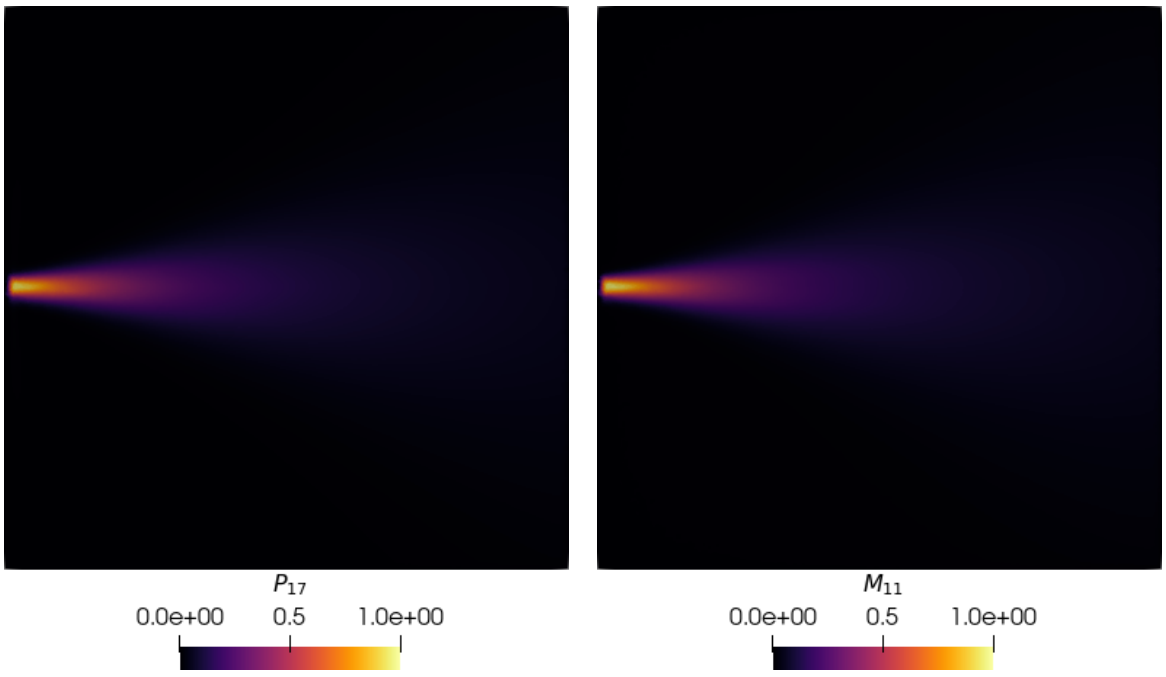
$$\psi(\mathbf{r}, \mathbf{\Omega}, E_{max}) = \frac{1}{(2\pi)^{3/2}\sigma_{\Omega_2}\sigma_{r_x}\sigma_{r_y}} \cdot \exp(-(\mu_{\Omega_2} - \Omega_2)^2/2\sigma_{\Omega_2}) \cdot \exp(-(\mu_{r_x} - r_x)^2/2\sigma_{r_x}) \cdot \exp(-(\mu_{r_y} - r_y)^2/2\sigma_{r_y}),$$

where $(\mu_{r_x}, \mu_{r_y}) = (0.5\text{cm}, 0\text{cm})$ is the beam position within the 1cm × 1cm domain and $\mu_{\Omega_2} = \frac{\pi}{2}\text{rad}$ is the beam direction. The remaining parameters are chosen as $\sigma_{r_x} = \sigma_{r_y} = 0.01$ and $\sigma_{\Omega_2} = 0.1$.

In fig. 6.2 the normalized dose computed with the S_{40}, P_{17} and M_{11} CSD solvers included in the KiT-RT framework is compared to a reference Monte Carlo solution. All solutions from the deterministic solvers are able to capture the beam structure including a higher dose and the correct beam width at the entrance and a gradual fanning out with increasing depth due to scattering.



(a) Monte Carlo (b) S_{40}

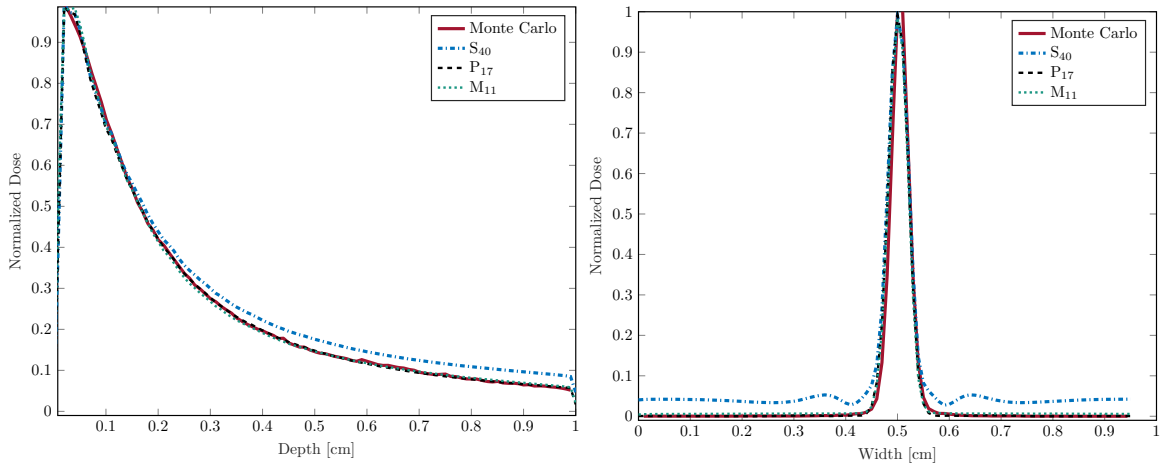


(c) P_{17} (d) M_{11}

Figure 6.2: Dose deposited by a 5 MeV electron beam in a 2D water box computed with (a) a Monte Carlo method (TOPAS MC) compared to KiT-RT implementations of (b) S_{40} (c) P_{17} and (d) M_{11} solvers.

Figure 6.3 further shows a lateral and longitudinal cut through the solutions. All three deterministic solvers agree well with Monte Carlo at the beam entrance. While P_{17} and M_{11} remain close to the Monte Carlo dose along the complete longitudinal cut, the dose computed with an S_{40} solver deviates slightly

at larger depths and does not fall off as steeply as the reference. In the lateral cut, all solvers exhibit the same beam width and peak height[1]. The S_N solution however oscillates slightly to the side of the beam and maintains a small positive dose value even in regions which should not be irradiated according to the other methods.



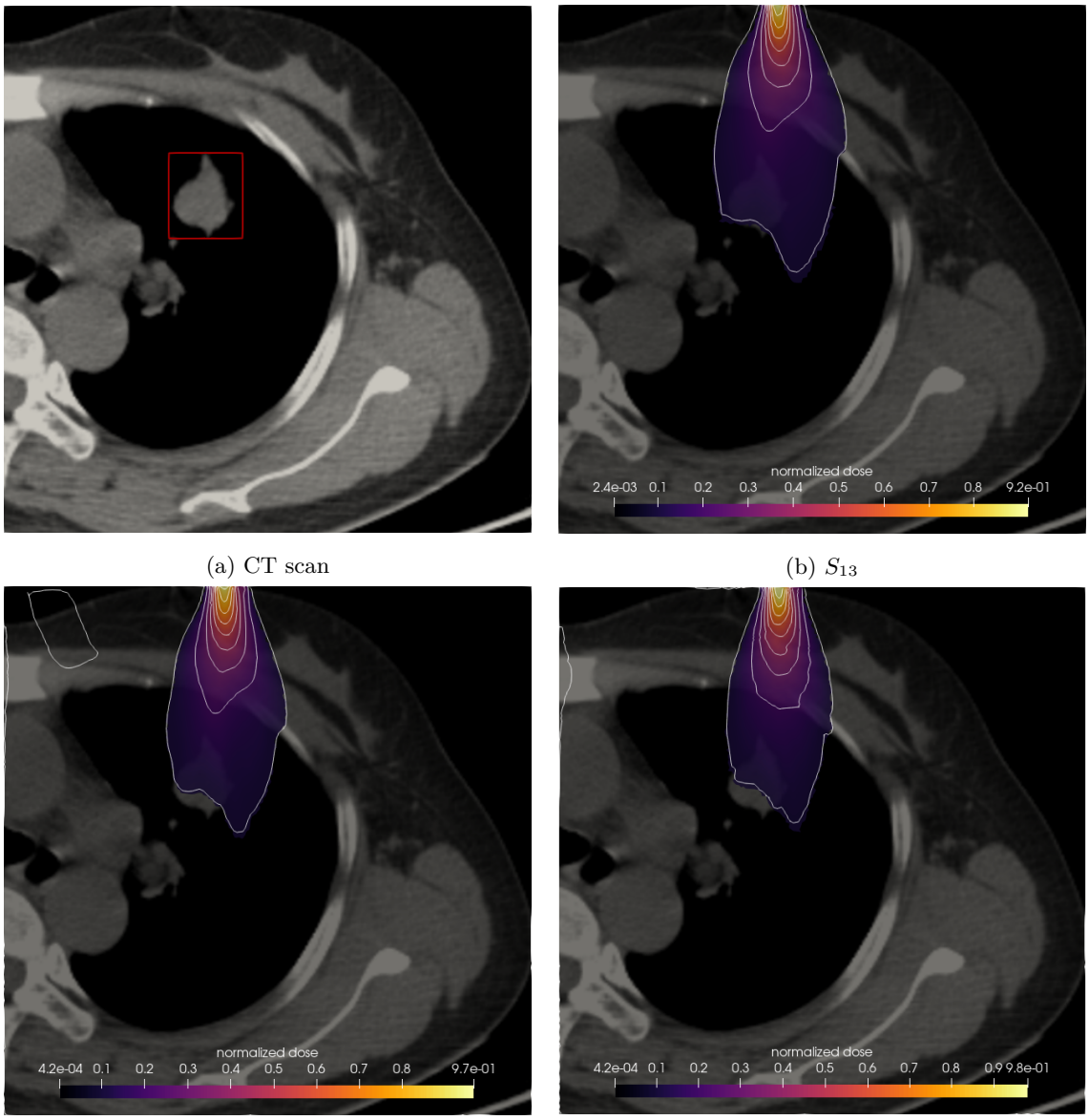
(a) (b)

Figure 6.3: Cuts through dose in 2D waterbox along (a) central beam axis ($x = 0.5$) and (b) at $y = 0.1$cm.

After validating the solvers using a simple test case, we now examine a realistic 2D CT scan of a lung patient as a proof of concept for the application of the framework to radiation therapy computations. The patient data was retrieved from an open source data set [127] in The Cancer Imaging Archive (TCIA) [39]. The patient is irradiated with an electron beam of $E_{\max} = 20$ MeV. We model this beam analogously to the previous test case, as a Gaussian, where $(\mu_{r_x}, \mu_{r_y}) = (2.5\text{cm}, 5.8\text{cm})$ is the beam position within the 6cm $\times$ 6cm domain and $\mu_{\Omega_2} = \frac{\pi}{2}$rad is the beam direction. The remaining parameters are chosen as $\sigma_{r_x} = \sigma_{r_y} = \sigma_{\Omega_2} = 0.1$. To determine a tissue density ρ for given gray-scale values of the CT image, we set the maximum density, represented by white pixels, to the density of bone $\rho_{\text{bone}} = 1.85$ g/cm^3. The remaining tissue is scaled such that the minimum pixel value of zero corresponds to a minimal density of $\rho_{\min} = 0.05$ g/cm^3. This corresponds approximately to the lower bound of observed lung densities [104].

Figure 6.4 compares the normalized dose for a CSD S_{13}, P_{13} and M_5 solver. While all methods show similar behaviour and are able to capture the effects of heterogeneities in the patient density, some differences, e.g., in the maximum depth of the S_{13} solution compared to P_{13} and M_5 or the shape of the lowest two isolines can be observed. The cross sections in figure 6.5 further show that the S_{13} dose has a lower maximum and higher minimum value than the M_5 and to a lesser extent also P_{13} solutions.

[1] Note that the difference at the peak between the methods stems partly from the different spatial grids used in KiT-RT vs. TOPAS MC. The rectangular grid cells in the MC code are aligned/perpendicular to the beam direction. This can lead to a more squared-off plateau in the MC solution, which is not reproduced by the irregular, triangular grid in KiT-RT.



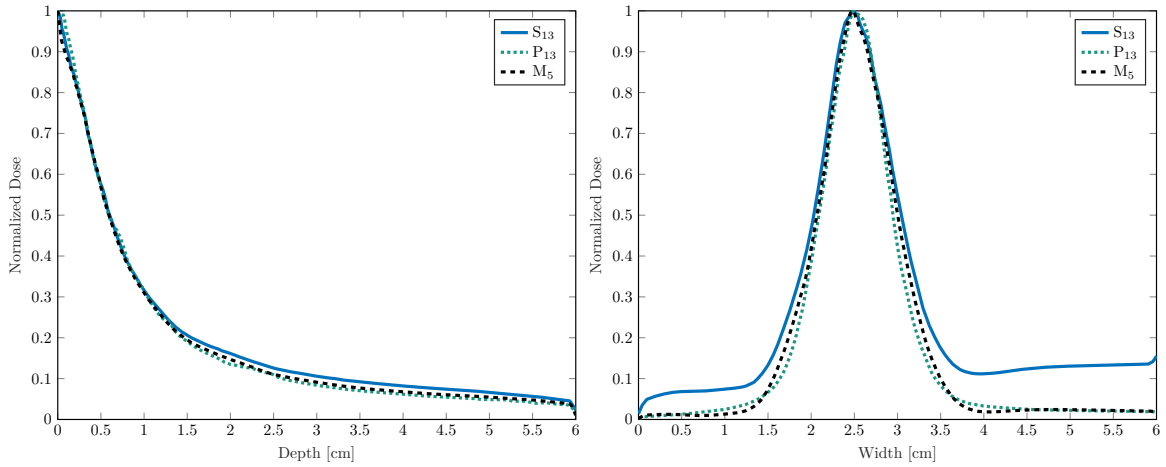
(a) CT scan (b) S_{13}

(c) P_{13} (d) M_5

Figure 6.4: Patient CT scan with lung tumor (red box) (a) as well as corresponding simulation results for the S_{13}, P_{13} and M_5 solver with spherical harmonics basis and partially regularized entropy.

6.4 Runtime and parallel scaling

In [112] the parallel performance of the three base solver implementations S_N, P_N and M_N is investigated based on a 1D line source benchmark testcase [62] with up to 16 workers on a fixed unstructured triangular mesh of size $n = 578290$. In the following the findings are briefly summarized : We observe almost linear scaling for up to 12 workers for all three methods. For higher amounts of workers, the S_N and to a smaller degree also P_N method lose efficiency, while the M_N method scales well also for 12-16 workers. While the performance of the continuous slowing down solver implementations has not been explicitly evaluated, it can be assumed to follow the corresponding base solver performance, since the same spatial,

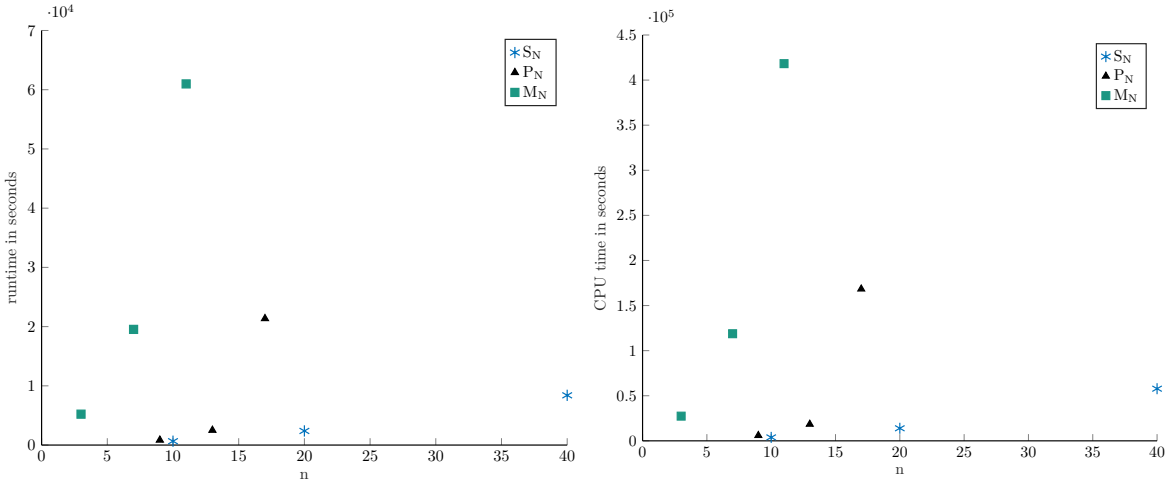


(a) (b)

Figure 6.5: (a) Vertical (at $x = 2.5$cm) and (b) horizontal (at $y = 5$cm) cross section through the normalized dose in the patient CT. Comparison of the S_{13}, P_{13} and partially regularized M_5 solver.

velocity and (pseudo) temporal discretizations are used.

A comparison of the run-times of the S_N, P_N and M_N solver for a fixed number of 8 workers is presented in fig. 6.6. The parallel implementation scales equally well for all presented methods, as can be seen by the similar structure of CPU time and absolute run time plots. It is apparent that the S_N method is the least computationally expensive, followed by the P_N and then the M_N method. The longer run-times of P_N and M_N are due to their slightly more complex set-up compared to the S_N method, which requires the computation of spherical harmonics basis functions and moments. The M_N method additionally solves an optimization problem to minimize the entropy, which further increases the computational costs.



(a) (b)

Figure 6.6: Runtimes (a) and CPU times (b) of the S_N, P_N and M_N method for CSD equations implemented in KiT-RT.

For increasing N, the S_N method is consistently the most computationally efficient of the compared methods in our experiments. However, the nodal nature of the angular discretization used in the S_N method can be problematic for strongly forward-peaked beams and the solution can exhibit ray-effects. In the previous section, we could already observe some artifacts and deviations in the S_N solution compared to the other methods and MC reference. Therefore, it is necessary to choose the best suited method on a case by case basis.

CHAPTER 7

Dynamical low-rank approximation for Boltzmann dose calculations

Deterministic Boltzmann solvers have the downside that they require fine numerical discretizations in order to achieve an accurate dose for directed beams in a heterogeneous medium. This negates their advantage over Monte Carlo methods and can not only lead to prohibitively long run times but also storage problems.

In the following chapter, we tackle the challenges arising from high-dimensional phase spaces in radiation therapy applications through dynamical low-rank approximation (DLRA) [102]. DLRA represents the solution by a low-rank ansatz. When the solution is an $n \times m$ time-dependent matrix with huge n and m, such a low-rank representation can be defined by a singular value decomposition (SVD) truncated to a (small) rank $\mathcal{R}$.[1] Time evolution equations for each of the low-rank factors of the SVD are then derived by minimizing the residual while maintaining the solution's low-rank structure. When the original method requires $O(n \cdot m)$ operations per time step, the updates of the low-rank factors only require $O(\mathcal{R}^2 \cdot (n+m))$ operations. Robust integrators for the time evolution of these factors are the matrix projector–splitting integrator (PSI) [137] and the recently developed "unconventional" basis-update & Galerkin low-rank matrix integrator of [36]. Unlike the unconventional integrator, the projector–splitting integrator can be extended to high order. However, since it includes a backward step in time, the projector–splitting integrator can yield instabilities for parabolic equations. Moreover, the backward time step can result in unstable schemes for hyperbolic problems, which is not an issue for the unconventional integrator [111].

The efficiency of DLRA highly depends on the rank required to capture important solution characteristics. Choosing this rank sufficiently high to guarantee a satisfactory solution quality while at best maintaining low computational costs requires a great amount of intuition. Furthermore, a fixed choice of the rank

[1] Note that the low-rank solution does not require a diagonal $\mathcal{R}$ by $\mathcal{R}$ singular value matrix. In fact, DLRA uses dense coefficient matrices.

does not capture the time evolution of the solution complexity. Rank adaptive DLRA integrators which pick the rank in an automated fashion during run time have for example been proposed in [47, 239, 35, 79, 87, 76]. In the following, an efficient choice of the rank for every energy is provided through the rank adaptive integrator of [35].

In addition to applying the DLRA to the CSD equation for radiation therapy, we introduce the following novelties:

1. *A stable and efficient time discretization of stiff scattering terms.* Following [162, 111], stiff scattering terms are split from the radiation transport equation. The unconventional integrator is used to time update the streaming part and the matrix projector-splitting integrator updates the scattering part. According to [111], the scattering part only requires an update of a single low-rank factor. By using an implicit time discretization on this part of the integrator and explicit updates on the remainder, we significantly reduce computational costs while allowing for a less restrictive CFL condition.

2. *A first-collision source method to reduce the rank and impose boundary conditions.* Boundary conditions in radiation therapy are often Dirichlet conditions of uncollided particles traveling into a single direction. To efficiently incorporate this information into our solution ansatz, we perform a collided-uncollided split, see e.g. [2, 72]. Here, collided and uncollided particles are treated in two separate equations. For the uncollided particles, an S_N method with a directed quadrature set resolving only the small number of relevant directions can be used. The collided part of the solution is represented through dynamical low-rank approximation. This strips away highly peaked particle distributions in the low-rank approximation, thereby potentially reducing the rank. Furthermore, since the density of uncollided particles is zero at the boundary, imposing boundary conditions becomes straightforward.

3. *A multilevel dynamical low-rank approximation.* The collided-uncollided split can be extended to L-collided splits. This can be interpreted as writing the solution as a telescoping sum. Expecting a reduced rank in each of these telescoping updates, individual dynamical low-rank approximations are derived for every term.

7.1 Mesoscopic transport models in radiation therapy

First, we recall the numerical discretization and splitting techniques used to determine the semi-discretized time/energy-dependent transport equations to which we can apply the dynamical low-rank approximation. Further, for the sake of conciseness, we introduce some simplified notation. The proposed method is based on the CSD equation (eq. (4.11)) with transformed energy, i.e., pseudo time t as well as transformed cross sections and particle density as defined in section 4.2. Omitting phase space dependencies, we now rearrange eq. (4.11) and call the right-hand side $R(t, \psi)$:

$$\partial_t \psi = -\mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\psi}{\rho} - \Sigma_t \psi + \int_{\mathbb{S}^2} \Sigma_s(t, \mathbf{\Omega} \cdot \mathbf{\Omega}') \psi(t, \mathbf{r}, \mathbf{\Omega}') \, \mathrm{d}\mathbf{\Omega}' := R(t, \psi) \, . \tag{7.1}$$

Furthermore, the in-scattering operator will be denoted as $\mathcal{Q}$, i.e.,

$$(\mathcal{Q}\psi)(t, \mathbf{r}, \mathbf{\Omega}) := \int_{\mathbb{S}^2} \Sigma_s(t, \mathbf{\Omega} \cdot \mathbf{\Omega}') \psi(t, \mathbf{r}, \mathbf{\Omega}') \, \mathrm{d}\mathbf{\Omega}' \, . \tag{7.2}$$

7.1.1 Collision source method

First collision source methods split the solution to the radiation transport equation eq. (7.1) into a collided and uncollided part, see e.g. [5, 72]. Let us write $\psi(t, \mathbf{r}, \mathbf{\Omega}) = \psi_u(t, \mathbf{r}, \mathbf{\Omega}) + \psi_c(t, \mathbf{r}, \mathbf{\Omega})$, where ψ_u represents uncollided and ψ_c represents collided particles. Then, the radiation transport equation (7.1) can be split into

$$\partial_t \psi_u = -\mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\psi_u}{\rho} - \Sigma_t \psi_u := R_u(t, \psi_u) \,, \tag{7.3a}$$

$$\partial_t \psi_c = -\mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\psi_c}{\rho} - \Sigma_t \psi_c + \mathcal{Q}\left(\psi_u + \psi_c\right) := R_c(t, \psi_u, \psi_c) \,. \tag{7.3b}$$

The first equation describes the dynamics of uncollided particles. Since collided particles cannot generate or deplete uncollided particles, eq. (7.3a) solely depends on ψ_u. Furthermore, as uncollided particles that scatter will no longer be uncollided, only the out-scattering term $-\Sigma_t \psi_u$ describes interactions with the background material. Uncollided particles that undergo a collision are added to the collided particles which are described by eq. (7.3b). Hence, the in-scattering term of uncollided particles $\mathcal{Q}(\psi_u)$ is treated as a source term in the collided equation. This methodology of representing the solution in terms of collided and uncollided particles can be developed further: Denoting the particles that have collided $\ell = 0, \cdots, L$ times as ψ_ℓ and particles that have collided more than L times as ψ_c, we have $\psi = \psi_0 + \psi_1 + \cdots + \psi_L + \psi_c$. Then, we obtain the equations

$$\partial_t \psi_0 = -\mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\psi_0}{\rho} - \Sigma_t \psi_0 := R_0(t, \psi_0) \,, \tag{7.4}$$

$$\partial_t \psi_\ell = -\mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\psi_\ell}{\rho} - \Sigma_t \psi_\ell + \mathcal{Q}\psi_{\ell-1} := R_\ell(t, \psi_{\ell-1}, \psi_\ell) \,, \qquad \text{for } \ell = 1, \cdots, L, \tag{7.5}$$

$$\partial_t \psi_c = -\mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \frac{\psi_c}{\rho} - \Sigma_t \psi_c + \mathcal{Q}\left(\psi_L + \psi_c\right) := R_c(t, \psi_L, \psi_c) \,. \tag{7.6}$$

Boundary conditions for the above equations can be imposed by noting that all particles entering from the boundary are uncollided. I.e., we have $\psi_0(E, \mathbf{r}, \mathbf{\Omega}) = \psi_{\text{BC}}(E, \mathbf{r}, \mathbf{\Omega})$ for $\mathbf{r} \in \partial D$ and $\mathbf{n} \cdot \mathbf{\Omega} < 0$. Since no collided particles will enter the spatial domain from the boundary, we have $\psi_\ell = 0$ and $\psi_c = 0$ at the boundary for $\mathbf{n} \cdot \mathbf{\Omega} < 0$.

7.1.2 Spatial and angular discretization

The dynamical low-rank approximation can be applied to time-dependent matrix differential equations. Thus, we discretize the phase space only with respect to space and angle. For the angular discretization, we use a P_N approach as introduced in section 5.3.1. The continuous slowing down approximation discretized only in angle according to the P_N method is given by eq. (5.12). The spatial variable is then discretized using a finite volume approach similar to section 5.3.3. In contrast to chapter 6 we now use a rectangular, structured grid. We further introduce the method for a two-dimensional spatial domain, i.e., the spatial variable becomes $\mathbf{r} = (x, y)^T$, an extension to three dimensions is straightforward. Then, we have

$$\partial_t \mathbf{u}(t, \mathbf{r}) = -\mathbf{A}_x \partial_x \frac{\mathbf{u}(t, \mathbf{r})}{\rho(\mathbf{r})} - \mathbf{A}_y \partial_y \frac{\mathbf{u}(t, \mathbf{r})}{\rho(\mathbf{r})} + \mathbf{G}\mathbf{u}(t, \mathbf{r}) \,. \tag{7.7}$$

The finite volume discretization splits the spatial domain D into $N_x \cdot N_y$ cells. In case of a structured quadrangle grid, the x and y domains are discretized into uniform one-dimensional grids $x_1 \leq x_2 \leq \cdots \leq$

x_{N_x+1} and $y_1 \leq y_2 \leq \cdots \leq y_{N_y+1}$ with grid size Δx and Δy respectively. Then, the cell of index (i,j) is defined on $I_{ij} := [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ on which the numerical solution is chosen as

$$\mathbf{u}_{ij}(t) \simeq \frac{1}{\Delta x \Delta y} \int_{I_{ij}} \mathbf{u}(t, \mathbf{r}) \, \mathrm{d}\mathbf{r} \, .$$

In the same manner, we compute the patient density ρ_{ij}. To simplify the presentation of the spatial discretization, let us collect $\mathbf{u}_{ij}(t) = (u_{ijk}(t))_{k=1}^m \in \mathbb{R}^m$ for $i = 1, \cdots, N_x$ and $j = 1, \cdots, N_y$ into a matrix $\mathbf{u}(t) \in \mathbb{R}^{n_x \times m}$, where $n_x := N_x \cdot N_y$. For this, we define the function $\mathrm{idx} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ as $\mathrm{idx}(i,j) = (i-1) \cdot N_x + j$. Then, the entries of $\mathbf{u}(t)$ are defined as $u_{\mathrm{idx}(i,j),k}(t) = u_{ijk}(t)$.

With this notation at hand, we can define a finite volume scheme for eq. (7.7) in compact notation. Let us define the sparse diffusion stencil matrices $\mathbf{L}_{x,y}^{(1)} \in \mathbb{R}^{n_x \times n_x}$ as

$$\begin{aligned} L^{(1)}_{x,\mathrm{idx}(i,j),\mathrm{idx}(i,j)} &= \frac{1}{\rho_{ij}\Delta x} \, , \quad L^{(1)}_{x,\mathrm{idx}(i,j),\mathrm{idx}(i\pm1,j)} = -\frac{1}{2\rho_{i\pm1,j}\Delta x} \, , \\ L^{(1)}_{y,\mathrm{idx}(i,j),\mathrm{idx}(i,j)} &= \frac{1}{\rho_{ij}\Delta y} \, , \quad L^{(1)}_{y,\mathrm{idx}(i,j),\mathrm{idx}(i,j\pm1)} = -\frac{1}{2\rho_{i,j\pm1}\Delta y} \, , \end{aligned}$$

as well as the sparse advection stencil matrices $\mathbf{L}_{x,y}^{(2)} \in \mathbb{R}^{n_x \times n_x}$ as

$$L^{(2)}_{x,\mathrm{idx}(i,j),\mathrm{idx}(i\pm1,j)} = \frac{\pm 1}{2\rho_{i\pm1,j}\Delta x} \, , \quad L^{(2)}_{y,\mathrm{idx}(i,j),\mathrm{idx}(i,j\pm1)} = \frac{\pm 1}{2\rho_{i,j\pm1}\Delta y} \, .$$

Furthermore with $\mathbf{A}_{x,y} = \mathbf{V}_{x,y}\mathbf{\Lambda}_{x,y}\mathbf{V}_{x,y}^T$, the Roe matrices $|\mathbf{A}_x|$, $|\mathbf{A}_y| \in \mathbb{R}^{m \times m}$ are defined as

$$|\mathbf{A}_x| := \mathbf{V}_x |\mathbf{\Lambda}_x| \mathbf{V}_x^T \, , \text{ and } \quad |\mathbf{A}_y| := \mathbf{V}_y |\mathbf{\Lambda}_y| \mathbf{V}_y^T \, .$$

Here, $\mathbf{V}_{x,y}$ collects the orthonormal eigenvectors of the symmetric matrices $\mathbf{A}_{x,y}$ and $\mathbf{\Lambda}_{x,y} = \mathrm{diag}(\lambda_1^{x,y}, \cdots \lambda_m^{x,y})$ are the corresponding real eigenvalues. Then, the semi-discrete finite volume update becomes a huge matrix differential equation of the form

$$\dot{\mathbf{u}}(t) = \mathbf{F}(\mathbf{u}(t)) + \mathbf{G}(t, \mathbf{u}(t), \mathbf{u}(t)) \, ,$$

where

$$\begin{aligned} \mathbf{F}(\mathbf{u}) &:= \mathbf{L}_x^{(2)} \mathbf{u} \mathbf{A}_x^T + \mathbf{L}_y^{(2)} \mathbf{u} \mathbf{A}_y^T + \mathbf{L}_x^{(1)} \mathbf{u} |\mathbf{A}_x|^T + \mathbf{L}_y^{(1)} \mathbf{u} |\mathbf{A}_y|^T, \\ \mathbf{G}(t, \mathbf{v}, \mathbf{u}) &:= -\Sigma_t(t)\mathbf{u} + \mathbf{v}\mathbf{\Sigma}(t) \, . \end{aligned}$$

The costs of evaluating the right-hand side are $C_{\mathrm{P}_N} \lesssim n_x \cdot m$, when accounting for the sparsity of all stencil, flux and Roe matrices which leads to linear costs in n_x and m of matrix products.

7.2 Main framework

This section gives a brief overview of dynamical low-rank approximation [102] for matrix differential equations $\dot{\mathbf{u}}(t) = \mathbf{F}(t, \mathbf{u}(t))$. Dynamical low-rank approximation represents and evolves the solution on a manifold of rank $\mathcal{R}$ matrices, which we denote by $\mathcal{M}_\mathcal{R}$. A low-rank representation is given by the SVD-like factorization

$$\mathbf{u}(t) \approx \mathbf{X}(t)\mathbf{S}(t)\mathbf{W}(t)^T, \tag{7.8}$$

where $\mathbf{X} \in \mathbb{R}^{n_x \times \mathcal{R}}$ and $\mathbf{W} \in \mathbb{R}^{m \times \mathcal{R}}$ are basis matrices with orthonormal column vectors and $\mathbf{S} \in \mathbb{R}^{\mathcal{R} \times \mathcal{R}}$ is a dense coefficient matrix. Time evolution equations for the factors can be defined by imposing

$$\dot{\mathbf{u}}(t) \in T_{\mathbf{u}(t)}\mathcal{M}_{\mathcal{R}} \qquad \text{such that} \qquad \|\dot{\mathbf{u}}(t) - \mathbf{F}(t, \mathbf{u}(t))\| = \min . \tag{7.9}$$

We use $T_{\mathbf{u}(t)}\mathcal{M}_{\mathcal{R}}$ to denote the tangent space of $\mathcal{M}_{\mathcal{R}}$ at $\mathbf{u}(t)$, i.e., the solution should remain of rank $\mathcal{R}$ over time while minimizing the defect. These conditions yield a time evolution equation for the low-rank solution [102, Lemma 4.1], which reads

$$\dot{\mathbf{u}}(t) = \mathbf{P}(\mathbf{u}(t))\mathbf{F}(t, \mathbf{u}(t)) . \tag{7.10}$$

The operator $\mathbf{P}$ is the orthogonal projection onto the tangent space, given by

$$\mathbf{P}\mathbf{g} = \mathbf{X}\mathbf{X}^T\mathbf{g} - \mathbf{X}\mathbf{X}^T\mathbf{g}\mathbf{W}\mathbf{W}^T + \mathbf{g}\mathbf{W}\mathbf{W}^T .$$

Evolution equations can then be derived for the factors $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{W}$, see [102]. However, the resulting equations depend on the inverse of the commonly ill-conditioned coefficient matrix, which substantially limits the permitted step size [99].

7.3 Robust fixed rank integrators

Two robust integrators have been proposed for the evolution equation (7.10). First, the matrix projector–splitting integrator [137] splits eq. (7.10) by a Lie-Trotter splitting technique, yielding

$$\dot{\mathbf{u}}_I(t) = \mathbf{F}(\mathbf{u}_I(t))\mathbf{W}\mathbf{W}^T , \qquad \mathbf{u}_I(t_0) = \mathbf{u}(t_0) , \tag{7.11a}$$

$$\dot{\mathbf{u}}_{II}(t) = -\mathbf{X}\mathbf{X}^T\mathbf{F}(\mathbf{u}_{II}(t))\mathbf{W}\mathbf{W}^T , \qquad \mathbf{u}_{II}(t_0) = \mathbf{u}_I(t_1) , \tag{7.11b}$$

$$\dot{\mathbf{u}}_{III}(t) = \mathbf{X}\mathbf{X}^T\mathbf{F}(\mathbf{u}_{III}(t)) , \qquad \mathbf{u}_{III}(t_0) = \mathbf{u}_{II}(t_1) . \tag{7.11c}$$

The resulting consecutive movement in the low-rank manifold ensures robustness irrespective of singular values and thereby allows for increased step sizes [137]. Defining the decompositions $\mathbf{u}_I = \mathbf{K}\mathbf{W}^T$ as well as $\mathbf{u}_{III} = \mathbf{X}\mathbf{L}$ gives the *matrix projector-splitting integrator*, which updates the low-rank factors $\mathbf{X}^0 = \mathbf{X}(t_0)$, $\mathbf{W}^0 = \mathbf{W}(t_0)$ and $\mathbf{S}^0 = \mathbf{S}(t_0)$ to time $t_1 = t_0 + \Delta t$:

1. ***K*-step**: Update $\mathbf{X}^0$ to $\mathbf{X}^1$ and $\mathbf{S}^0$ to $\widetilde{\mathbf{S}}^0$ via

   $$\dot{\mathbf{K}}(t) = \mathbf{F}(\mathbf{K}(t)\mathbf{W}^{0,T})\mathbf{W}^0 , \qquad \mathbf{K}(t_0) = \mathbf{X}^0\mathbf{S}^0 . \tag{7.12}$$

   Determine $\mathbf{X}^1$ and $\widetilde{\mathbf{S}}^0$ with $\mathbf{K}(t_1) = \mathbf{X}^1\widetilde{\mathbf{S}}^0$ by performing a QR decomposition.

2. ***S*-step**: Update $\widetilde{\mathbf{S}}^{\mathbf{0}}$ to $\widetilde{\mathbf{S}}^{\mathbf{1}}$ via

   $$\dot{\widetilde{\mathbf{S}}}(t) = -\mathbf{X}^{1,T}\mathbf{F}(\mathbf{X}^1\widetilde{\mathbf{S}}(t)\mathbf{W}^{0,T})\mathbf{W}^0 , \qquad \widetilde{\mathbf{S}}(t_0) = \widetilde{\mathbf{S}}^0 \tag{7.13}$$

   and set $\widetilde{\mathbf{S}}^1 = \widetilde{\mathbf{S}}(t_1)$.

3. ***L*-step**: Update $\mathbf{W}^0$ to $\mathbf{W}^1$ and $\widetilde{\mathbf{S}}^1$ to $\mathbf{S}^1$ via

   $$\dot{\mathbf{L}}(t) = \mathbf{X}^{1,T}\mathbf{F}(\mathbf{X}^1\mathbf{L}(t)) , \qquad \mathbf{L}(t_0) = \widetilde{\mathbf{S}}^1\mathbf{W}^{0,T} . \tag{7.14}$$

   Determine $\mathbf{W}^1$ and $\mathbf{S}^1$ with $\mathbf{L}(t_1) = \mathbf{S}^1\mathbf{W}^{1,T}$ by performing a QR decomposition.

Then, the time updated solution is $\mathbf{u}(t_1) = \mathbf{X}^1\mathbf{S}^1\mathbf{W}^{1,T}$. It has been noted in [111] that when the flux function takes the form $\mathbf{F}(t,\mathbf{u}(t)) = \mathbf{u}(t)\mathbf{G}(t)$, the K and S-steps cancel out and only the L-step determines the dynamics.

The second robust integrator is the *unconventional integrator*, which has recently been introduced in [36]. This integrator first performs basis updates of $\mathbf{X}$ and $\mathbf{W}$ in parallel and then updates the coefficient matrix $\mathbf{S}$ by a Galerkin step. It shares the robustness properties of the matrix projector–splitting integrator [36] and takes the form

1. K**-step**: Update $\mathbf{X}^0$ to $\mathbf{X}^1$ via

$$\dot{\mathbf{K}}(t) = \mathbf{F}(\mathbf{K}(t)\mathbf{W}^{0,T})\mathbf{W}^0\,, \qquad \mathbf{K}(t_0) = \mathbf{X}^0\mathbf{S}^0\,. \tag{7.15}$$

   Determine $\mathbf{X}^1$ with $\mathbf{K}(t_1) = \mathbf{X}^1\mathbf{R}$ and store $\mathbf{M} = \mathbf{X}^{1,T}\mathbf{X}^0$.

2. L**-step**: Update $\mathbf{W}^0$ to $\mathbf{W}^1$ via

$$\dot{\mathbf{L}}(t) = \mathbf{X}^{0,T}\mathbf{F}(\mathbf{X}^0\mathbf{L}(t))\,, \qquad \mathbf{L}(t_0) = \mathbf{S}^0\mathbf{W}^{0,T}\,. \tag{7.16}$$

   Determine $\mathbf{W}^1$ with $\mathbf{L}(t_1) = \mathbf{W}^1\widetilde{\mathbf{R}}$ and store $\mathbf{N} = \mathbf{W}^{1,T}\mathbf{W}^0$.

3. S**-step**: Update $\mathbf{S}^0$ to $\mathbf{S}^1$ via

$$\dot{\mathbf{S}}(t) = \mathbf{X}^{1,T}\mathbf{F}(\mathbf{X}^1\mathbf{S}(t)\mathbf{W}^{1,T})\mathbf{W}^1\,, \qquad \mathbf{S}(t_0) = \mathbf{M}\mathbf{S}^0\mathbf{N}^T \tag{7.17}$$

   and set $\mathbf{S}^1 = \mathbf{S}(t_1)$.

Unlike the projector–splitting integrator, the unconventional integrator only propagates the solution forward in time. Thereby, it inherits the structure of the full method, which can for example be used to ensure stability for hyperbolic and kinetic problems [111]. Therefore, the unconventional integrator yields the same semi-discrete evolution equations, irrespective of whether the problem is discretized first and then DLRA is applied to the resulting matrix ODE, or the DLRA projections are first performed on the continuous level and the resulting continuous evolution equations are then discretized as proposed in [54]. This equivalence holds when the same spatial and angular discretization is used in both approaches. A disadvantage of the unconventional integrator is that it cannot be interpreted as a classical splitting method which is why it currently is only first order accurate in time, if the exactness property [36, Theorem 1] does not hold.

7.4 Rank adaptive unconventional integrator

The unconventional integrator has recently been extended to allow for rank adaptivity [35]. That is, given a tolerance parameter ϑ, the integrator adapts the rank in time.

Starting from time t_0 where the solution has rank $\mathcal{R}_0$, the integrator gives the factored solution at time t_1 with rank $\mathcal{R}_1 \leq 2\mathcal{R}_0$. In the following, we use $\mathcal{R} = \mathcal{R}_0$ and use hats to denote matrices of rank $2\mathcal{R}$. Then the rank adaptive integrator reads

1. K**-step**: Update $\mathbf{X}^0 \in \mathbb{R}^{n_x\times\mathcal{R}}$ to $\widehat{\mathbf{X}}^1 \in \mathbb{R}^{n_x\times 2\mathcal{R}}$ via

$$\dot{\mathbf{K}}(t) = \mathbf{F}(\mathbf{K}(t)\mathbf{W}^{0,T})\mathbf{W}^0\,, \qquad \mathbf{K}(t_0) = \mathbf{X}^0\mathbf{S}^0\,. \tag{7.18}$$

Determine $\widehat{\mathbf{X}}^1$ with $[\mathbf{K}(t_1), \mathbf{X}^0] = \widehat{\mathbf{X}}^1\mathbf{R}$ and store $\widehat{\mathbf{M}} = \widehat{\mathbf{X}}^{1,T}\mathbf{X}^0 \in \mathbb{R}^{2\mathcal{R}\times\mathcal{R}}$.

2. ***L*-step**: Update $\mathbf{W}^0 \in \mathbb{R}^{m\times\mathcal{R}}$ to $\widehat{\mathbf{W}}^1 \in \mathbb{R}^{m\times 2\mathcal{R}}$ via

$$\dot{\mathbf{L}}(t) = \mathbf{X}^{0,T}\mathbf{F}(\mathbf{X}^0\mathbf{L}(t)) , \qquad \mathbf{L}(t_0) = \mathbf{S}^0\mathbf{W}^{0,T} . \tag{7.19}$$

Determine $\widehat{\mathbf{W}}^1$ with $[\mathbf{L}(t_1), \mathbf{W}^0] = \widehat{\mathbf{W}}^1\widetilde{\mathbf{R}}$ and store $\widehat{\mathbf{N}} = \widehat{\mathbf{W}}^{1,T}\mathbf{W}^0$.

3. ***S*-step**: Update $\mathbf{S}^0 \in \mathbb{R}^{\mathcal{R}\times\mathcal{R}}$ to $\widehat{\mathbf{S}}^1 \in \mathbb{R}^{2\mathcal{R}\times 2\mathcal{R}}$ via

$$\dot{\widehat{\mathbf{S}}}(t) = \widehat{\mathbf{X}}^{1,T}\mathbf{F}(\widehat{\mathbf{X}}^1\widehat{\mathbf{S}}(t)\widehat{\mathbf{W}}^{1,T})\widehat{\mathbf{W}}^1 , \qquad \widehat{\mathbf{S}}(t_0) = \widehat{\mathbf{M}}\mathbf{S}^0\widehat{\mathbf{N}}^T \tag{7.20}$$

and set $\widehat{\mathbf{S}}^1 = \widehat{\mathbf{S}}(t_1)$.

4. **Truncation**: Determine the SVD $\widehat{\boldsymbol{S}}^1 = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{Q}}^\top$ where $\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}(\sigma_j)$. For a given tolerance ϑ, choose the new rank $\mathcal{R}_1 \leq 2\mathcal{R}$ such that

$$\left(\sum_{j=\mathcal{R}_1+1}^{2\mathcal{R}} \sigma_j^2\right)^{1/2} \leq \vartheta .$$

Compute the new factors as follows: Let $\boldsymbol{S}^1$ be the $\mathcal{R}_1 \times \mathcal{R}_1$ diagonal matrix with the $\mathcal{R}_1$ largest singular values and let $\boldsymbol{P}_1 \in \mathbb{R}^{2\mathcal{R}\times\mathcal{R}_1}$ and $\boldsymbol{Q}_1 \in \mathbb{R}^{2\mathcal{R}\times\mathcal{R}_1}$ contain the first $\mathcal{R}_1$ columns of $\widehat{\boldsymbol{P}}$ and $\widehat{\boldsymbol{Q}}$, respectively. Finally, set $\boldsymbol{X}^1 = \widehat{\boldsymbol{X}}^1\boldsymbol{P}_1 \in \mathbb{R}^{m\times\mathcal{R}_1}$ and $\boldsymbol{W}^1 = \widehat{\boldsymbol{W}}^1\boldsymbol{Q}_1 \in \mathbb{R}^{n\times\mathcal{R}_1}$.

7.5 A robust collision source method for dynamical low-rank approximation

In this section, we present the main method, which aims at providing an efficient and robust alternative to conventional strategies. Key ingredients for the construction are 1) a collision source method to define a splitting of the original equation, 2) a further splitting of collision terms which are treated implicitly, 3) computing individual DLRA updates by using the unconventional integrator for collided particles.

7.5.1 Collided-uncollided split

Let us define a discretization in space and angle for the system of the ℓ-collided split eq. (7.4). We start by deriving moment equations for the collided particles in the collision source method and use an S_N method for uncollided particles. Without going into detail, we denote the S_N solution of the uncollided particles as $\boldsymbol{\psi}$ and the right-hand side of a time continuous S_N method for streaming (i.e., particles move without interacting with tissue) as $\mathbf{F}_S(\boldsymbol{\psi})$. Then the evolution equations for the semi-discrete solution become

$$\dot{\boldsymbol{\psi}}(t) = \mathbf{F}_S(\boldsymbol{\psi}(t)) - \Sigma_t(t)\boldsymbol{\psi}(t) , \tag{7.21a}$$

$$\dot{\mathbf{u}}_1(t) = \mathbf{F}(\mathbf{u}_1(t)) - \Sigma_t(t)\mathbf{u}_1(t) + \boldsymbol{\psi}(t)\mathbf{T}_M\boldsymbol{\Sigma}(t) , \tag{7.21b}$$

$$\dot{\mathbf{u}}_\ell(t) = \mathbf{F}(\mathbf{u}_\ell(t)) - \Sigma_t(t)\mathbf{u}_\ell(t) + \mathbf{u}_{\ell-1}(t)\boldsymbol{\Sigma}(t) , \qquad \text{for } \ell = 2, \cdots, L, \tag{7.21c}$$

$$\dot{\mathbf{u}}_c(t) = \mathbf{F}(\mathbf{u}_c(t)) - \Sigma_t(t)\mathbf{u}_c + (\mathbf{u}_L(t) + \mathbf{u}_c(t))\,\boldsymbol{\Sigma}(t) , \tag{7.21d}$$

where the matrix $\mathbf{T}_M$ maps the nodal solution onto its moments. These $L+2$ equations can be solved consecutively. Since radiation therapy commonly investigates the effects of particle beams that enter the computational domain from the boundary, the S_N equations from the uncollided particles can be solved efficiently by using a biased quadrature rule. I.e., the quadrature only encodes the small number of possible flight directions. After the first collision, particles move into all directions. To account for the increased complexity, we describe the collided solution through dynamical low-rank approximation. Following [111], we split streaming and scattering parts and use the matrix projector–splitting integrator to update in-scattering in equation (7.21d) as well as the unconventional integrator for the remainder.

To simplify our presentation, we start by discussing this strategy for eq. (7.21b) and then extend it to the remaining equations. Let us first split *streaming* and *scattering* in eq. (7.21b). Omitting the subscript 1 gives

$$\dot{\mathbf{u}}_I(t) = \mathbf{F}(\mathbf{u}_I(t)) \,, \qquad \mathbf{u}_I(t_0) = \boldsymbol{u}_1(t_0) \,, \tag{7.22a}$$

$$\dot{\mathbf{u}}_{II}(t) = -\Sigma_t(t)\mathbf{u}_{II}(t) + \boldsymbol{\psi}(t_1)\mathbf{T}_M\boldsymbol{\Sigma}(t) \,, \qquad \mathbf{u}_{II}(t_0) = \mathbf{u}_I(t_1) \,. \tag{7.22b}$$

The updated solution at time $t_1 = t_0 + \Delta t$ is then given as $\boldsymbol{u}_1(t_1) = \mathbf{u}_{II}(t_1)$. Note that the splitting method introduces an error of $O(\Delta t)$, which can be reduced by high-order splitting methods. We start with the derivation of the basis update and Galerkin step equations of the unconventional integrator for the streaming part (eq. (7.22a)). That is, we derive evolution equations for $\mathbf{X}_I(t), \mathbf{S}_I(t), \mathbf{W}_I(t)$ such that $\mathbf{u}_I(t) \approx \mathbf{X}_I(t)\mathbf{S}_I(t)\mathbf{W}_I(t)^T$. To simplify notation let us omit Roman indices in the following. The K-step equations (7.15) read

$$\begin{aligned}
\dot{\mathbf{K}}(t) &= \mathbf{F}(\mathbf{K}(t)\mathbf{W}^{0,T})\mathbf{W}^0 \\
&= \mathbf{L}_x^{(2)}\mathbf{K}(t)\mathbf{W}^{0,T}\mathbf{A}_x^T\mathbf{W}^0 + \mathbf{L}_y^{(2)}\mathbf{K}(t)\mathbf{W}^{0,T}\mathbf{A}_y^T\mathbf{W}^0 \\
&\quad + \mathbf{L}_x^{(1)}\mathbf{K}(t)\mathbf{W}^{0,T}|\mathbf{A}_x|^T\mathbf{W}^0 + \mathbf{L}_y^{(1)}\mathbf{K}(t)\mathbf{W}^{0,T}|\mathbf{A}_y|^T\mathbf{W}^0 \\
&= \mathbf{L}_x^{(2)}\mathbf{K}(t)\widehat{\mathbf{A}}_x^0 + \mathbf{L}_y^{(2)}\mathbf{K}(t)\widehat{\mathbf{A}}_y^0 + \mathbf{L}_x^{(1)}\mathbf{K}(t)|\widehat{\mathbf{A}}_x|^0 + \mathbf{L}_y^{(1)}\mathbf{K}(t)|\widehat{\mathbf{A}}_y|^0 \,,
\end{aligned} \tag{7.23}$$

where we use $\widehat{\mathbf{A}}_{x,y}^0 := \mathbf{W}^{0,T}\mathbf{A}_{x,y}^T\mathbf{W}^0$ and $|\widehat{\mathbf{A}}_{x,y}|^0 := \mathbf{W}^{0,T}|\mathbf{A}_{x,y}|^T\mathbf{W}^0$. The numerical costs to compute these matrices are of $O(\mathcal{R}^2 \cdot m)$ and evaluating the right-hand side of the K-step equations has costs of $O(\mathcal{R}^2 \cdot n_x)$. To point out that the spatial basis is not yet updated by the scattering step (eq. (7.22b)), we define the superscript $^1/_2$, i.e., the solution is given as $\mathbf{X}^{1/2} := \mathbf{X}_I(t_1)$.

The L-step equations (7.16) read

$$\begin{aligned}
\dot{\mathbf{L}}(t) &= \mathbf{X}^{0,T}\mathbf{F}(\mathbf{X}^0\mathbf{L}(t)) \\
&= \mathbf{X}^{0,T}\mathbf{L}_x^{(2)}\mathbf{X}^0\mathbf{L}(t)\mathbf{A}_x^T + \mathbf{X}^{0,T}\mathbf{L}_y^{(2)}\mathbf{X}^0\mathbf{L}(t)\mathbf{A}_y^T \\
&\quad + \mathbf{X}^{0,T}\mathbf{L}_x^{(1)}\mathbf{X}^0\mathbf{L}(t)|\mathbf{A}_x|^T + \mathbf{X}^{0,T}\mathbf{L}_y^{(1)}\mathbf{X}^0\mathbf{L}(t)|\mathbf{A}_y|^T \\
&= \widehat{\mathbf{L}}_x^{(2),0}\mathbf{L}(t)\mathbf{A}_x^T + \widehat{\mathbf{L}}_y^{(2),0}\mathbf{L}(t)\mathbf{A}_y^T + \widehat{\mathbf{L}}_x^{(1),0}\mathbf{L}(t)|\mathbf{A}_x|^T + \widehat{\mathbf{L}}_y^{(1),0}\mathbf{L}(t)|\mathbf{A}_y|^T \,,
\end{aligned} \tag{7.24}$$

where we use $\widehat{\mathbf{L}}_{x,y}^{(2),0} := \mathbf{X}^{0,T}\mathbf{L}_{x,y}^{(2)}\mathbf{X}^0$ and $\widehat{\mathbf{L}}_{x,y}^{(1),0} := \mathbf{X}^{0,T}\mathbf{L}_{x,y}^{(1)}\mathbf{X}^0$. The numerical costs to compute these matrices are of $O(\mathcal{R}^2 \cdot n_x)$ and evaluating the right-hand side of the L-step equations has costs of $O(\mathcal{R}^2 \cdot m)$. To point out that the directional basis is not yet updated by the scattering step (eq. (7.22b)), we define $\mathbf{W}^{1/2} := \mathbf{W}_I(t_1)$.

The S-step equations (7.17) read

$$\begin{aligned}\dot{\mathbf{S}}(t) &= \mathbf{X}^{1/2,T}\mathbf{F}(\mathbf{X}^{1/2}\mathbf{S}(t)\mathbf{W}^{1/2,T})\mathbf{W}^{1/2} \\ &= \mathbf{X}^{1/2,T}\mathbf{L}_x^{(2)}\mathbf{X}^{1/2}\mathbf{S}(t)\mathbf{W}^{1/2,T}\mathbf{A}_x^T\mathbf{W}^{1/2} + \mathbf{X}^{1/2,T}\mathbf{L}_y^{(2)}\mathbf{X}^{1/2}\mathbf{S}(t)\mathbf{W}^{1/2,T}\mathbf{A}_y^T\mathbf{W}^{1/2} \\ &\quad + \mathbf{X}^{1/2,T}\mathbf{L}_x^{(1)}\mathbf{X}^{1/2}\mathbf{S}(t)\mathbf{W}^{1/2,T}|\mathbf{A}_x|^T\mathbf{W}^{1/2} + \mathbf{X}^{1/2,T}\mathbf{L}_y^{(1)}\mathbf{X}^{1/2}\mathbf{S}(t)\mathbf{W}^{1/2,T}|\mathbf{A}_y|^T\mathbf{W}^{1/2} \\ &= \widehat{\mathbf{L}}_x^{(2),1/2}\mathbf{S}(t)\widehat{\mathbf{A}}_x^{1/2} + \widehat{\mathbf{L}}_y^{(2),1/2}\mathbf{S}(t)\widehat{\mathbf{A}}_y^{1/2} + \widehat{\mathbf{L}}_x^{(1),1/2}\mathbf{S}(t)|\widehat{\mathbf{A}}_x|^{1/2} + \widehat{\mathbf{L}}_y^{(1),1/2}\mathbf{S}(t)|\widehat{\mathbf{A}}_y|^{1/2} . \end{aligned} \tag{7.25}$$

The numerical costs to compute update matrices and to evaluate the right-hand side of the S-step equations are of $O(\mathcal{R}^2 \cdot (n_x + m))$. To point out that the coefficient matrix is not yet updated by the scattering step (eq. (7.22b)), we define $\mathbf{S}^{1/2} := \mathbf{S}_I(t_1)$.

In a second step, we use the updated factors as the initial condition, i.e., $\mathbf{X}_{II}(t_0) = \mathbf{X}^{1/2}$, $\mathbf{S}_{II}(t_0) = \mathbf{S}^{1/2}$ and $\mathbf{W}_{II}(t_0) = \mathbf{W}^{1/2}$. Again Roman numbers are omitted in the following. However, we do include a subscript 1 to denote that we are solving for the factors of $\mathbf{u}_1$. For the scattering step (eq. (7.22b)), determining the K, L and S-steps is straightforward and leads to

$$\dot{\mathbf{K}}_1(t) = -\Sigma_t(t)\mathbf{K}_1(t) + \boldsymbol{\psi}(t_1)\mathbf{T}_M\boldsymbol{\Sigma}(t)\mathbf{W}_1^{1/2}, \tag{7.26a}$$

$$\dot{\mathbf{L}}_1(t) = -\Sigma_t(t)\mathbf{L}_1(t) + \mathbf{X}_1^{1/2,T}\boldsymbol{\psi}(t_1)\mathbf{T}_M\boldsymbol{\Sigma}(t) , \tag{7.26b}$$

$$\dot{\mathbf{S}}_1(t) = -\Sigma_t(t)\mathbf{S}_1(t) + \mathbf{X}_1^{1,T}\boldsymbol{\psi}(t_1)\mathbf{T}_M\boldsymbol{\Sigma}(t)\mathbf{W}_1^1 . \tag{7.26c}$$

The time updated solution after streaming and scattering is given as $\mathbf{u}_1(t_1) = \mathbf{X}_1(t_1)\mathbf{S}_1(t_1)\mathbf{W}_1(t_1)^T$. When the directed S_N quadrature set has n_q nodes, computational costs are of $O(\mathcal{R}\cdot n_q \cdot (n_x + m))$. Note that since radiation therapy uses highly peaked particle beams as boundary conditions or source terms, only a limited number of directions needs to be resolved by the quadrature, i.e., n_q is expected to be small.

In the same manner, evolution equations for the factors of the solutions to the remaining moment equations (7.21c) and (7.21d) are derived. Here, the streaming update can be determined with eq. (7.23), eq. (7.24) and eq. (7.25) (except for eq. (7.21a), since for $\boldsymbol{\psi}$ we use a directed S_N method instead of a dynamical low-rank approximation). The scattering update for a general $\ell = 2, \cdots, L$ which we denote by a subscript reads

$$\begin{aligned}\dot{\mathbf{K}}_\ell(t) &= -\Sigma_t(t)\mathbf{K}_\ell(t) + \mathbf{X}_{\ell-1}^1\mathbf{S}_{\ell-1}^1\mathbf{W}_{\ell-1}^{1,T}\boldsymbol{\Sigma}(t)\mathbf{W}_\ell^{1/2} , \\ \dot{\mathbf{L}}_\ell(t) &= -\Sigma_t(t)\mathbf{L}_\ell(t) + \mathbf{X}_\ell^{1/2,T}\mathbf{X}_{\ell-1}^1\mathbf{S}_{\ell-1}^1\mathbf{W}_{\ell-1}^{1,T}\boldsymbol{\Sigma}(t) , \\ \dot{\mathbf{S}}_\ell(t) &= -\Sigma_t(t)\mathbf{S}_\ell(t) + \mathbf{X}_\ell^{1,T}\mathbf{X}_{\ell-1}^1\mathbf{S}_{\ell-1}^1\mathbf{W}_{\ell-1}^{1,T}\boldsymbol{\Sigma}(t)\mathbf{W}_\ell^1 . \end{aligned}$$

Computational costs are of $O(\mathcal{R}^2 \cdot (n_x + m))$. Lastly, for the fully collided solution $\mathbf{u}_c$ we perform a further splitting step. Omitting the subscript c, we have

$$\dot{\mathbf{u}}_I(t) = \mathbf{F}(t, \mathbf{u}_I(t)) , \qquad \mathbf{u}_I(t_0) = \boldsymbol{u}_1(t_0) , \tag{7.28a}$$

$$\dot{\mathbf{u}}_{II}(t) = \mathbf{u}_L(t_1)\boldsymbol{\Sigma}(t) , \qquad \mathbf{u}_{II}(t_0) = \mathbf{u}_I(t_1) , \tag{7.28b}$$

$$\dot{\mathbf{u}}_{III}(t) = -\Sigma_t(t)\mathbf{u}_{III}(t) + \mathbf{u}_{III}(t)\boldsymbol{\Sigma}(t) , \qquad \mathbf{u}_{III}(t_0) = \mathbf{u}_{II}(t_1) . \tag{7.28c}$$

In this case, the K, S, L-equations for inscattering from $\mathbf{u}_L$, i.e., equation (7.28b) read (omitting Roman

indices)

$$\dot{\mathbf{K}}_c(t) = \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \boldsymbol{\Sigma}(t) \mathbf{W}_c^{1/2}, \tag{7.29a}$$

$$\dot{\mathbf{L}}_c(t) = \mathbf{X}_c^{1/2,T} \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \boldsymbol{\Sigma}(t) \,, \tag{7.29b}$$

$$\dot{\mathbf{S}}_c(t) = \mathbf{X}_c^{1,T} \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \boldsymbol{\Sigma}(t) \mathbf{W}_c^1 \,. \tag{7.29c}$$

For the in-scattering and out-scattering of the collided flux, i.e., equation (7.28c) we use the matrix projector–splitting integrator. Following [111], only the L-step needs to be computed and we are left with

$$\dot{\mathbf{L}}_c(t) = -\Sigma_t(t) \mathbf{L}_c(t) + \mathbf{L}_c(t) \boldsymbol{\Sigma}(t). \tag{7.30}$$

The costs for the collided particles are again $O(\mathcal{R}^2 \cdot (n_x + m))$. According to the derived steps, the scheme then consecutively updates the uncollided particles $\boldsymbol{\psi}$, the factors of $\mathbf{u}_1, \cdots, \mathbf{u}_L$ and lastly the factors of $\mathbf{u}_c$. In every step, the factors are first updated by a streaming step, followed by a scattering step. Lastly, when updating the factors of the collided flux, the additional L-step is performed to account for self-scattering. This procedure is repeated until a final time (or minimal energy) is reached. A flow chart to visualize the presented method is given in figure 7.1.

7.5.2 Time (or energy) discretization

The presented equations still continuously depend on the pseudo-time (or energy) t. To treat stiff scattering terms, we use an implicit time update method for the scattering equations. The remainder uses explicit time discretizations. More specifically, we use implicit and explicit Euler time-discretizations in this work. Let us start with $\mathbf{X}_\ell^n = \mathbf{X}_\ell^0$, $\mathbf{S}_\ell^n = \mathbf{S}_\ell^0$ and $\mathbf{W}_\ell^n = \mathbf{W}_\ell^0$, where ℓ denotes the individual collision steps, i.e., $\ell \in \{1, \cdots, L-1, L, c\}$. The streaming update is the same for all collision steps. Hence, when omitting a specific collision index, we obtain

$$\mathbf{K}^{1/2} = \mathbf{K}^0 + \Delta t \left(\mathbf{L}_x^{(2)} \mathbf{K}^0 \widehat{\mathbf{A}}_x^0 + \mathbf{L}_y^{(2)} \mathbf{K}^0 \widehat{\mathbf{A}}_y^0 + \mathbf{L}_x^{(1)} \mathbf{K}^0 |\widehat{\mathbf{A}}_x|^0 + \mathbf{L}_y^{(1)} \mathbf{K}^0 |\widehat{\mathbf{A}}_y|^0 \right) \,, \qquad \mathbf{X}^{1/2} \mathbf{R}_1 = \mathbf{K}^1 \,, \tag{7.31a}$$

$$\mathbf{L}^{1/2} = \mathbf{L}^0 + \Delta t \left(\widehat{\mathbf{L}}_x^{(2),0} \mathbf{L}^0 \mathbf{A}_x^T + \widehat{\mathbf{L}}_y^{(2),0} \mathbf{L}^0 \mathbf{A}_y^T + \widehat{\mathbf{L}}_x^{(1),0} \mathbf{L}^0 |\mathbf{A}_x|^T + \widehat{\mathbf{L}}_y^{(1),0} \mathbf{L}^0 |\mathbf{A}_y|^T \right) \,, \qquad \mathbf{W}^{1/2} \mathbf{R}_2 = \mathbf{L}^{1,T} \,, \tag{7.31b}$$

$$\mathbf{S}^{1/2} = \widetilde{\mathbf{S}}^0 + \Delta t \left(\widehat{\mathbf{L}}_x^{(2),1/2} \widetilde{\mathbf{S}}^0 \widehat{\mathbf{A}}_x^{1/2} + \widehat{\mathbf{L}}_y^{(2),1/2} \widetilde{\mathbf{S}}^0 \widehat{\mathbf{A}}_y^{1/2} + \widehat{\mathbf{L}}_x^{(1),1/2} \widetilde{\mathbf{S}}^0 |\widehat{\mathbf{A}}_x|^{1/2} + \widehat{\mathbf{L}}_y^{(1),1/2} \widetilde{\mathbf{S}}^0 |\widehat{\mathbf{A}}_y|^{1/2} \right) \,, \tag{7.31c}$$

where $\widetilde{\mathbf{S}}^0 = \mathbf{X}^{1/2,T} \mathbf{X}^0 \mathbf{S}^0 \mathbf{W}^{0,T} \mathbf{W}^{1/2}$ and flux matrices are computed before evaluating the right-hand side. The collision equations differ for the $\ell = 1$ and $\ell \in \{2, \cdots, L\}$ collided fluxes as well as the collided flux. For $\ell = 1$, we have

$$\mathbf{K}_1^1 = \frac{1}{1 + \Delta t \Sigma(t_1)} \left(\mathbf{K}_1^{1/2} + \Delta t \boldsymbol{\psi}(t_1) \mathbf{T}_M \boldsymbol{\Sigma}(t_1) \mathbf{W}_1^{1/2} \right) \,, \qquad \mathbf{X}_1^1 \mathbf{R}_1 = \mathbf{K}_1^1 \,, \tag{7.32a}$$

$$\mathbf{L}_1^1 = \frac{1}{1 + \Delta t \Sigma(t_1)} \left(\mathbf{L}_1^{1/2} + \Delta t \mathbf{X}_1^{1/2,T} \boldsymbol{\psi}(t_1) \mathbf{T}_M \boldsymbol{\Sigma}(t_1) \right) \,, \qquad \mathbf{W}_1^1 \mathbf{R}_2 = \mathbf{L}_1^{1,T} \,, \tag{7.32b}$$

$$\mathbf{S}_1^1 = \frac{1}{1 + \Delta t \Sigma(t_1)} \left(\widetilde{\mathbf{S}}_1^{1/2} + \Delta t \mathbf{X}_1^{1,T} \boldsymbol{\psi}(t_1) \mathbf{T}_M \boldsymbol{\Sigma}(t_1) \mathbf{W}_1^1 \right) \,. \tag{7.32c}$$

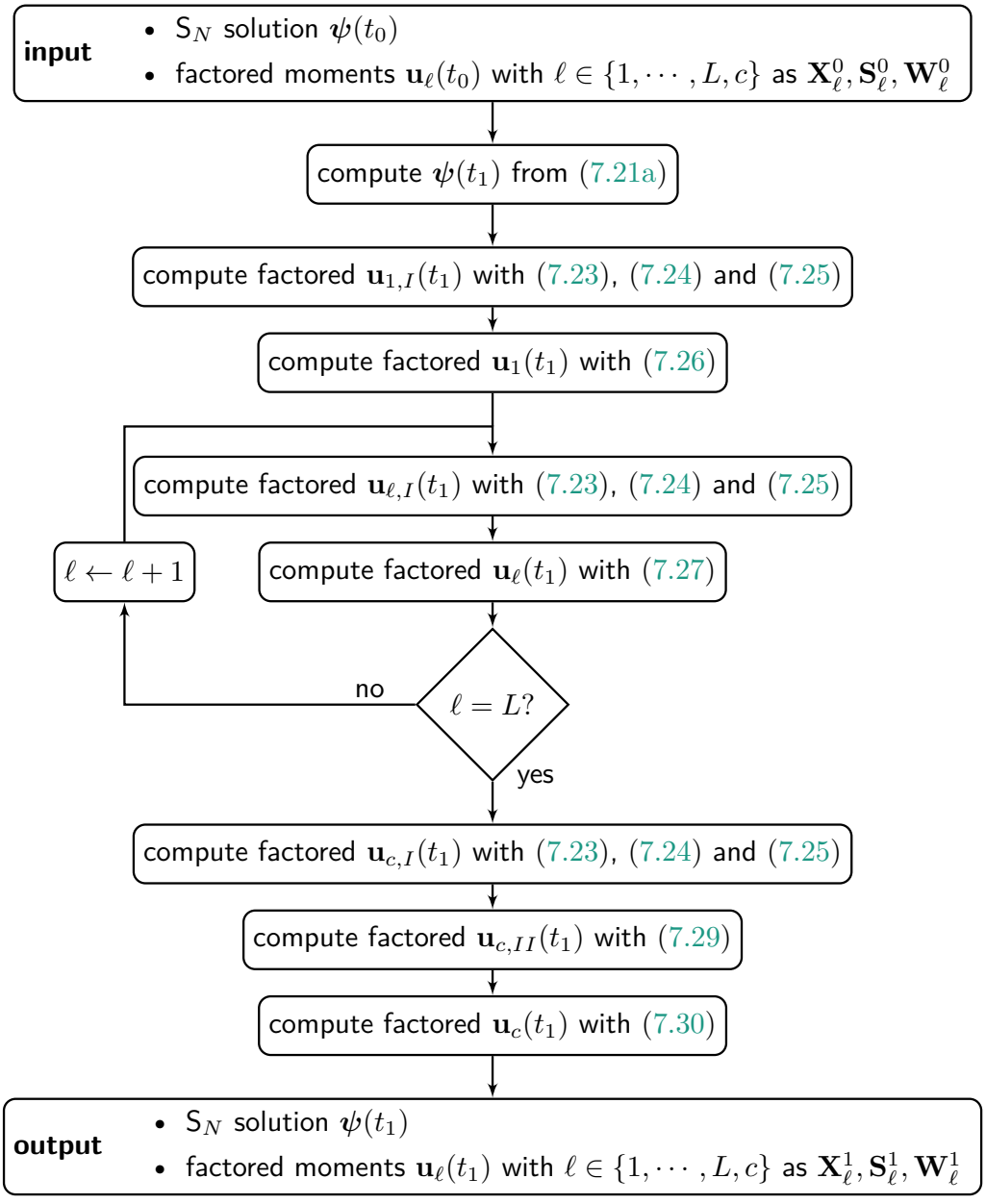


Figure 7.1: Flowchart of the presented method.

For $\ell \in \{2, \cdots, L\}$ we have

$$\mathbf{K}_\ell^1 = \frac{1}{1+\Delta t \Sigma(t_1)} \left(\mathbf{K}_\ell^{1/2} + \Delta t \mathbf{X}_{\ell-1}^1 \mathbf{S}_{\ell-1}^1 \mathbf{W}_{\ell-1}^{1,T} \mathbf{\Sigma}(t_1) \mathbf{W}_\ell^{1/2} \right) , \qquad \mathbf{X}_\ell^1 \mathbf{R}_1 = \mathbf{K}_\ell^1 , \tag{7.33a}$$

$$\mathbf{L}_\ell^1 = \frac{1}{1+\Delta t \Sigma(t_1)} \left(\mathbf{L}_\ell^{1/2} + \Delta t \mathbf{X}_\ell^{1/2,T} \mathbf{X}_{\ell-1}^1 \mathbf{S}_{\ell-1}^1 \mathbf{W}_{\ell-1}^{1,T} \mathbf{\Sigma}(t_1) \right) , \qquad \mathbf{W}_\ell^1 \mathbf{R}_2 = \mathbf{L}_\ell^{1,T} , \tag{7.33b}$$

$$\mathbf{S}_\ell^1 = \frac{1}{1+\Delta t \Sigma(t_1)} \left(\widetilde{\mathbf{S}}_\ell^{1/2} + \Delta t \mathbf{X}_\ell^{1,T} \mathbf{X}_{\ell-1}^1 \mathbf{S}_{\ell-1}^1 \mathbf{W}_{\ell-1}^{1,T} \mathbf{\Sigma}(t_1) \mathbf{W}_\ell^1 \right) . \tag{7.33c}$$

The collided flux is then updated through

$$\mathbf{K}_c^1 = \mathbf{K}_c^{1/2} + \Delta t \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \mathbf{\Sigma}(t_1) \mathbf{W}_c^{1/2} , \qquad \mathbf{X}_c^1 \mathbf{R}_1 = \mathbf{K}_c^1 , \tag{7.34a}$$

$$\mathbf{L}_c^1 = \mathbf{L}_c^{1/2} + \Delta t \mathbf{X}_c^{1/2,T} \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \mathbf{\Sigma}(t_1) , \qquad \widetilde{\mathbf{W}}_c^1 \mathbf{R}_2 = \mathbf{L}_c^{1,T} , \tag{7.34b}$$

$$\bar{\mathbf{S}}_c^1 = \widetilde{\mathbf{S}}_c^{1/2} + \Delta t \mathbf{X}_c^{1,T} \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \mathbf{\Sigma}(t_1) \widetilde{\mathbf{W}}_c^1, \tag{7.34c}$$

$$\widetilde{\mathbf{L}}_c^1 = \bar{\mathbf{S}}_c^1 \widetilde{\mathbf{W}}_c^{1,T} (\mathbf{I} + \Sigma_t(t_1) \Delta t \mathbf{I} - \Delta t \mathbf{\Sigma}(t_1))^{-1} , \qquad \mathbf{W}_c^1 \mathbf{S}_c^{1,T} = \widetilde{\mathbf{L}}_c^{1,T} . \tag{7.34d}$$

Note that since $\mathbf{\Sigma}$ is a diagonal matrix, the inversion in eq. (7.34d) is given explicitly without having to solve a linear system of equations. The time updated solution is then given by $\mathbf{X}_\ell^{n+1} = \mathbf{X}_\ell^1$, $\mathbf{S}_\ell^{n+1} = \mathbf{S}_\ell^1$ and $\mathbf{W}_\ell^{n+1} = \mathbf{W}_\ell^1$, where $\ell \in \{1, \cdots, L-1, L, c\}$.

Remark 7.1. *The proposed idea of multilevel DLRA can be applied in various settings with various strategies. The core ingredient is to write the solution as a sum of different contributions*

$$\mathbf{u}(t) = \mathbf{u}_1(t) + \mathbf{u}_2(t) + \cdots + \mathbf{u}_L(t) .$$

Strategies to write the solution as a sum of different components can be the use of telescoping identities, a split into symmetric and anti-symmetric solution contributions, a splitting of the original phase space, e.g. particles that move forward and backward and many more. In a second step, evolution equations for every component need to be derived. Third, every component $\mathbf{u}_i$ is represented through a low-rank factorization and evolution equations for every factor are derived with DLRA.

7.5.3 Boundary conditions

Boundary conditions for dynamical low-rank approximation are not straightforward to impose. Methods to realize boundary conditions preserve relevant basis functions by excluding such functions from the DLRA approximation [see e.g. 151, 110] or through rank adaptivity [87]. In this work, we shift the complicated structure of the boundary condition onto a solution component that can be efficiently treated with a conventional method. A main advantage of using an S_N method for uncollided particles, besides having to resolve a limited number of beam directions, is that its boundary conditions can be treated in a straightforward manner. Since at the beam position we have $\mathbf{n} \cdot \mathbf{\Omega} < 0$ for all directions resolved by the S_N method, we can impose $\psi_0 = \psi_{\text{BC}}$ at all spatial points and ordinates. Hence, we have a simple Dirichlet condition, which can be incorporated into the S_N method with the use of ghost cells. For the collided solution components, we assume $\mathbf{u}_\ell = \mathbf{0}$ and $\mathbf{u}_c = \mathbf{0}$ at the boundary. This encodes the fact that no collided particles enter from the boundary and collided particles that reach the boundary from within the computational domain will not reenter and can therefore be removed. This Dirichlet condition can be imposed with DLRA in a straightforward manner, see e.g. [110, Section 4].

7.6 L^2-stability of the proposed scheme

The derived method is robust in that its time step restriction (or CFL number) does not depend on small singular values of the coefficient matrix or stiff terms arising in the scattering step. By the choice of the splitting steps, we ensure that this stability is achieved without having to invert matrices or solve a nonlinear problem, which is commonly the case for implicit time integration methods.

In [114] the following theorems are proven, showing that the proposed method is L^2-stable under a time step restriction:

Lemma 7.1. *Assume that the CFL condition*

$$\frac{\lambda_{max}(\mathbf{A}_{x,y})}{\rho_{min}} \frac{\Delta t}{\Delta x} \leq \frac{1}{2} \tag{7.35}$$

holds true. Then, the streaming scheme in eq. (7.31) is L^2-stable, i.e.,

$$\|\mathbf{X}^{1/2}\mathbf{S}^{1/2}\mathbf{W}^{1/2,T}\|_F \leq \|\mathbf{X}^0\mathbf{S}^0\mathbf{W}^{0,T}\|_F \ .$$

Theorem 7.1. *Assume that the CFL conditions from eq. (7.35) and*

$$\max_k \frac{1}{1 + \Delta t\Sigma_t(t) - \Delta t\Sigma_{kk}(t)} \leq 1 \tag{7.36}$$

hold true for all pseudo-times $t \in [0, T]$. Then, the scheme is L^2-stable, in the sense that with $\mathbf{u}_c^1 := \mathbf{X}_c^1\mathbf{S}_c^1\mathbf{W}_c^{1,T}$ and $\mathbf{u}_\ell^1 := \mathbf{X}_\ell^1\mathbf{S}_\ell^1\mathbf{W}_\ell^{1,T}$ we have

$$\|\mathbf{u}_c^1\|_F + \sum_{\ell=1}^{L} \|\mathbf{u}_\ell^1\|_F + \|\boldsymbol{\psi}(t_1)\|_F \leq \|\mathbf{u}_c^0\|_F + \sum_{\ell=1}^{L} \|\mathbf{u}_\ell^0\|_F + \|\boldsymbol{\psi}(t_0)\|_F \ .$$

The proofs are omitted in this thesis, all details can be found in [114].

7.7 Extension to rank adaptivity

In a last step, we discuss the extension of the proposed scheme to the rank adaptive integrator of [35] and how it simplifies for a forward Euler time discretization. The core ingredient of this method is to extend the time updated basis with the basis at time t_0. Hence, for the streaming step, the updated basis becomes $\widehat{\mathbf{X}}^{1/2} = [\mathbf{X}^0, \overline{\mathbf{X}}^{1/2}]$, where $\overline{\mathbf{X}}^{1/2}$ is chosen such that the column range of $\widehat{\mathbf{X}}^{1/2}$ contains $\mathbf{K}(t_1)$ from the streaming K-step and the basis is orthonormal, i.e., $\mathbf{X}^{0,T}\overline{\mathbf{X}}^{1/2} = \mathbf{0}$ and $\overline{\mathbf{X}}^{1/2,T}\overline{\mathbf{X}}^{1/2} = \mathbf{I}$. Thus, the matrix to compute the initial condition of the S-step reads $\widehat{\mathbf{M}} = \widehat{\mathbf{X}}^{1/2,T}\mathbf{X}^0 = [\mathbf{I}, \mathbf{0}]^T \in \mathbb{R}^{2\mathcal{R}_0 \times \mathcal{R}_0}$, where $\mathcal{R}_0$ is the rank at time t_0. In the same manner, we have $\widehat{\mathbf{N}} = [\mathbf{I}, \mathbf{0}]^T \in \mathbb{R}^{2\mathcal{R}_0 \times \mathcal{R}_0}$. Hence, as pointed out in [35], we have $\widehat{\mathbf{X}}^{1/2}\widehat{\mathbf{S}}(t_0)\widehat{\mathbf{W}}^{1/2,T} = \mathbf{X}^0\mathbf{S}(t_0)\mathbf{W}^{0,T} \in \mathcal{M}_{\mathcal{R}_0}$. The fact that the initial condition of the S-step is of rank $\mathcal{R}_0$ can be used to reduce computational costs when using an explicit Euler step. The S-step of the rank adaptive integrator for the streaming step then reads

$$\widehat{\mathbf{S}}^{1/2} = [\mathbf{I}, \mathbf{0}]^T \ \mathbf{S}^0 \ [\mathbf{I}, \mathbf{0}] + \Delta t \left(\overline{\mathbf{L}}_x^{(2),1/2}\mathbf{S}^0\overline{\mathbf{A}}_x^{1/2} + \overline{\mathbf{L}}_y^{(2),1/2}\mathbf{S}^0\overline{\mathbf{A}}_y^{1/2} + \overline{\mathbf{L}}_x^{(1),1/2}\mathbf{S}^0|\overline{\mathbf{A}}_x|^{1/2} + \overline{\mathbf{L}}_y^{(1),1/2}\mathbf{S}^0|\overline{\mathbf{A}}_y|^{1/2} \right) \ ,$$

where the flux matrices are given by

$$\overline{\mathbf{A}}_{x,y}^{1/2} := \mathbf{W}^{0,T}\mathbf{A}_{x,y}^T\widehat{\mathbf{W}}^{1/2} \in \mathbb{R}^{\mathcal{R}_0 \times 2\mathcal{R}_0} \ , \qquad |\overline{\mathbf{A}}_{x,y}|^{1/2} := \mathbf{W}^{0,T}|\mathbf{A}_{x,y}|^T\widehat{\mathbf{W}}^{1/2} \in \mathbb{R}^{\mathcal{R}_0 \times 2\mathcal{R}_0} \ ,$$
$$\overline{\mathbf{L}}_{x,y}^{(2),1/2} := \widehat{\mathbf{X}}^{1/2,T}\mathbf{L}_{x,y}^{(2)}\mathbf{X}^0 \in \mathbb{R}^{2\mathcal{R}_0 \times \mathcal{R}_0} \ , \qquad \overline{\mathbf{L}}_{x,y}^{(1),1/2} := \widehat{\mathbf{X}}^{1/2,T}\mathbf{L}_{x,y}^{(1)}\mathbf{X}^0 \in \mathbb{R}^{2\mathcal{R}_0 \times \mathcal{R}_0} \ .$$

Thus, the flux matrices that need to be computed have $2\mathcal{R}_0^2$ entries. When using more general time integration schemes to solve the S-step for the rank adaptive integrator, the flux matrices have $4\mathcal{R}_0^2$ entries. We determine the solution factors after the streaming update $\mathbf{X}^{1/2}, \mathbf{S}^{1/2}$ and $\mathbf{W}^{1/2}$ as well as the rank $\mathcal{R}_{1/2}$ through the truncation step of the rank adaptive integrator. For components $\ell \in \{1, 2, \cdots, L, c\}$, the scattering steps have modified S-step equations

$$\begin{aligned}
\widehat{\mathbf{S}}_1^1 &= \frac{1}{1+\Delta t \Sigma(t_1)} \left([\mathbf{I}, \mathbf{0}]^T \ \mathbf{S}_1^{1/2} \ [\mathbf{I}, \mathbf{0}] + \Delta t \widehat{\mathbf{X}}_1^{1,T} \boldsymbol{\psi}(t_1) \mathbf{T}_M \boldsymbol{\Sigma}(t_1) \widehat{\mathbf{W}}_1^1 \right) , \\
\widehat{\mathbf{S}}_\ell^1 &= \frac{1}{1+\Delta t \Sigma(t_1)} \left([\mathbf{I}, \mathbf{0}]^T \ \mathbf{S}_\ell^{1/2} \ [\mathbf{I}, \mathbf{0}] + \Delta t \widehat{\mathbf{X}}_\ell^{1,T} \mathbf{X}_{\ell-1}^1 \mathbf{S}_{\ell-1}^1 \mathbf{W}_{\ell-1}^{1,T} \boldsymbol{\Sigma}(t_1) \widehat{\mathbf{W}}_\ell^1 \right) , \quad \text{for } \ell = 2, \cdots, L , \\
\widehat{\mathbf{S}}_c^1 &= [\mathbf{I}, \mathbf{0}]^T \ \mathbf{S}_c^{1/2} \ [\mathbf{I}, \mathbf{0}] + \Delta t \widehat{\mathbf{X}}_c^{1,T} \mathbf{X}_L^1 \mathbf{S}_L^1 \mathbf{W}_L^{1,T} \boldsymbol{\Sigma}(t_1) \widehat{\mathbf{W}}_c^1 .
\end{aligned}$$

For $\ell \in \{1, \cdots, L, c\}$ we use $\widehat{\mathbf{X}}_\ell^1 = [\mathbf{X}_\ell^{1/2}, \overline{\mathbf{X}}_\ell^1]$, where $\overline{\mathbf{X}}_\ell^1$ is chosen such that the column range of $\widehat{\mathbf{X}}_\ell^1$ contains $\mathbf{K}_\ell(t_1)$ from the scattering K-step and the basis is orthonormal, i.e., $\mathbf{X}_\ell^{1/2,T} \overline{\mathbf{X}}_\ell^1 = \mathbf{0}$ and $\overline{\mathbf{X}}_\ell^{1,T} \overline{\mathbf{X}}_\ell^1 = \mathbf{I}$. The directional basis $\widehat{\mathbf{W}}_\ell^1$ is defined analogously. Note that for the scattered particles, we need to do a final L-step (7.34d) after having updated the coefficient. In this case, the truncation step of the rank adaptive integrator yields $\widetilde{\mathbf{W}}_c^1$ and $\bar{\mathbf{S}}_c^1$. Since eq. (7.34d) is constructed through the fixed-rank projector–splitting integrator, it will not modify the rank.

Besides allowing for a dynamic choice of the rank, the rank adaptive integrator remains L^2 stable.

Lemma 7.2. *Assume that the CFL condition from eq. (7.35) holds true. Then, the streaming scheme of the rank adaptive integrator is L^2-stable, i.e.,*

$$\|\mathbf{X}^{1/2} \mathbf{S}^{1/2} \mathbf{W}^{1/2,T}\|_F \leq \|\mathbf{X}^0 \mathbf{S}^0 \mathbf{W}^{0,T}\|_F . \tag{7.37}$$

Theorem 7.2. *Assume that the CFL conditions from eq. (7.35) and eq. (7.36) hold true for all pseudo-times $t \in [0, T]$. Then, the scheme is L^2-stable, in the sense that with $\mathbf{u}_c^1 := \mathbf{X}_c^1 \mathbf{S}_c^1 \mathbf{W}_c^{1,T}$ and $\mathbf{u}_\ell^1 := \mathbf{X}_\ell^1 \mathbf{S}_\ell^1 \mathbf{W}_\ell^{1,T}$ we have*

$$\|\mathbf{u}_c^1\|_F + \sum_{\ell=1}^{L} \|\mathbf{u}_\ell^1\|_F + \|\boldsymbol{\psi}(t_1)\|_F \leq \|\mathbf{u}_c^0\|_F + \sum_{\ell=1}^{L} \|\mathbf{u}_\ell^0\|_F + \|\boldsymbol{\psi}(t_0)\|_F .$$

Again, the full stability proofs can be found in [114].

7.8 Implementation

The proposed method including options for rank adaptivity and a multi-level collided-uncollided split has been implemented in julia and is openly available on GitHub. First results presented in [114] used a 2D electron code running on CPUs. The current version includes GPU-based solvers using CUDA and also supports three-dimensional spatial domains as well as proton dose calculations.

7.9 Results: Electron transport

In the following, we demonstrate numerical experiments to compare conventional and the proposed methods. First, we present results for electron transport published in [114]. All results can be reproduced with the openly available code framework [113].

7.9.1 2D results

Line source benchmark

To demonstrate the applicability of the proposed collision source method for dynamical low-rank approximation in general radiation transport applications, we first take a look at the time-dependent radiation transport equation for the *line source* benchmark [62, 63]

$$\begin{aligned} &\partial_t \psi + \mathbf{\Omega} \cdot \nabla_{\mathbf{r}} \psi + \Sigma_s \psi = \frac{\Sigma_s}{4\pi} \int_{\mathbb{S}^2} \psi \, \mathrm{d}\mathbf{\Omega} \,, \qquad (\mathbf{r}, \mathbf{\Omega}) \in [-1.5, 1.5]^2 \times \mathbb{S}^2 \,, \\ &\psi(t_0) = \frac{1}{4\pi\sigma^2} \exp\left(-\frac{\|\mathbf{r}\|^2}{4\sigma^2} \right) \,, \end{aligned} \tag{7.38}$$

where $\Sigma_s = 1$ and $\sigma = 0.03$. This equation can be recovered from the continuous slowing down approximation when choosing $\rho \equiv 1$ and treating the energy variable as time. The line source benchmark is a common test case for radiation transport problems, exposing deficiencies of different methods. A comparison of conventional methods for this benchmark can for example be found in [64]. Common methods require high computational costs or parameter tuning to yield a satisfactory approximation. Uses of dynamical low-rank approximation for this benchmark are [172, 171, 35], where it is observed that high ranks are needed to achieve a desired level of accuracy. Nevertheless, in comparison to classical methods, DLRA yields reduced run times and memory requirements. We use the following computational parameters for our calculations:

$n_x = N_x \cdot N_y = 40000$	number of spatial cells
$m = (N+1)^2 = 484$	number of spherical basis functions
$n_q = 968$	number of quadrature points for uncollided flux
$t_{\text{end}} = 1$	end time

We use a CFL number of 0.5 according to eq. (7.35). That is, the time step size is chosen as $\Delta t = \rho_{\min} \frac{\Delta x}{2}$. The scalar flux $\Phi(t = 1, \mathbf{r}) = \int_{\mathbb{S}^2} \psi(t = 1, \mathbf{r}, \mathbf{\Omega}) \, \mathrm{d}\mathbf{\Omega}$ computed with different methods can be found in Figure 7.2. We observe a significant increase in the solution quality when using $L = 4$ instead of $L = 1$ levels. The level 4 approximation with a tolerance parameter of $\vartheta = 0.3$ agrees well with the P_N solution. Here, we use the term P_N to indicate the use of an S_N solver for uncollided particles as well as a P_N solver for the remainder. While P_N takes 5408 seconds to compute the scalar flux at time $t = 1$, the DLRA methods with $L = 1$ and $L = 4$ levels only require 1009 and 1278 seconds respectively. Since particles move into all directions, a main factor in this run time is the S_N solution. Taking a look at the rank evolution in time, depicted in Figure 7.3, we see that most information is carried by the uncollided flux as well as solution components with little collisions.

Treatment planning for lung patient

In the following, we examine the application of the proposed method to a realistic 2D CT scan of a lung patient. The patient is radiated with an electron beam of $E_{\max} = 21$ MeV. We model this beam as

$$\begin{aligned} \psi_{\text{in}}(E, \mathbf{r}, \mathbf{\Omega}) = 10^5 \cdot \exp(-(\Omega_{1,*} - \Omega_1)^2 / \sigma_{\Omega_1}) \cdot \exp(-(E_{\max} - E)^2 / \sigma_E) \\ \cdot \exp(-(r_x^* - r_x)^2 / \sigma_x) \cdot \exp(-(r_y^* - r_y)^2 / \sigma_y) \,, \end{aligned}$$

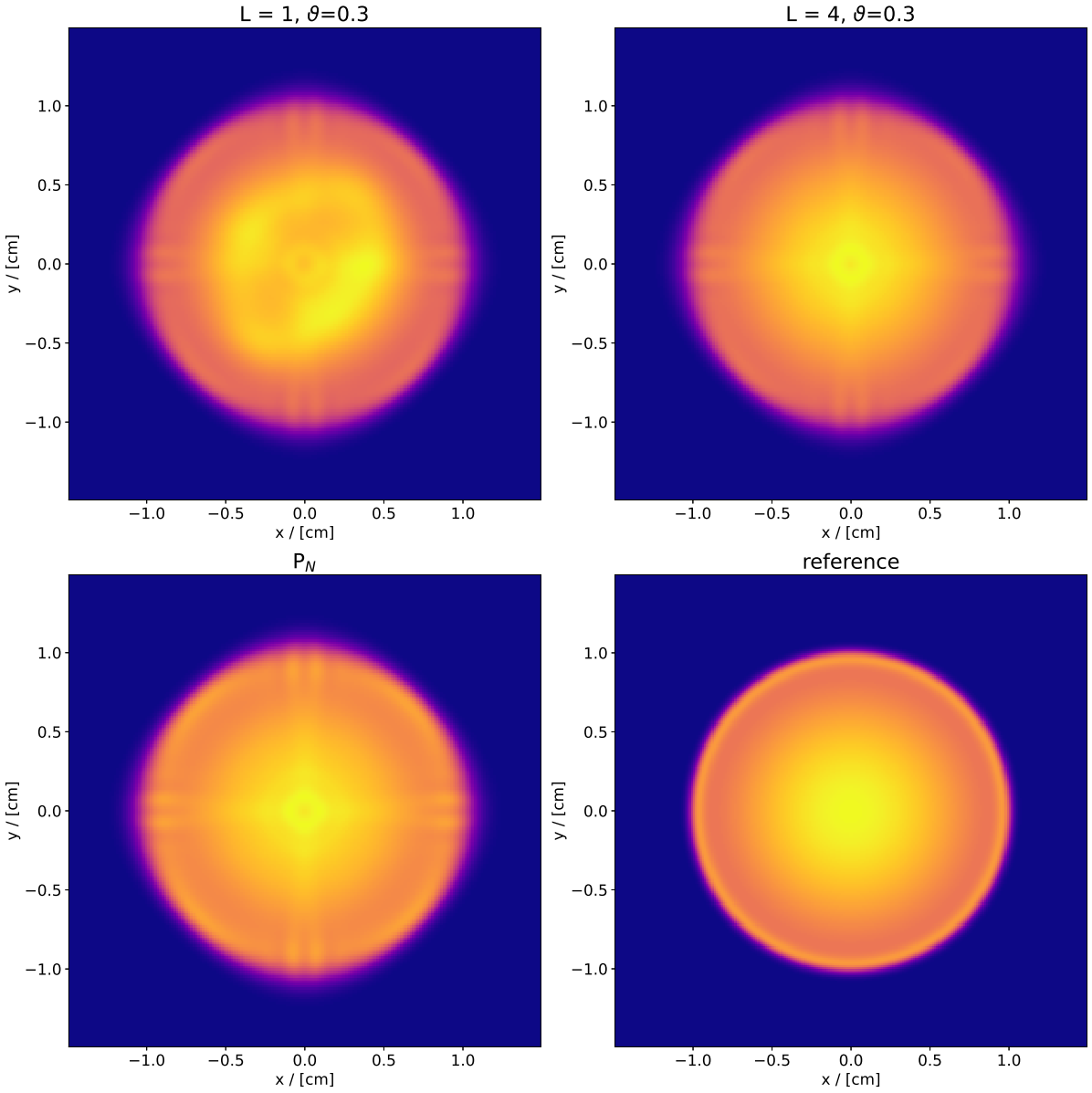
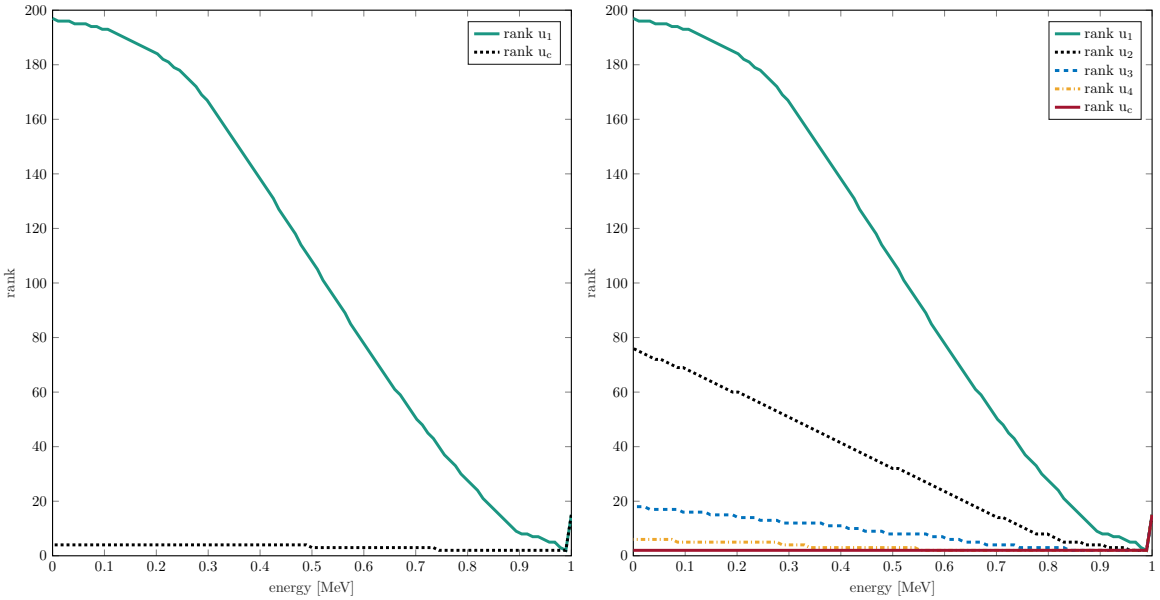


Figure 7.2: Scalar flux $\Phi(t=1,\mathbf{r}) = \int_{\mathbb{S}^2} \psi(t=1,\mathbf{r},\mathbf{\Omega})\,\mathrm{d}\mathbf{\Omega}$ with different methods and analytic reference solution.



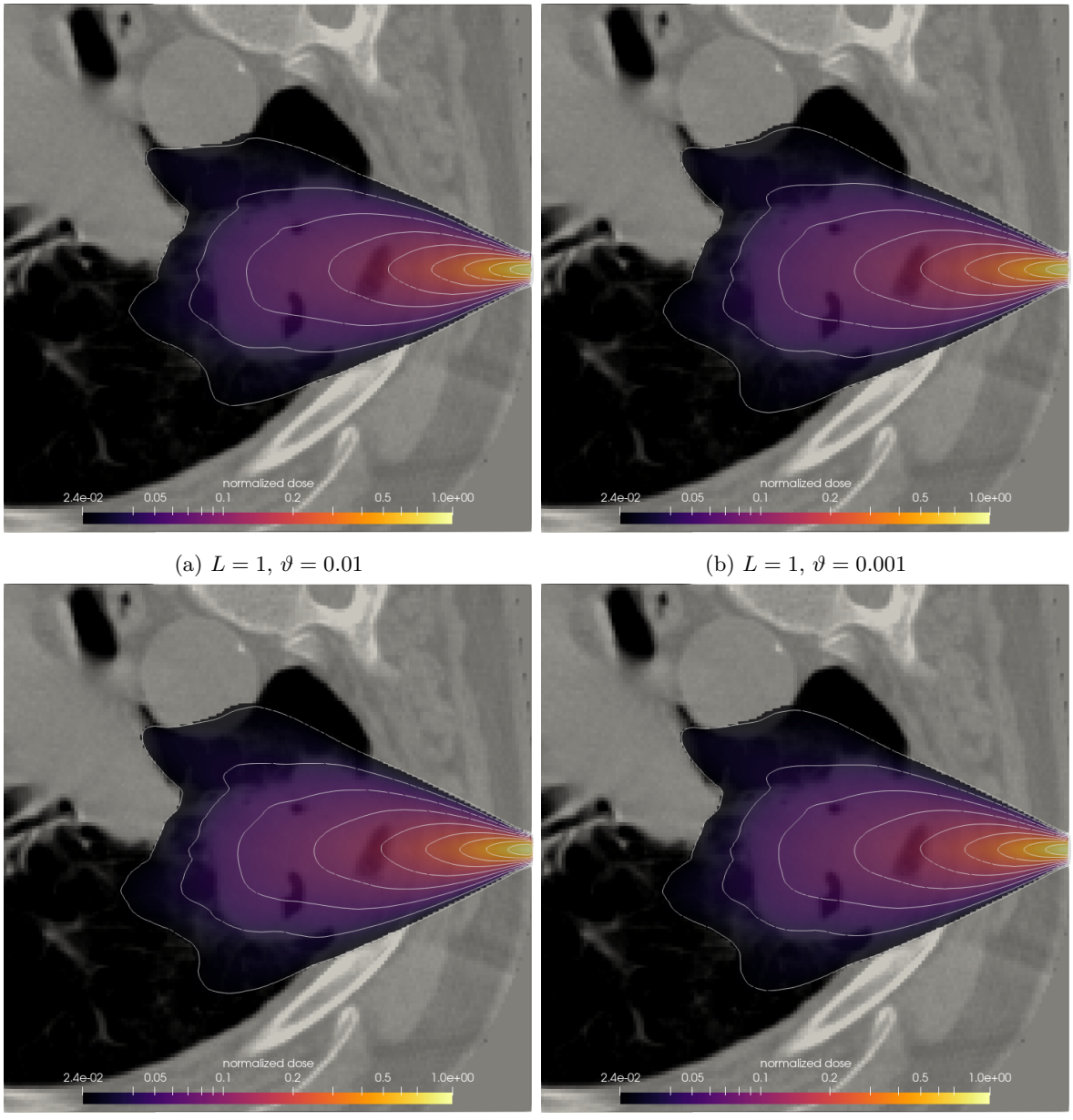
(a) $\vartheta = 0.3, L = 1$ (b) $\vartheta = 0.3, L = 4$

Figure 7.3: Rank evolution in energy for different numbers of intermediate collision levels L.

which is used as boundary condition for the uncollided particle flux. To determine a tissue density ρ for given gray-scale values of the CT image, we assume a value of one, i.e., a white pixel, to consist of bone material with density $\rho_{\text{bone}} = 1.85$ g/cm^3. The remaining tissue is scaled such that a pixel value of zero corresponds to a minimum density of $\rho_{\text{min}} = 0.05$ g/cm^3. Air around the patient is filled with material since this region does not impact the dose distribution. Note, that CT-calibration curves or density look-up tables can be used for a more sophisticated mapping of pixel values to density [see e.g. 188, 189]. The chosen settings are the same as in section 7.9.1. Since we are using a directed particle beam as boundary condition for the uncollided particles, the number of quadrature points n_q reduces by over 59 percent. The remaining parameters are:

$n_q = 396$	number of quadrature points for uncollided flux
$E_{\text{max}} = 21$	energy of beam in MeV
$r_x^* = 7.25, r_y^* = 14.5$	spatial mean of particle beam in cm
$\Omega_{1,*} = 1$	directional mean of particle beam
$\sigma_{\Omega_1}^{-1} = 75$	inverse directional beam variance
$\sigma_x^{-1} = \sigma_y^{-1} = 20$	inverse spatial beam variance
$\sigma_E^{-1} = 100$	inverse energy variance

We again use a CFL number of 0.5 according to eq. (7.35), leading to $\Delta t = \rho_{\text{min}} \frac{\Delta x}{2}$. For this setting, we compute the full P_N solution, the proposed dynamical low-rank method with a fixed rank of 50 consisting only of collided and uncollided particles as well as the rank adaptive version with $L = 1$ intermediate levels. Due to its reduced computational costs, the DLRA methods show a significantly reduced run time. While the full P_N method runs for 47329 seconds, the DLRA methods have a run time of 4373, 3917 and 5392 seconds respectively.



(a) $L = 1$, $\vartheta = 0.01$

(b) $L = 1$, $\vartheta = 0.001$

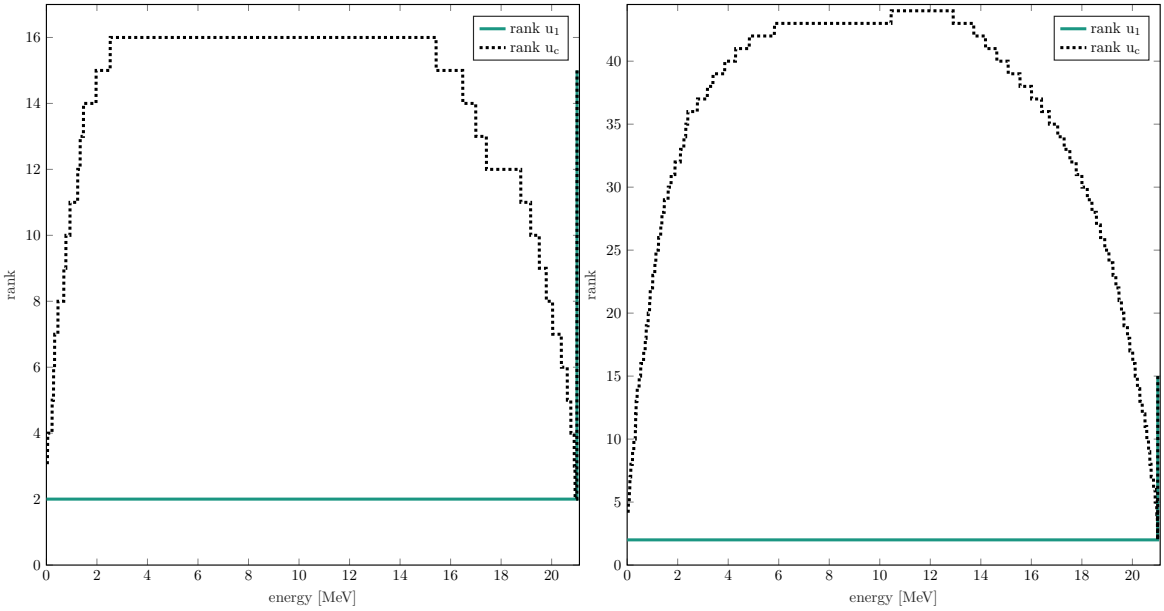
(c) $L = 0$, fixed rank $r = 50$

(d) P_N

Figure 7.4: Dose distribution with different methods.

The resulting dose distribution can be found in Figure 7.4. All considered variations of the proposed method are able to capture the effect of heterogeneities in the patient density and agree very well with the P_N solution in the relevant dose areas. The efficiency of the method concerning both time and memory is a step towards feasibility for practical applications. This includes the generation of optimal treatment plans with gradient-based optimization methods. It is observed that choosing a low refinement tolerance $\vartheta = \bar{\vartheta} \cdot \|\mathbf{S}\|_F$ with $\bar{\vartheta} = 0.01$ leads to a very slight difference to the full solution for the smallest two isolines. However, these deviations are barely visible to the naked eye.

The corresponding ranks at different energies for $\bar{\vartheta} \in \{0.001, 0.01\}$ are depicted in Figure 7.5. It is observed that the rank of collided particles remains small at low and high energies. At intermediate energies, the rank reaches its maximum. A possible explanation for this is, that advection (which is dominant at the beginning) and diffusion (which is dominant at the end), are both individually low rank. Thus, higher ranks are mainly required at medium energies, where a complex interplay of both effects takes place. Particles that have collided once can be described with a small rank throughout the simulation. Note that particles which have collided once are only present for energies below 21 MeV. The reason for this is that particles enter the patient tissue with maximum energy and are then directly subject to energy loss and scattering. The first four dominant spatial modes at the lowest energy are depicted



(a) $\bar{\vartheta} = 0.01$ (b) $\bar{\vartheta} = 0.001$

Figure 7.5: Rank evolution in energy for different $\vartheta = \bar{\vartheta} \cdot \|\mathbf{S}\|_F$.

in Figure 7.6 and the first four dominant directional modes are shown in Figure 7.7. These modes have been computed by an SVD of the coefficient matrix $\mathbf{S} = \mathbf{UDV}^T$. We then plot the first four columns of $\mathbf{XU}$ and $\mathbf{WV}$. The directional basis carries the information that particles are predominantly traveling into the x-direction, i.e., in direction of the particle beam. The spatial basis encodes that particles with low energies are situated at the left of the CT scan and can mostly be found in high-density tissue. Due to the low energy of these particles, no significant contribution to the overall dose distribution is observed.

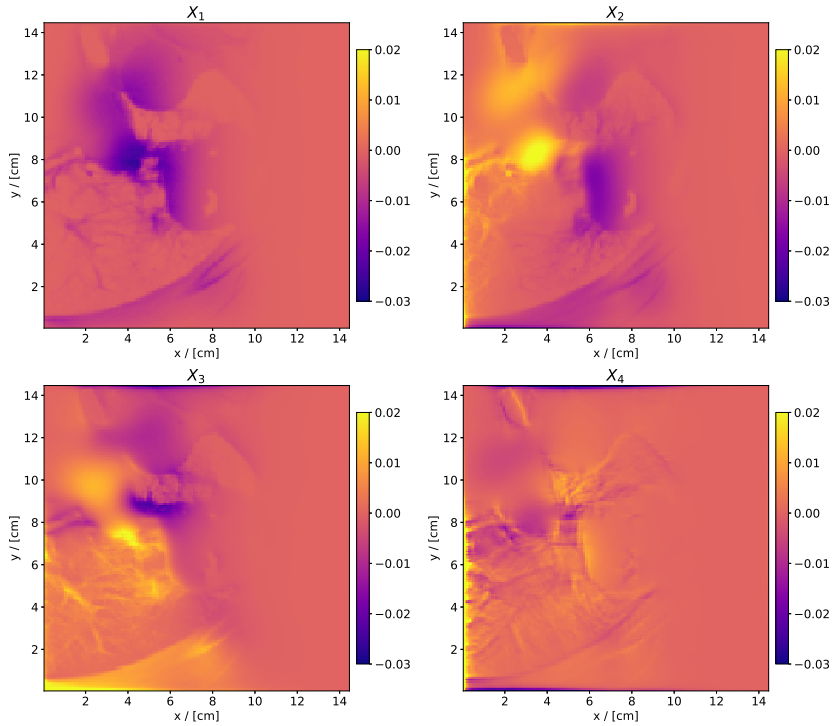


Figure 7.6: First four dominant spatial modes with fixed rank integrator.

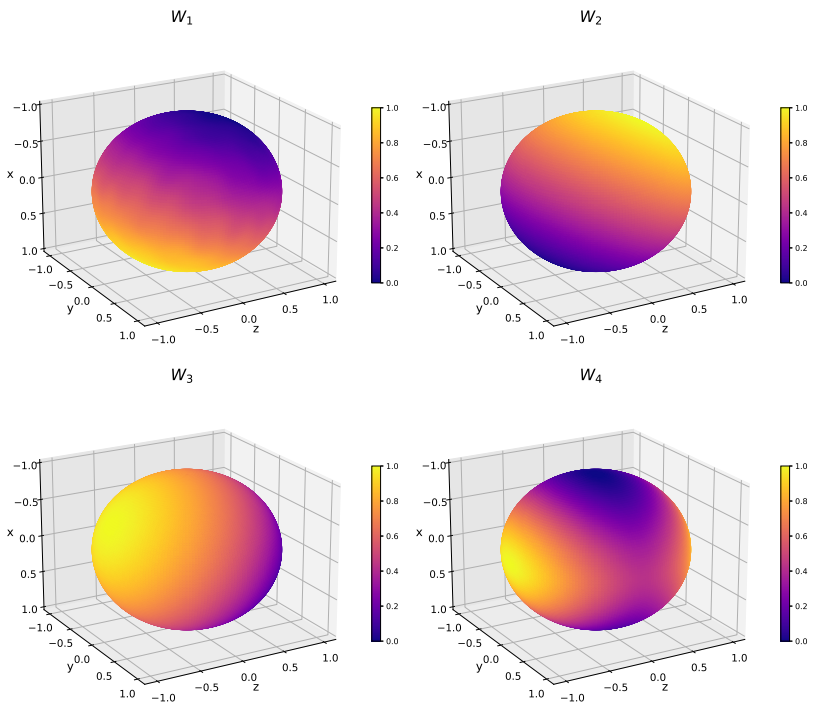


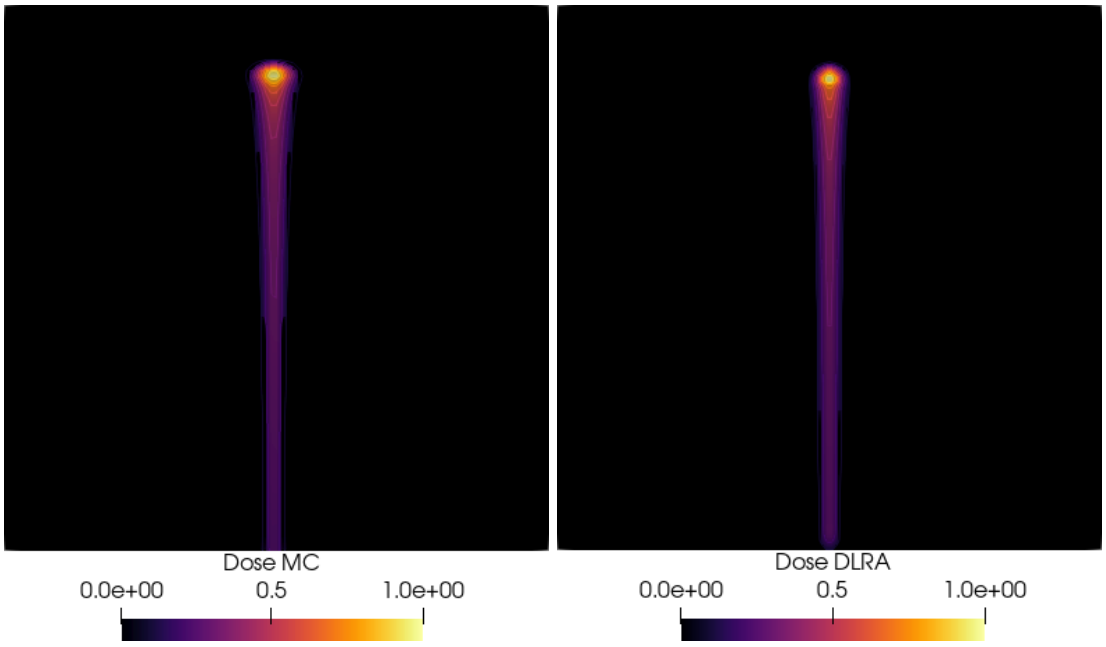
Figure 7.7: First four dominant directional modes with fixed rank integrator.

7.10 Results: Proton transport

Due to the structure of electron dose deposition, electron beams are only used for a small fraction of treatment cases, where the tumor is close to the surface of the skin. Therefore, we now consider the more widely applicable proton transport equation as described in section 4.4. Initially, we only include nuclear scattering cross sections from [13] as well as the stopping power from [17] and consider elastic Coulomb scattering as described by Rutherford without corrections for Multiple Coulomb Scattering (eq. (4.12)).

7.10.1 Cross-validation in 2D

First, we consider a $7.25 \times 7.25\text{cm}^2$ 2D spatial domain with a Gaussian beam directed forward, i.e. $\Omega_{1,*} = 1$. We set $E_{max} = 90$ MeV, $\sigma_x = \sigma_y = 0.1\text{cm}, \sigma_E = 0.01, \sigma_{\Omega_{1,*}=0.01}$ and use a rank of $\mathcal{R} = 5$. Figure 7.8 compares the dose distributions between the DLRA and a MC dose engine for computations in a water phantom ($\rho = 1$ everywhere). The dynamical low-rank approximation is able to capture the beam width, dose magnitude and particle range correctly. Figure 7.9 further shows that the Monte Carlo solution is in very good agreement with the DLRA on the central beam axis, both for the homogeneous water phantom and when inserting a box with higher density $\rho_{\text{bone}} = 1.85$ g/cm^3 at 3.625 cm depth behind the beam entry. Some deviations from the MC reference can however be observed in the lateral dose spread near the Bragg peak. Here, the MC solution fans out due to the increased scattering of slow protons. The lack of this effect in the DLRA solution can most likely be attributed to the simplified physics model, which relies on coarse approximations and merely includes a subset of all particle interactions.



(a) MC (b) DLRA

Figure 7.8: Dose computed using (a) Monte Carlo and (b) DLRA for a beam in a water box.

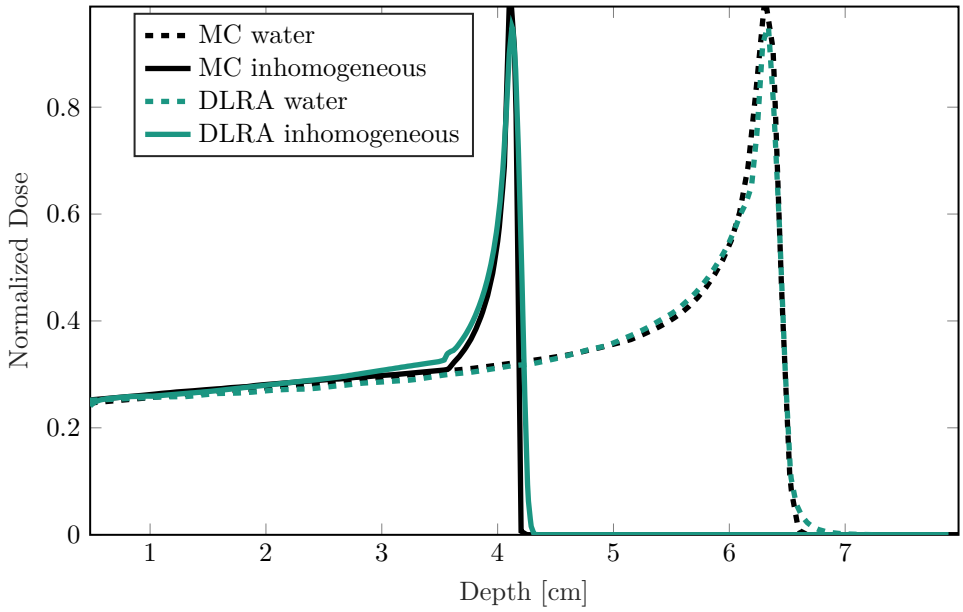


Figure 7.9: Central beam axis cut through dose computed using Monte Carlo and DLRA for a beam in a water box and a heterogeneous target (water box with a bone insert).

The runtime of the DLRA approach (here still without GPUs) is reduced significantly from 26707s to 653s compared to a full deterministic P_n solver. Note, that for protons we require a much lower rank than in our previous electron dose calculations. Here, we observe that ranks $\mathcal{R} \leq 5$ are usually sufficient. A possible explanation for this behavior is the fact that protons experience much less scattering during the build-up of the beam. Intuitively, advection appears dominant until shortly before the Bragg peak, where it is directly superseded by diffusion. Thus both effects, that are individually low rank, are more disjoint than during electron transport.

7.10.2 3D extension

Lastly, we consider a three-dimensional water phantom and compare to results from the Monte Carlo dose engine TOPAS. The Gaussian beam is defined analogously to the previous section with $\sigma_z = \sigma_x = \sigma_y = 0.1$ and we use rank $\mathcal{R} = 3$. The physical model is now extended to also include the range straggling approximation described in section 4.1. Figure 7.10 shows a 3D plot of the delivered dose and fig. 7.11 the integrated depth dose in comparison to the Monte Carlo method. We see that DLRA agrees well with the Monte Carlo reference in terms of the overall structure and range. The DLRA however produces an overly sharp Bragg peak compared to the dose plateau during build-up. This effect can stem from the lack of nuclear scattering in the cross sections used for the dynamical low-rank approximation [25]. In order to verify this suspicion, fig. 7.11 also shows the results of a Monte Carlo simulation with only electromagnetic (Coulomb) interactions. Here, the integrated depth dose curve indeed agrees better with our DLRA results and now predicts only very slightly higher values during the dose build-up. Note, however, that since we are looking at integrated depth dose curves, the observed effects or the remaining difference could also stem from lateral deviations. Further investigations, e.g. into the approximation error introduced by using the Rutherford formula instead of a multiple Coulomb scattering model, are necessary for more certainty in this regard.

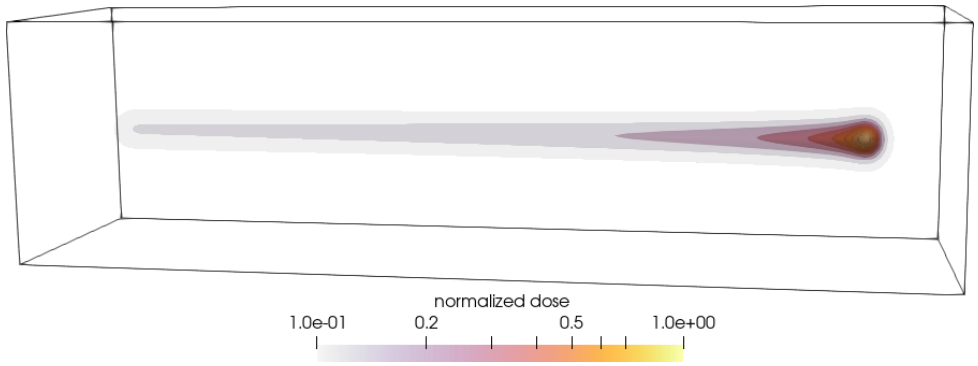


Figure 7.10: 3D plot of proton dose distribution computed with DLRA.

The run time of the DLRA with rank 3, on a $(100 \times 100 \times 400)$ cell spatial domain with 37^2 moments for the P_n angular discretization is 4 minutes and 13 seconds, which is comparable to a Monte Carlo simulation with 10^5 particles (6 minutes 30 seconds) and significantly less than MC with 10^6 particles (60 minutes 9 seconds).

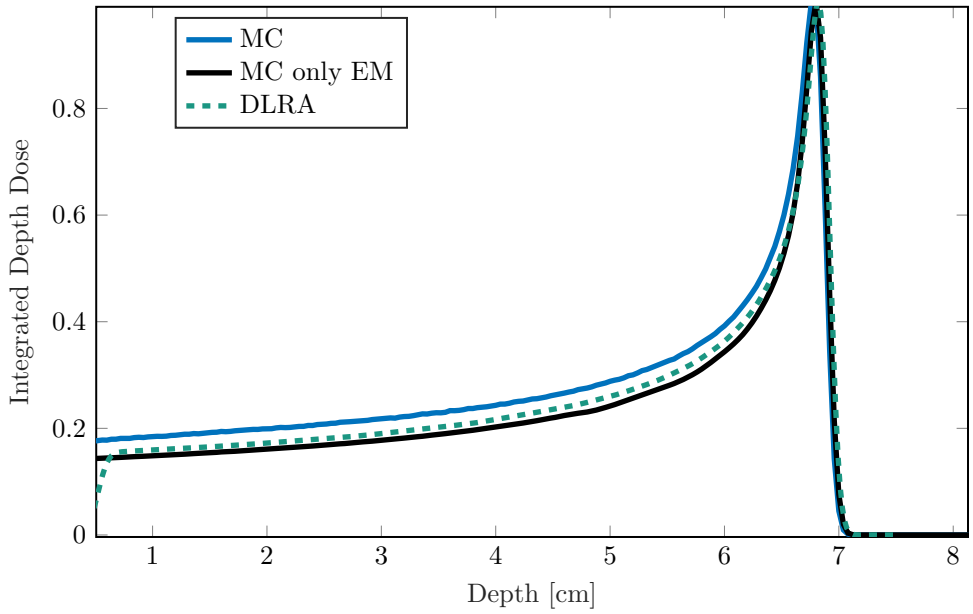


Figure 7.11: Comparison of MC with complete physical model and only electro-magnetic interactions vs. DLRA integrated depth dose for a beam in a 3D water box.

Note, that professional Monte Carlo dose calculations are highly optimized and parallelized. Therefore, despite their slow convergence, the run times remain relatively low when computing in parallel. During treatment planning, typically several million particles have to be simulated for the numerous pencil beams. The achievable speed-up for an IMRT plan, therefore, depends largely on the required rank and discretization when considering a weighted sum of Gaussian beams. The reduced storage and energy consumption however serve as additional arguments for the use of DLRA. Our results further indicate that the accuracy of DLRA dose computations could be improved by incorporating more exact physical models including nuclear scattering.

CHAPTER 8

Results: Comparison of the numerical solution methods

In this last chapter of part II, we want to directly compare the previously discussed deterministic methods with a reference Monte Carlo code for a simple radiation therapy test case. Since the Monte Carlo method is prevalent in the medical physics community, while the deterministic Boltzmann solvers as well as improvements like the DLRA are almost exclusively used in more mathematically focused research, these are seldomly juxtaposed.

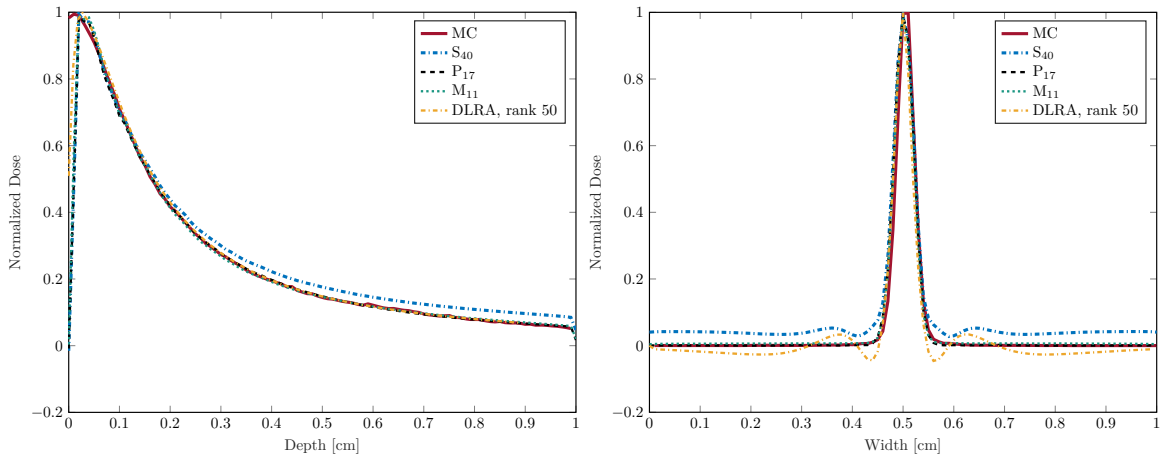
We consider a simple Gaussian electron beam in a 2D $1 \times 1\text{cm}^2$ homogeneous water box ($\rho \equiv 1$). We model the beam as

$$\begin{aligned}\psi_{\text{in}}(E, \mathbf{r}, \mathbf{\Omega}) = 10^5 \cdot \exp(-(\Omega_{1,*} - \Omega_1)^2/\sigma_{\Omega_1}) \cdot \exp(-(E_{\max} - E)^2/\sigma_E) \\ \cdot \exp(-(r_x^* - r_x)^2/\sigma_x) \cdot \exp(-(r_y^* - r_y)^2/\sigma_y) \ ,\end{aligned}$$

where $E_{max} = 5$ MeV, $\Omega_{1,*} = 1$, $\sigma_x = \sigma_y = 0.01\text{cm}, \sigma_E = 0.01, \sigma_{\Omega_{1,*}} = 0.1$. The dose is computed using the KiT-RT implementations of S_{40}, P_{17} and M_{11} solvers as well as the dynamical low-rank approximation using S_N for the uncollided particles and P_{17} for the at least once collided particles with fixed rank $r = 50$. All methods are compared to a Monte Carlo reference computed using TOPAS MC [175] with 10^6 simulated particle histories.

Figure 8.1 compares the dose solutions along a lateral and longitudinal cut. Along the central beam axis, the P_{17}, M_{11} and dynamical low-rank solver agree well with the Monte Carlo reference, while the S_{40} solver increasingly overestimates the dose with growing depth. In the lateral cut, all methods are able to correctly predict the beam width. However, the S_{40} and dynamical low-rank method exhibit slight oscillations to the side of the beam. Further, the S_{40} method again overestimates the solution in the low-dose regions. The parallel between S_{40} and DLR is likely due to the use of the S_N method for the uncollided particle equation in the dynamical low-rank approach. An improvement of the DLR

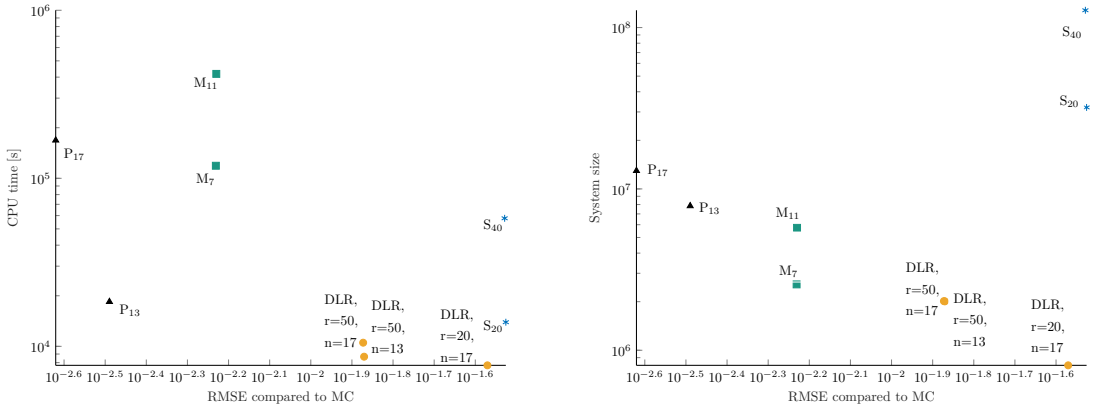
results could thus be achieved by using a method with fewer artifacts, such as an analytical ray-tracing or method of characteristics approach.



(a) (b)

Figure 8.1: (a) Vertical (at $x = 0.1$cm) and (b) horizontal (at $y = 0.5$cm) cross section through the normalized dose in a waterbox. Comparison of the KiT-RT S_{40}, P_{17}, M_{11} solver, a DLRA solver with rank 50 based on P_{17} and a reference Monte Carlo solution with 10^6 simulated particles.

Figure 8.2 shows the trade-off between accuracy in terms of the root mean square error (RMSE) compared to the Monte Carlo reference and computation time or memory consumption.



(a) (b)

Figure 8.2: Comparison of different KiT-RT S_N, P_N and M_N solvers as well as DLRA solvers with different ranks based on $P_{13/17}$ in terms of (a) accuracy (measured in root mean squared error (RMSE) vs. CPU time and (b) accuracy vs. system size/required memory.

Here, it becomes apparent, that all methods lie within a similar range of accuracy. The P_N and M_N solvers show the highest agreement with the reference, however, computation time is relatively high for the M_N method due to the need to solve an optimization problem in each solver iteration. While the

dynamical low-rank method shows slightly higher RMSE values than the P_N and M_N method, both run-time and system size are reduced. As chapter 7 has demonstrated, this complexity reduction is significantly higher for proton dose computations, where a much lower rank of 3-5 is sufficient. Further, due to the structure of the DLRA complexity formulas, the effect can be expected to be larger in higher spatial dimensions.

Overall, deterministic dose calculations are an alternative to Monte Carlo in terms of accuracy. However, standard S_N, P_N and M_N solvers are typically not able to compete with the speed and memory requirements of an optimized, highly parallelized MC code. The dynamical low-rank approximation is a promising approach to increase the efficiency of deterministic dose calculation methods, especially for proton therapy.

CHAPTER 9

Discussion

The preceding chapters defined a mathematical model of radiation transport and solved the corresponding system of equations in order to obtain the deposited dose. We recalled the most common dose calculation methods used in medical physics: The Monte Carlo method and the pencil beam algorithm. While the Monte Carlo method is seen as the gold standard in terms of accuracy, it is computationally expensive. The pencil beam algorithm on the other hand is a heavy approximiation, relying on simplified assumptions in handling density and medium heterogeneities. For this reason, we consider the class of Boltzmann solvers as a deterministic alternative with accuracy comparable to Monte Carlo transport. While numerical methods focused on solving the discretized transport equation are widely researched in mathematical transport theory, they are far less common in the application field of medical physics.

KiT-RT

In chapter 6, we first present a collection of deterministic transport solvers for radiation therapy applications. The open-source code framework aims at making this class of dose calculation methods accessible for medical physics researchers as well as allowing applied mathematicians to easily implement and test their methods in a realistic and relevant setting. The implemented S_N, P_N and M_N solvers agree well with results obtained with conventional radiation therapy codes. Due to the use of polymorphism, we are able to guarantee a straightforward extension to further numerical methods, which facilitates the investigation of novel radiation therapy solvers and their comparison to conventional methods. However, the framework so far does not include attempts to lower the computational costs associated with high-dimensional dose calculations. Further, the solvers are limited to a 2D spatial domain and electron transport. Future work could incorporate the methodology from chapter 7 to reduce memory and run-time and extend the implementation to 3D as well as proton or photon therapy.

Dynamical low-rank approximation

In chapter 7, we proposed a dynamical low-rank approximation for use in radiation therapy. Instead of computing the full solution, dynamical low-rank approximation evolves a low-rank factorization of the solution in time, thereby significantly reducing computational times and memory requirements. The method can be understood as a Galerkin method, which automatically and dynamically picks basis functions to capture relevant information of the solution. Through rank adaptivity, the rank of the solution approximation (that is, the number of basis functions) decreases or increases in time according to the solution complexity.

Furthermore, a collided-uncollided split is used to evolve only collided particles with a dynamical low-rank approximation. Thereby, we potentially reduce the required rank while facilitating the implementation of boundary conditions. This approach can be extended to L-collided splits, which can be interpreted as writing the solution as a telescoping sum. By applying DLRA to each update in this sum, we can decrease the solution complexity in a multilevel fashion. Additionally, we propose an efficient implicit time discretization for scattering terms, which allows for increased time step sizes while not requiring the inversion of matrices. The proposed method is shown to be L^2-stable under a CFL restriction which only depends on the streaming term. We apply the proposed method to electron and proton dose calculations in 2D and 3D spatial domains. Here, we observe good agreements with a Monte Carlo reference solution. While run times could already be reduced by more than an order of magnitude in a simple implementation for electrons, the approach is shown to be especially effective for proton dose calculations. Due to the structure of proton beams, extremely low ranks are sufficient. In combination with a parallelized GPU implementation, this leads to significantly lower memory requirements as well as run-times which are 1-2 orders of magnitude lower than in the electron case, even in higher spatial dimensions. To further improve the accuracy of dose calculations, more exact physical models are required, which explicitly consider multiple Coulomb scattering and include nuclear scattering or range/energy straggling. Further, an S_N method with a directed quadrature set was used for the uncollided particles. Here, an analytical approach such as ray-tracing could be used to potentially improve efficiency and reduce the sensitivity to numerical artifacts.

In future work, we aim at using the proposed forward method to facilitate optimization and uncertainty quantification in radiation therapy. Our results promise an approach that is efficient enough for use in dose optimization, while still taking into account the most relevant physical interactions.

Preview

In the next part of this thesis, we will introduce uncertainties to the transport equation and explore efficient ways of modeling as well as propagating them through dose calculations. With view to the observations made in this chapter, we will focus on two dose calculation methods: First, we will consider how uncertainty quantification can be reconciled with the expensive gold standard MC method. Second, we will explore the possibility to include uncertainty quantification directly in the novel dynamical low-rank approximation for radiation therapy.

Part III

Uncertainty quantification

Uncertainties can significantly affect the dose delivery in radiation therapy. This can hinder plan fulfillment, reducing treatment effectiveness and causing unwanted side effects. In this part, we therefore introduce uncertain parameters to the transport equation and discuss how to mathematically model as well as efficiently propagate uncertainties. We introduce novel approaches to incorporate higher-dimensional or time-dependent uncertainties into a multivariate Gaussian phase space model. Further, two novel uncertainty propagation methods are discussed: 1. For Monte Carlo dose calculations based on importance sampling and the reuse of particle trajectories and 2. for deterministic dose calculations based on the dynamical low-rank approximation for tensor differential equations.

CHAPTER 10

Sources of uncertainty in radiation therapy

In application sciences such as radiation therapy, uncertainties play a key role in the actual fulfillment of a computed plan. Due to the structure of its dose delivery curve, proton therapy is especially sensitive to uncertainties and a displacement of the dose peak into a risk volume can have dire consequences [132]. For this reason, we are interested in quantifying the effect of uncertainties on the delivered dose and the consequent plan fulfillment.

To this end, we introduce an uncertain parameter to the problem eq. (4.1), which we assume to be a vector of random variables $\boldsymbol{\Delta} : \Omega_\Delta \to \mathbb{R}^d$ with probability density function (pdf) $p_\Delta : \mathbb{R}^d \to \mathbb{R}_+$. The corresponding uncertain linear Boltzmann equation takes the following form.

$$\Omega \cdot \nabla\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E, \boldsymbol{\Delta}) + \Sigma_t(\boldsymbol{r}, E, \boldsymbol{\Delta})\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E, \boldsymbol{\Delta}) = Q^{Sca}(\boldsymbol{r}, \boldsymbol{\Omega}, E, \boldsymbol{\Delta}) + S(\boldsymbol{r}, \boldsymbol{\Omega}, E, \boldsymbol{\Delta}) \,, \tag{10.1}$$

We want to determine uncertainty statistics such as the expected value and variance of the dose estimate with respect to these uncertainties. First, however, let's consider which uncertainties are prevalent in radiation therapy and how they can be modeled mathematically.

10.1 Spatial uncertainties

Among the most important sources of uncertainties in particle therapy are spatial displacements of the patient anatomy relative to the irradiating beam [169, 130, 174, 132, 133]. In most cases, treatment plans are optimized with respect to one initial pre-treatment CT. Afterward, the patient position as well as the anatomy are assumed to be fixed throughout the complete treatment. This can lead to two types of uncertainties: So-called *inter-fractional* uncertainties are caused for example by errors in the set-up or positioning of the patient at the beginning of each the treatment or *fraction* (treatment session). *Intra-fractional* uncertainties on the other hand occur during the irradiation and can be caused by patient

movements, organ motion or respiratory motion. Since we only consider single fractions, inter- and intrafractional uncertainties coincide. Due to the varying densities a beam passes through depending on its spatial position, spatial uncertainties do not lead to a static shift, but can also affect the particle range and distort the structure of the dose distribution.

Intrafractional uncertainties can further lead to *interplay* effects when the application pattern of pencil beams in IMRT and patient or organ motion interact.

10.2 (Calculational) Range uncertainties

The second type of uncertainty, which is important for dose calculations, is caused by a variety of factors concerning the patient density, ranging from CT conversion errors to imaging artifacts or changes in the patient geometry [217, 166, 132, 145].
Here, we will focus mostly on so-called *calculational* range uncertainties. The tissue density which is required for dose calculations cannot be measured directly. For this reason, the CT image is converted into density values and used to derive the material stopping power [see e.g. 188, 189]. However, this conversion is not straightforward and often subject to unknown or erroneous parameters. In [132], it is estimated that an uncertainty of $\pm 3\%$ has to be anticipated in the density conversion. Other sources cite even higher deviations of $2.7 - 4.6\%$ [166].

10.3 Further sources of uncertainty

Apart from set-up and range errors, there are numerous additional sources of uncertainty including the delineation of volumes of interest [e.g. 156, 218], organ motion [e.g. 129], tumor shrinkage or growth [e.g. 146, 128] or the biological tumor response [e.g. 52]. The subdivision of the treatment into fractions, spread over the space of several weeks, can also cause different random and systematic errors [122, 136]. These types of errors require tailored models and propagation methods, which will not be discussed in depth within this thesis. The models for spatial uncertainties discussed in section 11.4 can be used to approximate target or organ motion in simple geometries. Further investigations would however be necessary to assess the magnitude of the approximation error introduced by the simplifications. Chapter 15 will also address the direct incorporation of time-dependent CT scans, which can reflect anatomical deformations happening in the time frame of the recorded data.

CHAPTER 11

Uncertainty modeling

The first challenge in uncertainty quantification is deriving an adequate mathematical model of the observed uncertainties. Adequate in this case entails both that the form of the model is suitable for an efficient use with common uncertainty propagation methods and that the description is as close to reality as possible. While the focus in research often lies more on uncertainty propagation methods, the uncertainty model is as much if not more responsible for the quality of the uncertainty analysis. An unrealistic or faulty model will never yield helpful insights, independent of the method used to generate them.

In the following, we will first review the standard uncertainty assumptions used in radiation therapy. Subsequently, we will develop models which are able to reflect more complex and realistic uncertainties, all while maintaining their usability within the propagation methods introduced in chapters 12 to 15. Here, we will focus on the two main sources of uncertainty introduced in chapter 10, i.e. set-up and range errors, many of the concepts could however also be adapted to different types of uncertainty using suitable probability distribution functions and parameters. Note that in the following we will adopt a stochastic view of particle dynamics as introduced in section 5.1. Here, we see particles as following a random walk given by a black box according to their random initial phase space parameters $\boldsymbol{Z} = (\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)$. We will investigate how uncertainties affect the phase space distribution and derive probabilistic representations of the phase space under uncertainty. In terms of a numerical view of particle dynamics, the phase space distributions define the boundary conditions in the transport equation.

11.1 Modeling set-up uncertainties

Set-up uncertainties correspond to a shift of the patient position or equivalently the positions of primary particles relative to the patient. They can be modeled by random shifts in the initial particle positions

$\boldsymbol{r}_0$. The initial phase space of the deterministic solver $\boldsymbol{Z} = (\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)$ is thus affected by the error additively, i.e., the input $\boldsymbol{Z(\Delta)}$ under uncertainty is

$$\boldsymbol{Z(\Delta)} = \boldsymbol{Z} + \boldsymbol{\Delta}\ , \text{where } \boldsymbol{Z} = (\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0) \text{ and } \boldsymbol{\Delta} = (\boldsymbol{\Delta_r}, \boldsymbol{0}, 0). \tag{11.1}$$

We follow the common assumption, that errors, as well as input parameters, are normally distributed for each pencil beam $b = 1, .., B$ [233, 217, 174, 10, 61, 226]:

$$p_\Delta = \mathcal{N}(\boldsymbol{\mu}_\Delta, \boldsymbol{\Sigma}_\Delta)\ ,\ S_0^b = \mathcal{N}(\boldsymbol{\mu}_Z^b, \boldsymbol{\Lambda}^b)\ . \tag{11.2}$$

In this case we can define the density $\mathcal{S}(\boldsymbol{Z(\Delta)})$ of the input including the uncertain factors through convolution of p_Δ and $S_0 = \sum_{b=1}^B w_b S_0^b$ (comp. eq. (5.8)):

$$\mathcal{S} := \sum_{b=1}^{B} w_b \cdot \mathcal{N}(\boldsymbol{\mu}_Z^b + \boldsymbol{\mu}_\Delta, \boldsymbol{\Lambda}^b + \boldsymbol{\Sigma}_\Delta)\ . \tag{11.3}$$

An individual error realization $\boldsymbol{\delta}_k \sim p_\Delta$ then formally just corresponds to a shift of the original primary positions, which now follow the distribution

$$S_k := \mathbb{P}(\boldsymbol{Z} + \boldsymbol{\Delta} \mid \boldsymbol{\Delta} = \boldsymbol{\delta}_k) = \sum_{b=1}^{B} w_b \cdot \mathcal{N}(\boldsymbol{\mu}_Z^b + \boldsymbol{\delta}_k, \boldsymbol{\Lambda}^b)\ , \tag{11.4}$$

corresponding to the nominal distribution shifted by $\boldsymbol{\delta}_k$.

Remark 11.1 (Standard assumptions)**.** *The most frequently used uncertainty model for set-up errors assumes one global error* $\boldsymbol{\Delta_r} = (\Delta_{r_x}, \Delta_{r_y}, \Delta_{r_z}) \sim \mathcal{N}(\boldsymbol{\mu}_{\Delta_r}, \boldsymbol{\Sigma}_{\Delta_r})$*. This would correspond to movements or mispositioning of the patient which affect the complete body or at least the area of interest around the tumor.*

11.2 Modeling range uncertainties

In contrast to set-up uncertainties, which directly affect the phase space parameters $(\boldsymbol{r}, \boldsymbol{\Omega}, E)$, range uncertainties modify the density values $\rho(\boldsymbol{r})$. They can be caused by different effects described in chapter 10, here we however only consider *calculational* range uncertainties, which occur when determining the tissue density or relative electron densities/stopping powers from the voxel values in a CT image. These conversion errors are typically assumed to manifest as the same percentage-wise deviation of densities in each voxel:

$$\rho(\boldsymbol{r}, \Delta_\rho) = (1 + \Delta_\rho) \cdot \rho(\boldsymbol{r}),$$

where $\Delta_\rho \in [1, -1]$ represents the relative deviation.

Using common uncertainty propagation methods, it would be sufficient to again assume for example a Gaussian distribution of Δ_ρ with a standard deviation of $\pm 3\%$ [132], to compute the required uncertainty statistics. However, for the methods which will be presented in chapters 13 and 14, it is advantageous to

model uncertainties in terms of the phase space parameters, such that we can derive a joint distribution of $\boldsymbol{Z}(\boldsymbol{\Delta})$.

In the following, we therefore derive an approximate model of range uncertainties using the initial particle energy E_0. We exploit that the largest dose uncertainty is induced near the range of a beam [25], although the uncertain density variation affects the whole trajectory. The deviation in density is thus equated to a deviation in range. Range can be expressed in terms of the initial energy of particles, using the Bragg-Kleemann rule

$$R = \alpha \cdot E_0^p \,, \tag{11.5}$$

where R is the range, E_0 is the initial energy and α and p are application-specific parameters. For the case of the slow-down of therapeutic protons in water, values of $\alpha = 0.022$ mm/MeV^p and $p = 1.77$ can be chosen [216].

The initial energy spectrum of a scanned pencil beam at the exit of the nozzle can be approximately represented by a Gaussian [25, 100, 212, 201]. We can use this connection to model range uncertainties through random variations of the initial energy [compare treatment of range straggling in 170]. To achieve this, we use a first order Taylor expansion. In statistics, this approach is also known as the *delta method* [e.g. 167].

Let's assume range uncertainties are normally distributed and range is otherwise deterministic, i.e. $\Delta_R \sim \mathcal{N}(0, \sigma_R^2)$ and $R(\Delta_R) \sim \mathcal{N}(\mu_R, \sigma_R^2)$ [comp. 132, 238]. Note, that when considering range straggling, the particle range itself is also considered to be random [18, 24, 152]. However, since this is typically modeled using a normal distribution [19], the derivation of the range distribution under uncertainty is straightforward. For the random variable R with mean μ_R and variance σ_R^2, the mean and variance of a differentiable function $g(R)$ can be approximated using a Taylor approximation around $\mu_R = \mathbb{E}[R]$. A first order Taylor expansion gives:

$$g(R) \approx g(\mu_R) + g'(\mu_R)(R - \mu_R) \tag{11.6}$$

$$\begin{aligned} \mathbb{E}[g(R)] &\approx g(\mu_R) + g'(\mu_R)\mathbb{E}[R - \mu_R] \\ &= g(\mu_R) \end{aligned} \tag{11.7}$$

$$\begin{aligned} Var(g(R)) &\approx E\left[\left(g(R) - g(\mu_R)\right)^2\right] \\ &\approx E\left[\left(g'(R)(R - \mu_R)\right)^2\right] \\ &= g'(\mu_R)^2 Var(R). \end{aligned} \tag{11.8}$$

Now we can determine the parameters μ_{E_0}, σ_{E_0} of the energy distribution due to range uncertainties by choosing $g(R) := (\frac{1}{\alpha}R)^{\frac{1}{p}} = E_0$ according to the Bragg-Kleemann rule (see eq. (11.5)).

$$\mu_{E_0} = \mathbb{E}[g(R)] \approx g(\mu_R) = (\tfrac{1}{\alpha}\mu_R)^{\frac{1}{p}} \tag{11.9}$$

$$\begin{aligned} \sigma_{E_0}^2 &= Var(g(R)) \approx g'(\mu_R)^2 Var(R) \\ &= \left(\tfrac{1}{p\alpha}(\mu_R \tfrac{1}{\alpha})^{\frac{1}{p}-1}\right)^2 \sigma_R^2 \end{aligned} \tag{11.10}$$

We then further assume, that the energy E_0 is normally distributed and set the parameters of this normal distribution to μ_{E_0} and $\sigma^2_{E_0}$. Thus, the randomness in range is approximated through an energy distribution $E_0 \sim \mathcal{N}(\mu_{E_0}, \sigma^2_{E_0})$ and we can derive joint phase space and uncertainty distributions analogous to section 11.1. Note, that considering the functional relation given by the Bragg-Kleemann rule and the fact that R is normally distributed the assumption that E_0 is also normally distributed is a coarse approximation. We will investigate the associated approximation error numerically in chapter 14 and section 11.5.

11.3 Multivariate models

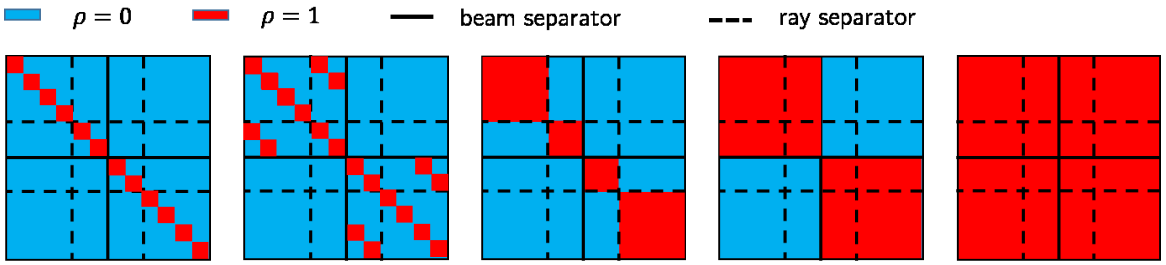
In radiation therapy, often numerous beamlets with different energy levels, positions and incoming angles are used to achieve better coverage of the tumor. Recall that the term *beamlet* refers to a fine proton beam with specific energy directed at a certain position in the tumor. An ensemble of beamlets coming from the same geometrical set-up of nozzle and patient is then referred to as *beam*, a group of beamlets with the same lateral positioning within a beam as *ray* and a group of beamlets with the same energy within a beam as *energy level*. In this case, contrary to remark 11.1 it may be more reasonable to assume beamlet-specific correlation patterns, e.g., depending on the differences in time and position of application of the individual beamlets. Additionally, treatments are usually administered in several sittings or fractions, which can also affect how errors are correlated.

We therefore now differentiate between uncertainties in each beamlet and define the error in beamlet b as $\delta^b = (\delta^b_x, \delta^b_y, \delta^b_z) \sim \mathcal{N}(\mu^b_\delta, \Sigma^b)$ and their covariance by

$$C = \begin{pmatrix} & \Sigma^1 & & \rho^{12}_{xx} & \rho^{12}_{xy} & \rho^{12}_{xz} & & \rho^{1B}_{xx} & \rho^{1B}_{xy} & \rho^{1B}_{xz} \\ & & & \rho^{12}_{yx} & \rho^{12}_{yy} & \rho^{12}_{yz} & \cdots & \rho^{1B}_{yx} & \rho^{1B}_{yy} & \rho^{1B}_{yz} \\ & & & \rho^{12}_{zx} & \rho^{12}_{zy} & \rho^{12}_{zz} & & \rho^{1B}_{zx} & \rho^{1B}_{zy} & \rho^{1B}_{zz} \\ \rho^{21}_{xx} & \rho^{21}_{xy} & \rho^{21}_{xz} & & & & & & & \\ \rho^{21}_{yx} & \rho^{21}_{yy} & \rho^{21}_{yz} & & \Sigma^2 & & & & & \\ \rho^{21}_{zx} & \rho^{21}_{zy} & \rho^{21}_{zz} & & & & & & & \\ & \vdots & & & & & \ddots & & \vdots & \\ \rho^{B1}_{xx} & \rho^{B1}_{xy} & \rho^{B1}_{xz} & & & & & & & \\ \rho^{B1}_{yx} & \rho^{B1}_{yy} & \rho^{B1}_{yz} & & \cdots & & & & \Sigma^B & \\ \rho^{B1}_{zx} & \rho^{B1}_{zy} & \rho^{B1}_{zz} & & & & & & & \end{pmatrix}.$$

Here ρ^{ab}_{xy} is the covariance between the set-up errors of beamlet a in dimension x and beamlet b in dimension y. It is then possible to obtain error realizations for different choices of the covariance matrix, using a known probability distribution or, alternatively, a Copula with mean μ and covariance matrix C.

Simple correlation models, assuming full correlation between the beamlets belonging to the same beam or ray, have been investigated by different authors [10, 226, 178]. Figure 11.1 illustrates the covariance matrix C for five examples of such simple correlation models.



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

Figure 11.1: Correlation matrices for different assumptions. Rows and columns of the matrices correspond to the individual pencil beams, and beam and ray separators indicate sections of pencil beams with the same irradiation angle and lateral position, respectively. (a) No correlation between beamlets, (b) correlation of energy levels within one beam, (c) ray-wise correlation, (d) all pencil beams with the same lateral position, i.e. hitting the same material are fully correlated and (e) beamwise correlation, pencil beams with the same irradiation angle are fully correlated.

In case the covariance matrix is singular (perfect correlation between some pencil beams), the dimension of the uncertain vector can be reduced and one joint error can be sampled for the respective perfectly correlated pencil beams.

As multivariate Gaussian distributions of the initial phase space parameters can be handled very efficiently by the propagation methods proposed in chapters 13 and 14, we want to explore how this framework can be used to model more complex correlation patterns. This can, for example, be used to incorporate knowledge about the radiation process and even model certain time dependencies causing interplay effects [e.g. 193, 202]. In the following, we demonstrate the derivation of the covariance matrix C for two examples of autocorrelated set-up errors as well as time series data from a simple chest displacement data set and a 4D CT scan.

11.4 Time-dependent models

We assume a multivariate Gaussian error model, where each dimension corresponds to the error in an individual beamlet with a Gaussian marginal distribution. The covariance matrix of this multivariate model can reflect the spatial as well as time-dependent (auto-)correlation patterns. Uncertainty quantification methods such as importance reweighting for MC (chapter 14) or analytical probabilistic modeling (APM) for pencil beam dose calculations (chapter 13) then allow an efficient quantification of the uncertainties directly from the error model.

We use the time-dependency inherent in the beam application process to derive covariance matrices representing intrafraction motion interplaying with the beam application pattern. Each beamlet $b = 1, ..., B$ is therefore associated with a point in time during treatment $t(b)$ at which particles are emitted from the position and with the energy associated with this beamlet. The correlations between uncertainties of each beamlet can then be derived from those between the errors at the associated time points. Previous works have already demonstrated that specific types of intrafraction motion may be modeled as time series

using, e.g., random walks, auto-regressive (ARMA) models or Gaussian processes [86, 50, 83]. Then the required correlations can be determined using the autocovariance function. The process is illustrated in fig. 11.2.

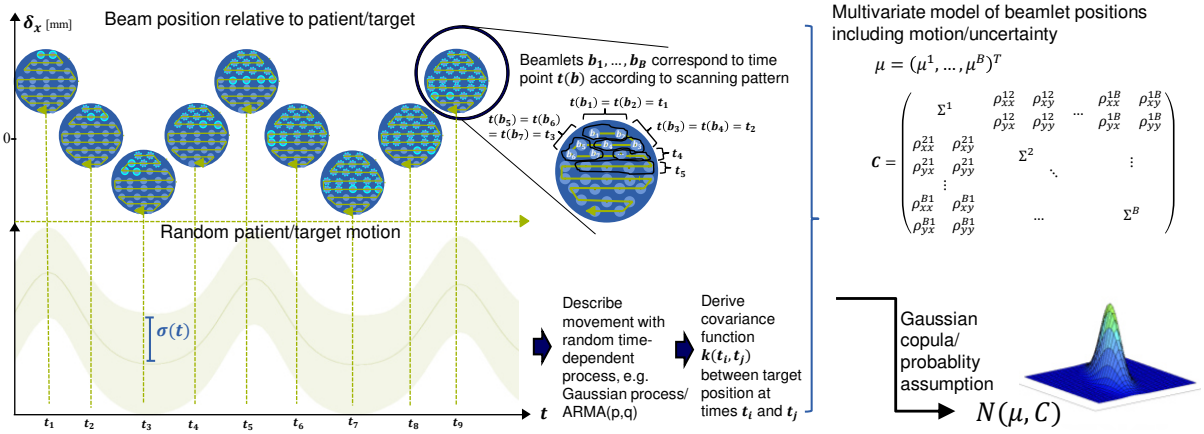


Figure 11.2: A covariance matrix reflecting the correlations of beamlet positions can be derived from the beam scanning pattern and covariances between errors at different time points. Illustration published in [204].

11.4.1 Autocorrelated set-up errors

To demonstrate how the covariance matrix C can be obtained for a given correlation model, we first consider the case of autocorrelated set-up errors, which follow an AR(1) process. While irradiation itself usually only takes a few milliseconds, beamlets are applied sequentially according to their energy level and up to 1-2 seconds can pass between them. Thus, assuming the patient is exhibiting random movement with some probability at all points in time, the patient position corresponding to beamlets in adjacent energy groups should be correlated to a higher degree than those of beamlets with a larger difference in energies. Since first order autoregressive models are sufficient to capture interfractional set-up errors [86], we choose such a model for the random patient movements during treatment.

We thus describe movements by an autoregressive process of order one, meaning that the overall deviation at time t depends on the deviation at time $t-1$ as well as a random variable $U_t \overset{iid}{\sim} \mathcal{N}(\mu_\delta, \Sigma)$, which represents the movement error introduced between $t-1$ and t:

$$\delta_t = \alpha \delta_{t-1} + U_t \ . \tag{11.11}$$

Here, the factor $|\alpha| < 1$ controls how strongly the position at the previous time step affects the current position. It can, for example, be chosen depending on how much time passes between the time steps. In our case, each beam b corresponds to a point in time $t(b)$ according to its energy and we assume that there is no additional correlation between the (x,y,z)-coordinates, i.e.,

$$\Sigma = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \ . \tag{11.12}$$

From this description of the process, it is now possible to obtain the covariance matrix C using the Autocovariance Function (ACVF):

$$\rho_{ab}^{ij} = Cov(\delta_{t(i)}, \delta_{t(j)}) = \begin{cases} \frac{\alpha^{|t(i)-t(j)|}\sigma_x^2}{1-\alpha^2} & , a = b = x \\ \frac{\alpha^{|t(i)-t(j)|}\sigma_y^2}{1-\alpha^2} & , a = b = y \\ \frac{\alpha^{|t(i)-t(j)|}\sigma_z^2}{1-\alpha^2} & , a = b = z \\ 0 & , a \neq b \end{cases} . \tag{11.13}$$

Note that the variance is constant in time and the covariances only depend on the time difference. To obtain a joint probability function we use a Gaussian copula, which for Gaussian marginals reduces to the multivariate normal distribution $\mathcal{N}(\mu_\delta, C)$ with $\mu_\delta = (\mu_\delta^1, \mu_\delta^2, ..., \mu_\delta^B)$ and $C =$

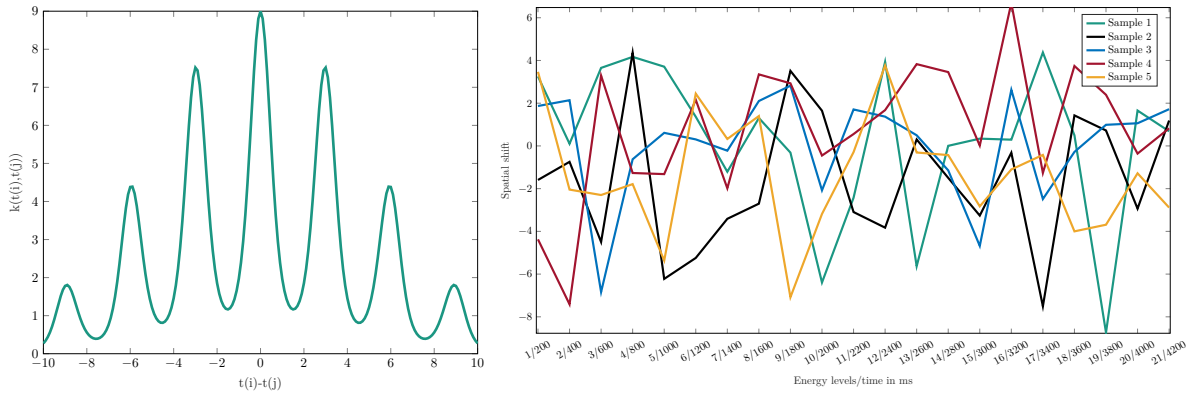
$$\begin{pmatrix} \frac{\sigma_x^2}{1-\alpha^2} & 0 & 0 & \cdots & \frac{\alpha^{|t(B)-t(1)|}\sigma_x^2}{1-\alpha^2} & 0 & 0 \\ 0 & \frac{\sigma_y^2}{1-\alpha^2} & 0 & \cdots & 0 & \frac{\alpha^{|t(B)-t(1)|}\sigma_y^2}{1-\alpha^2} & 0 \\ 0 & 0 & \frac{\sigma_z^2}{1-\alpha^2} & \cdots & 0 & 0 & \frac{\alpha^{|t(B)-t(1)|}\sigma_z^2}{1-\alpha^2} \\ \vdots & & & \ddots & & \vdots & \\ \frac{\alpha^{|t(1)-t(B)|}\sigma_x^2}{1-\alpha^2} & 0 & 0 & \cdots & \frac{\sigma_x^2}{1-\alpha^2} & 0 & 0 \\ 0 & \frac{\alpha^{|t(1)-t(B)|}\sigma_y^2}{1-\alpha^2} & 0 & \cdots & 0 & \frac{\sigma_y^2}{1-\alpha^2} & 0 \\ 0 & 0 & \frac{\alpha^{|t(1)-t(B)|}\sigma_z^2}{1-\alpha^2} & \cdots & 0 & 0 & \frac{\sigma_z^2}{1-\alpha^2} \end{pmatrix} .$$

This AR(1)-correlation model gives an intuitive description of the random movements of a patient over the time of irradiation. However, there are also errors caused by periodic movements such as respiratory motion or heartbeats. Respiratory motion is often modeled using Gaussian processes with an appropriate covariance kernel [50], however, there is also literature exploring the use of a periodic autoregressive moving average (ARMA) model [83]. Both could be incorporated into our uncertainty framework, we however confine ourselves to showing an example for a Gaussian process with a (local) periodic kernel here.

The Gaussian process (GP) implements the concept of data in a time series following a joint multivariate Gaussian distribution, where the covariance kernel defines the dependencies between different data points. As the type of movement we are interested in has a dominant periodic component, which may however vary over time, we choose the following kernel $k(\delta_{t(i)}, \delta_{t(j)})$, proposed similarly by [51],

$$k(\delta_{t(i)}, \delta_{t(j)}) = \sigma^2 \cdot \exp\left(-\frac{2\sin^2(\pi(t(i)-t(j))^2/p)}{l_1^2}\right) \cdot \exp\left(-\frac{(t(i)-t(j))^2}{2l_2^2}\right) . \tag{11.14}$$

Here, σ is the variance of the stationary process, p is the period and l_1 and l_2 are scale parameters. The first exponential term describes an exact periodic repetition, whereas the second term is a local kernel that introduces the variations over time. Figure 11.3 illustrates the covariance kernel as well as



(a) (b)

Figure 11.3: Local periodic covariance kernel with $\sigma = 3$, $p = 3$, $l_1 = 1$ and $l_2 = 5$ (a) and five realizations of the associated GP (b)

five exemplary realizations for a choice of $\sigma = 3$, $p = 3$, $l_1 = 1$ and $l_2 = 5$. The process starts at a random point in the cycle and periodically repeats itself with some random variations. From eq. (11.14) the derivation of the covariance matrix C is straightforward with the entries

$$\rho_{ab}^{ij} = Cov(\delta_{t(i)}, \delta_{t(j)}) = \begin{cases} \sigma_x^2 \cdot \exp(-\frac{2\sin^2(\pi(t(i)-t(j))/p)}{l_1^2}) \cdot \exp(-\frac{(t(i)-t(j))^2)}{2l_2^2}) & , a = b = x \\ \sigma_y^2 \cdot \exp(-\frac{2\sin^2(\pi(t(i)-t(j))/p)}{l_1^2}) \cdot \exp(-\frac{(t(i)-t(j))^2)}{2l_2^2}) & , a = b = y \\ \sigma_z^2 \cdot \exp(-\frac{2\sin^2(\pi(t(i)-t(j))/p)}{l_1^2}) \cdot \exp(-\frac{(t(i)-t(j))^2)}{2l_2^2}) & , a = b = z \\ 0 & , a \neq b \end{cases} . \quad (11.15)$$

The errors for all pencil beams can now be sampled from the multivariate Gaussian $\mathcal{N}(\mu_\delta, C)$ and will depend on the time at which their corresponding energy level is applied.

The two examples above demonstrate, that it is possible to implement complex movement patterns and even to some degree model time dependencies, by changing the way error realizations for the beamlet positions are sampled.

11.4.2 Fitting models to data

We have seen in the previous section that – with some simplifying assumptions – it is possible to model time-dependent uncertainties using the covariance structure of a multivariate Gaussian distribution of the initial phase space. In the following, we want to apply the introduced concepts to realistic data and highlight a possible workflow from patient data to personalized treatment planning.

We consider two sets of observed time series data $\mathcal{T}_O \in \mathbb{R}^{Txd}$ describing patient or tumor movement

within a certain measurement period with T observation times and d-dimensional data points. To derive the covariance between uncertainties at the beam application times $t(b),\ b = 1, ..., B$, we want to fit a time series model with associated autocovariance function or kernel. In the following, we will use the example of a Gaussian process.

Kernel fitting and hyperparameter search

Fitting kernels or Gaussian processes to time series data is a problem for which a wide range of methods and research is available, not least due to the recent popularity of machine learning and the use of Gaussian processes in this context [180, 183, 182].

Here we use *maximum likelihood estimation* to find the set of hyperparameters $\boldsymbol{\theta} \in \mathcal{D}_\theta \subseteq \mathbb{R}^d$, which maximize the probability for observing the given set of data points. This likelihood is given by

$$\begin{aligned} \mathcal{L} = p(\mathcal{T}_O \mid \boldsymbol{\theta}) &= \mathcal{N}(\mathcal{T}_O \mid \boldsymbol{\theta}) \\ &= \mathcal{N}(\mathcal{T}_O \mid \mu, \Sigma^\theta), \end{aligned} \tag{11.16}$$

where $\Sigma^\theta_{ij} = k_\theta(t_i, t_j),\ i, j = 1, ..., T$ is the covariance kernel with hyperparameters $\boldsymbol{\theta}$. Note, that in the following we assume that the function k_θ is known and we merely optimize the choice of hyperparameters. This is a reasonable assumption here since we have some insight into the structure of the respiratory data. However, depending on the application and type of data this decision is not trivial. Numerous works [e.g. 51, 211, 1] therefore investigate approaches for an automated choice of kernel functions. This is however beyond the scope of this thesis.

Given the likelihood function $\mathcal{L}$, the maximum likelihood estimator solves

$$\arg\max_\theta \mathcal{L}(\boldsymbol{\theta}) = \mathcal{N}(\mathcal{T}_O \mid \boldsymbol{\theta}) \tag{11.17}$$

or rather the equivalent problem using the logarithmic likelihood function due to its monotonicity

$$\arg\max_\theta \log\left(\mathcal{N}(\mathcal{T}_O \mid \boldsymbol{\theta})\right) = -\frac{1}{2}(\mathcal{T}_O - \mu)^T {\Sigma^\theta}^{-1}(\mathcal{T}_O - \mu) - \frac{1}{2}\log(|\Sigma^\theta|) - \frac{d}{2}\log(2\pi). \tag{11.18}$$

Setting the partial derivatives for hyper parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_{N_\theta})$ to zero, we arrive at the extrema candidates

$$\frac{\partial}{\partial\theta_i} p = \frac{1}{2}(\mathcal{T}_O - \mu)^T {\Sigma^\theta}^{-1} \frac{\partial \Sigma^\theta}{\partial \theta_i} {\Sigma^\theta}^{-1}(\mathcal{T}_O - \mu) - \frac{1}{2} tr\left({\Sigma^\theta}^{-1}\frac{\partial \Sigma^\theta}{\partial \theta_i}\right) \overset{!}{=} 0. \tag{11.19}$$

Note, that this approach can not only be used to derive an autocovariance kernel for given data, but also constitutes a basis for Bayesian inference, i.e. for observed data $\mathcal{T}_O$ in a time series, one could predict the most likely next observations. This is discussed in more detail in [181]. Future work could explore the possibility of a use for in-line radiotherapy, where the energy and position of the gantry are adapted during the irradiation according to predictions of tumor movement.

Example 1: Fitting to surrogate chest displacement signal

The first dataset we consider has been retrieved from an open source platform [153] and contains one-dimensional chest displacement data of 75 patients for high, medium and low breathing rates, measured

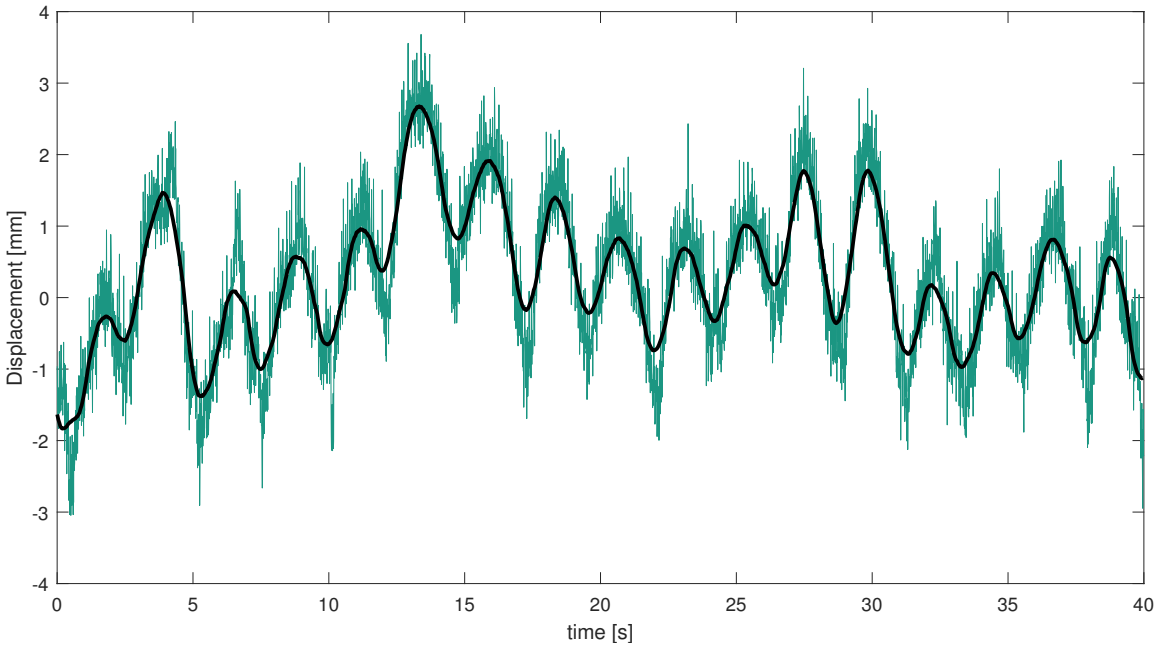


Figure 11.4: Raw (green) and smoothed (black) data for the surrogate signal of a patient's respiratory pattern.

at a rate of $1/256s$ over the course of one minute. In order to directly obtain tumor movement data, invasive procedures or a large number of CT scans (which are associated with further dose to the patient) are necessary. For this reason, surrogate signals of the chest displacement are used in their place. These can be recorded via an external marker block.

We will initially optimize for the hyperparameters σ and l of the squared exponential (SE) kernel [183] as defined by

$$k(t_i, t_j) = \sigma^2 \exp\left(-\frac{(t_i - t_j)^2}{2l^2}\right), \tag{11.20}$$

where σ models the amplitude of the uncertainty and the specific length l the frequency of the variations. Since the original data is recorded at a high frequency and contains a lot of noise, we first smooth the signals using a Gaussian-weighted moving average as illustrated in fig. 11.4.

The optimization of hyperparameters then yields

$$\sigma_{\text{opt.}} = 1.7167 \qquad l_{\text{opt.}} = 0.9146. \tag{11.21}$$

We further apply the same strategy for different kernel types from [183]: the exponential kernel, Matérn kernel, rational quadratic kernel and local periodic kernel. The different kernels as well as their optimal hyperparameter values are summarized in table 11.1. Sampling from the corresponding multivariate Gaussian distribution we can obtain uncertainty patterns that reflect the amplitude and period of the initial data set (see fig. 11.5).

Using the derived covariance kernel with the optimized parameters, the required covariance matrix C describing beamlet correlations can be directly obtained by evaluating the kernel function at the beam

application times $t(b),\ b = 1, ..., B$. Figure 11.6 shows the beamlet covariance kernel and an excerpt from the corresponding covariance matrix for a treatment plan with 1408 beamlets in 21 energy groups.

Table 11.1: Hyperparameter optimization for the surrogate signal from patient n°4, for different kernel types, where $r = \sqrt{(t_i - t_j)^2}$.

Kernel	Function	Optimized Hyperparameters
Exponential	$k(t_i, t_j\ ;\ \sigma, l) = \sigma^2 \exp(-\frac{r}{l})$	$[\sigma = 1.1837, l = 1.6127]$
Matérn$_{3/2}$	$k(t_i, t_j\ ;\ \sigma, l) = \sigma^2 \left(1 + \frac{\sqrt{3}r}{l}\right) \exp\left(-\frac{\sqrt{3}r}{l}\right)$	$[\sigma = 1.1329, l = 0.5972]$
Rational Quadratic	$k(t_i, t_j\ ;\ \sigma, l, \alpha) = \sigma^2 \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}$	$[\sigma = 1.3192, l = 0.4288, \alpha = 0.2317]$
Local periodic	$k(t_i, t_j\ ;\ \sigma, p, l_1, l_2) = \sigma^2 \exp\left(\frac{2\sin^2(\pi(t_i - t_j).^2/p)}{l_1^2}\right) \cdot \exp\left(-\frac{(t_i - t_j)^2}{2l_2^2}\right)$	$[\sigma = 1.0014 p = 0.9967, l_1 = 1.0002, l_2 = 0.9997]$

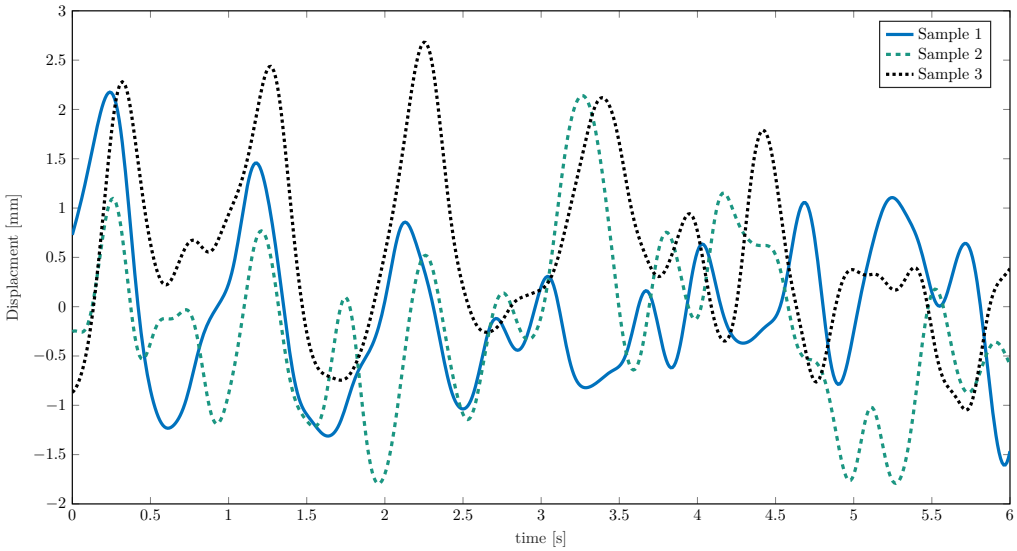
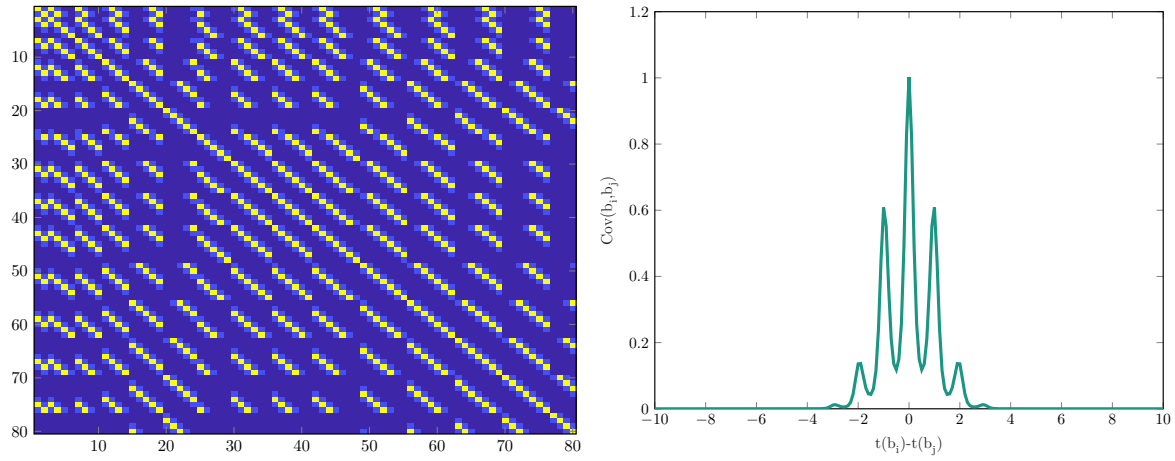


Figure 11.5: 6 second extract of three displacement patterns sampled from the Gaussian process with a local periodic kernel and optimized hyperparameters.

Having demonstrated how to incorporate simple one-dimensional data into our time-dependent uncertainty model, we now proceed to show a proof-of-concept for higher dimensional 4D cone beam CT (CBCT) imaging data.

Example 2: Fitting To Patient's CBCT

In addition to standard three-dimensional CT images, which are obtained at the beginning of each treatment planning process, so called *4D CTs* are sometimes acquired. A 4D CT contains several 3D images of the patient anatomy at different points in time. When considering respiratory motion, images corresponding to the equivalent stages in the breathing cycle are merged to obtain a representative time series. In the following, we use 4D lung CT data [88] from the open-source cancer imaging archive (TCIA) [40], which covers 10 different time points during a breathing cycle (see fig. 11.7). Besides image data, the chosen dataset further includes tumor and organ delineations in the form of contours for each slice



(a) (b)

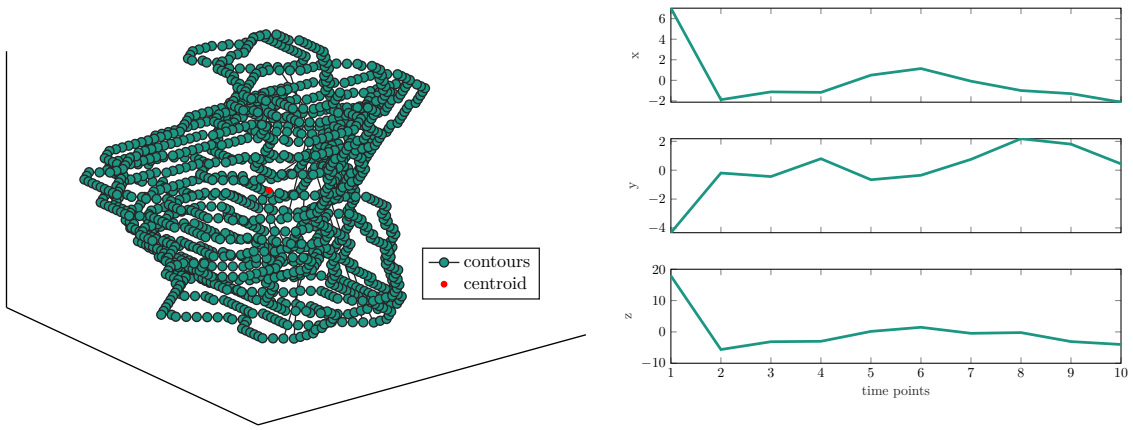
Figure 11.6: (a) Excerpt from the covariance matrix and (b) beamlet covariance function for a local periodic model fit to the chest displacement data set.

Figure 11.7: Slice of 4D CT taken from the first fraction of the breathing pattern.

of the 3D CTs. In order to track the tumor movement, we approximate the position of the tumor by its centroid. Note, that this representation can only approximate positional shifts of the complete tumor structure and disregards information on deformations that might affect tumor shape or size. Given a set of points $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\} \in \mathbb{R}^3$ defining the target contour, the centroid c is given by,

$$\mathbf{c} = \frac{\mathbf{r}_1 + \cdots + \mathbf{r}_n}{n} \tag{11.22}$$

An example of a centroid determined from tumor contours as well as the evolution of the centroid position in x, y and z direction over time is shown in fig. 11.8.



(a) (b)

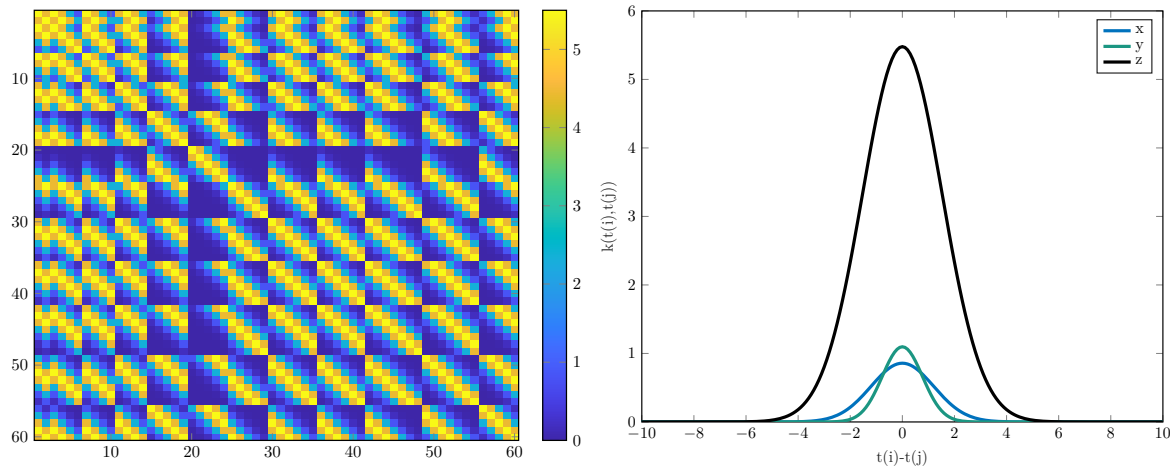
Figure 11.8: (a) Tumor contours and computed centroid (b) displacement of tumor centroid separated into x, y and z coordinates for an exemplary patient CT from [88].

Assuming no correlation between spacial dimensions, we optimize the hyperparameters of a squared exponential kernel for each dimension individually. The optimized parameters are summarized in table 11.2.

Table 11.2: Optimized hyperparameters corresponding to the process depicted in fig. 11.8

Squared Exp. Kernel Hyperparameters	x	y	z
σ	0.9261	1.0476	2.3406
l	1.2009	0.7712	1.5304

A covariance matrix for the beamlets can be derived analogously to the previous sections. For example, using a squared exponential kernel and assuming a time lag of 20ms between each energy level yields the covariance structures presented in fig. 11.9 .



(a) (b)

Figure 11.9: (a) Excerpt from the beamlet covariance matrix (according to z-axis) and (b) kernel function for a squared exponential model fit to the 4D lung CT data set (table 11.2).

11.5 Model validation

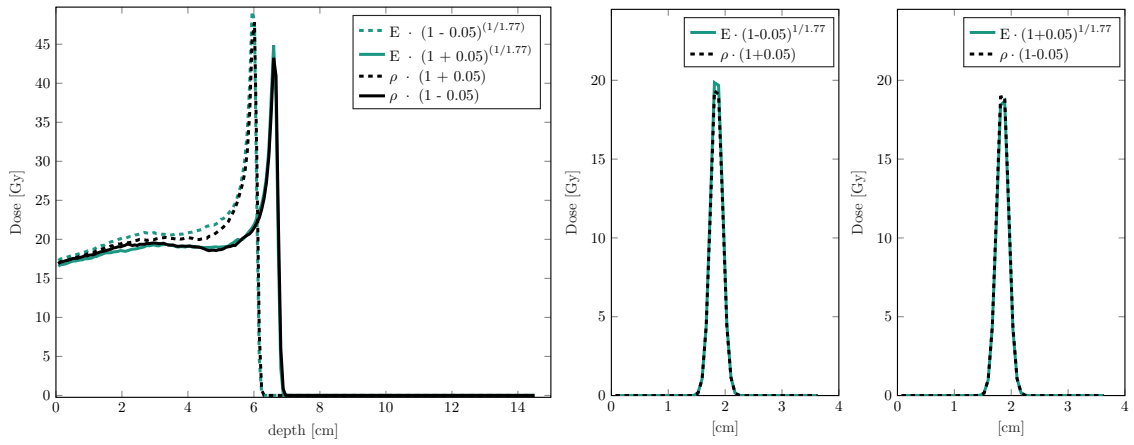
In the previous sections, several approximations have been introduced to model uncertainties for an efficient use in uncertainty propagation methods. In the following, we want to examine the validity of these approximations.

11.5.1 Range-energy approximation

First, we consider the approximation of calculational range uncertainties using the distribution of initial particle energies. To investigate the error introduced by this simplification, we scale the density of a water phantom and a CT scan (downsampled for the sake of simplicity) up and down by ±5% and compare the results to an equivalent scaling of the initial energy according to the model introduced in section 11.2. For the computations in the water phantom, a single pencil beam with 90 MeV is used, while the CT is irradiated with 134.68 MeV.

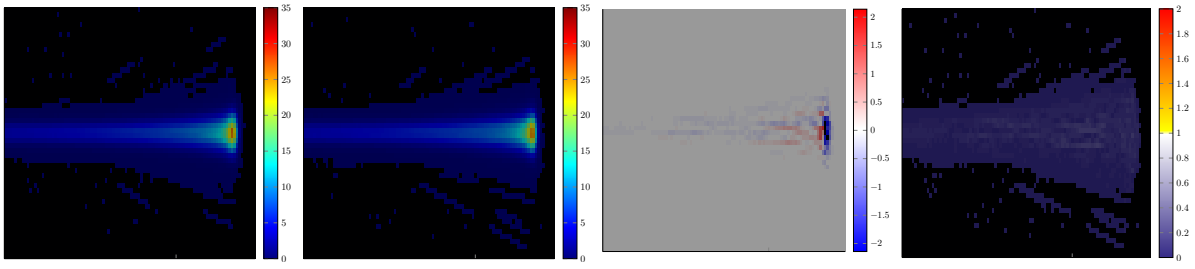
Figures 11.10 and 11.13 show that the approximation is able to accurately reflect the change in the particle range in water as well as the heterogeneous CT scan. It however introduces some deviations during the dose build-up and around the Bragg peak. These deviations do not appear to be amplified by heterogeneities. Further, the agreement according to a $\gamma_{3\,\mathrm{mm}/3\,\%}$ is at least 99.9% (see figs. 11.11 and 11.14). While some minor variations can also be caused by the stochastic nature of the Monte Carlo solver used to compute the dose distributions, figs. 11.11 and 11.14 also show some systematic differences around the peak/range of the beams.

In chapter 14 we apply the derived energy distribution for the quantification of expected dose and dose variance in the presence of range uncertainties. We observe, that the model errors do not affect the accuracy of the expected value significantly, but introduce deviations similar to those in fig. 11.11 and fig. 11.14 when computing the variance.



(a) (b)

Figure 11.10: Comparison of dose cross sections in a water box for density scaled with factor (1 ± 0.05) and approximation using energy scaled with $(1 \mp 0.05)^{(1/p)}$, $p = 1.77$ (according to section 11.2). (a) Cuts through central beam axis, (b) Lateral cut across beams at $z = 3.625$[mm].

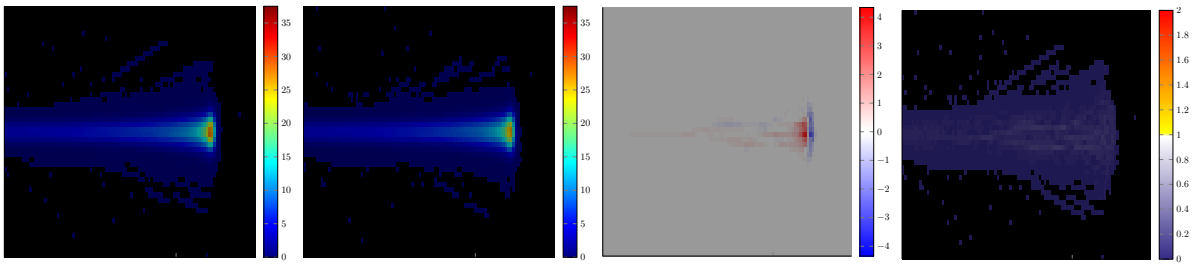


(a) Transformed energy, $E \cdot (1 + 0.05)^{(1/1.77)}$

(b) Scaled down density, $\rho \cdot (1 - 0.05)$

(c) Difference between (a) and (b)

(d) $\gamma_{3\,\mathrm{mm}/3\,\%}$-map, passrate 100%



(e) Transformed energy, $E \cdot (1 - 0.05)^{(1/1.77)}$

(f) Scaled up density, $\rho \cdot (1 + 0.05)$

(g) Difference between (e) and (f)

(h) $\gamma_{3\,\mathrm{mm}/3\,\%}$-map, passrate 100%

Figure 11.11: Comparisons of dose in water for density scaled with factor (1 ± 0.05) and approximation using energy scaled with $(1 \mp 0.05)^{(1/p)}$, $p = 1.77$ (according to section 11.2).

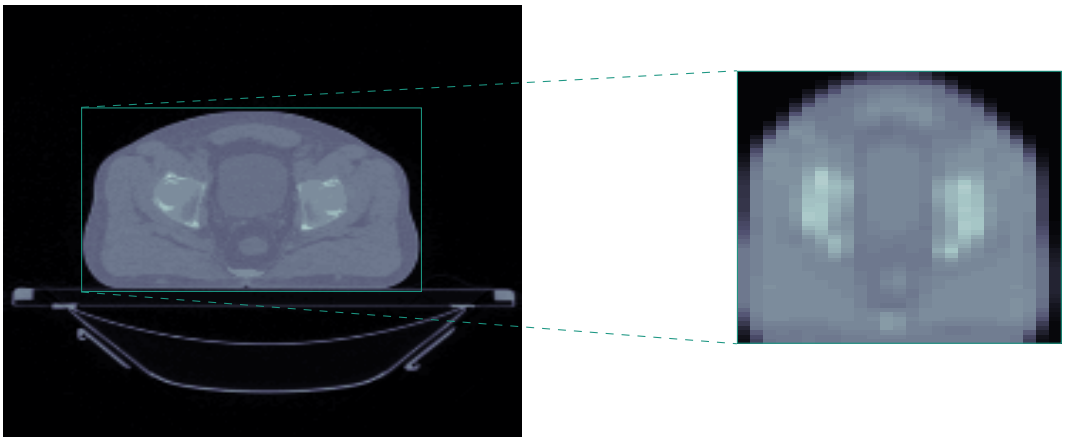
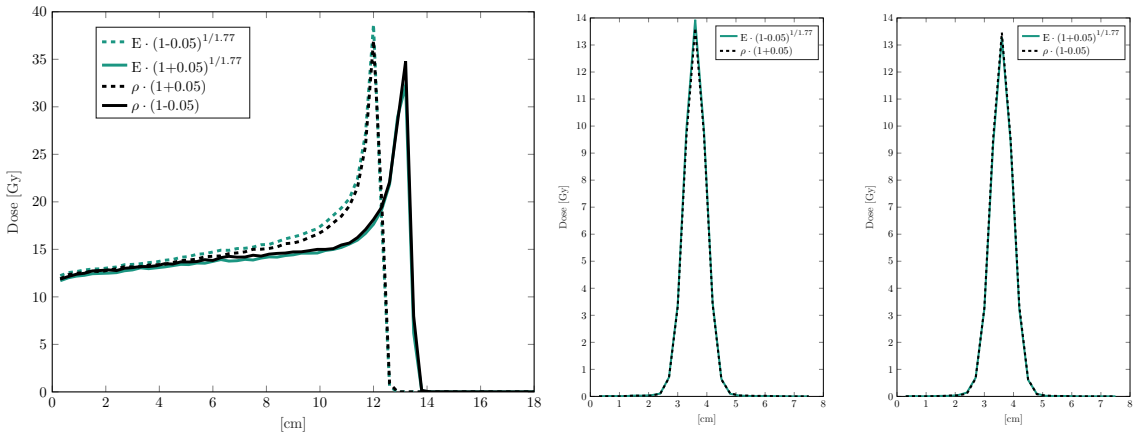


Figure 11.12: Original and downsampled version of the prostrate CT scan from [42], axial plane at 36 [mm].

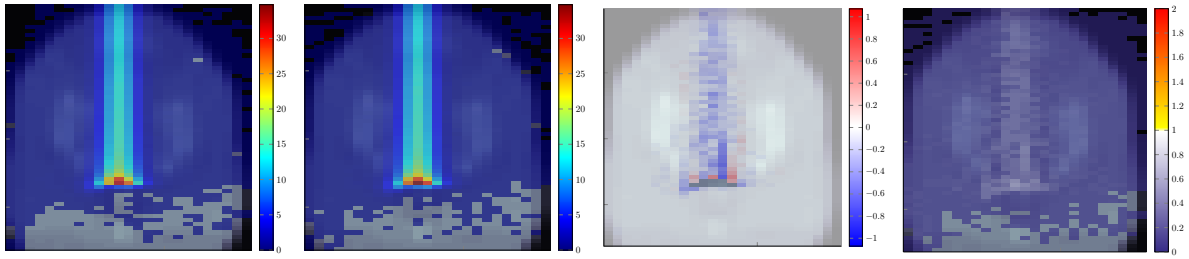


(a) (b)

Figure 11.13: Comparison of dose cross sections in a low-resolution heterogeneous CT scan (fig. 11.12) for density scaled with factor (1 ± 0.05) and approximation using energy scaled with $(1 \mp 0.05)^{(1/p)}$, $p = 1.77$ (according to section 11.2). (a) Cuts through central beam axis, (b) Lateral cut across beams at $z = 4.5$[mm].

11.5.2 Representation of time-dependent uncertainties via their autocovariance function

Next, we want to investigate the approximation of time-dependent spatial uncertainties through the covariance matrix of beamlets applied at different time points during the treatment. First, our implementation was validated against Monte Carlo results for individual realizations of positional shifts of the patient over time. This ensures, that shifting just the beamlet applied at the considered time points is equivalent to a shift of the complete patient geometry. For a single ray with 6 energy levels in a water box, shifted by ± 2 mm every 10 ms, the dose results in each time step reached 100 % agreement in a

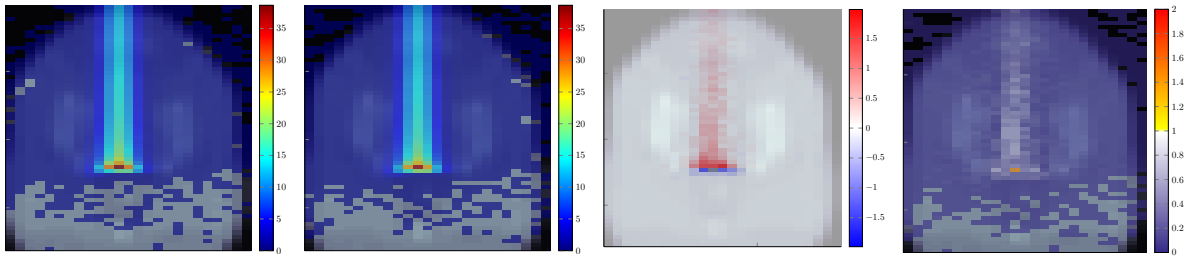


(a) Transformed energy, $E \cdot (1 + 0.05)^{(1/1.77)}$

(b) Scaled down density, $\rho \cdot (1 - 0.05)$

(c) Difference between (a) and (b)

(d) $\gamma_{3\,\mathrm{mm}/3\,\%}$-map, passrate 100%



(e) Transformed energy, $E \cdot (1 - 0.05)^{(1/1.77)}$

(f) Scaled up density, $\rho \cdot (1 + 0.05)$

(g) Difference between (e) and (f)

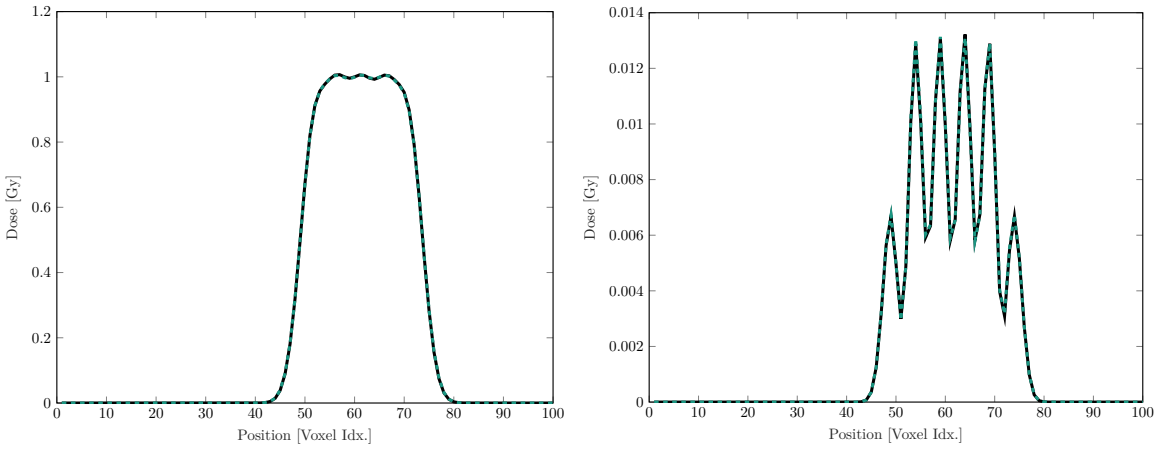
(h) $\gamma_{3\,\mathrm{mm}/3\,\%}$-map, passrate 99.908%

Figure 11.14: Comparisons of dose in a low-resolution heterogeneous CT scan (fig. 11.12) for density scaled with factor (1 ± 0.05) and approximation using energy scaled with $(1 \mp 0.05)^{(1/p)}$, $p = 1.77$ (according to section 11.2).

γ-analysis with tolerances of $2\,\mathrm{mm}/2\,\%$ in dose and location, respectively.

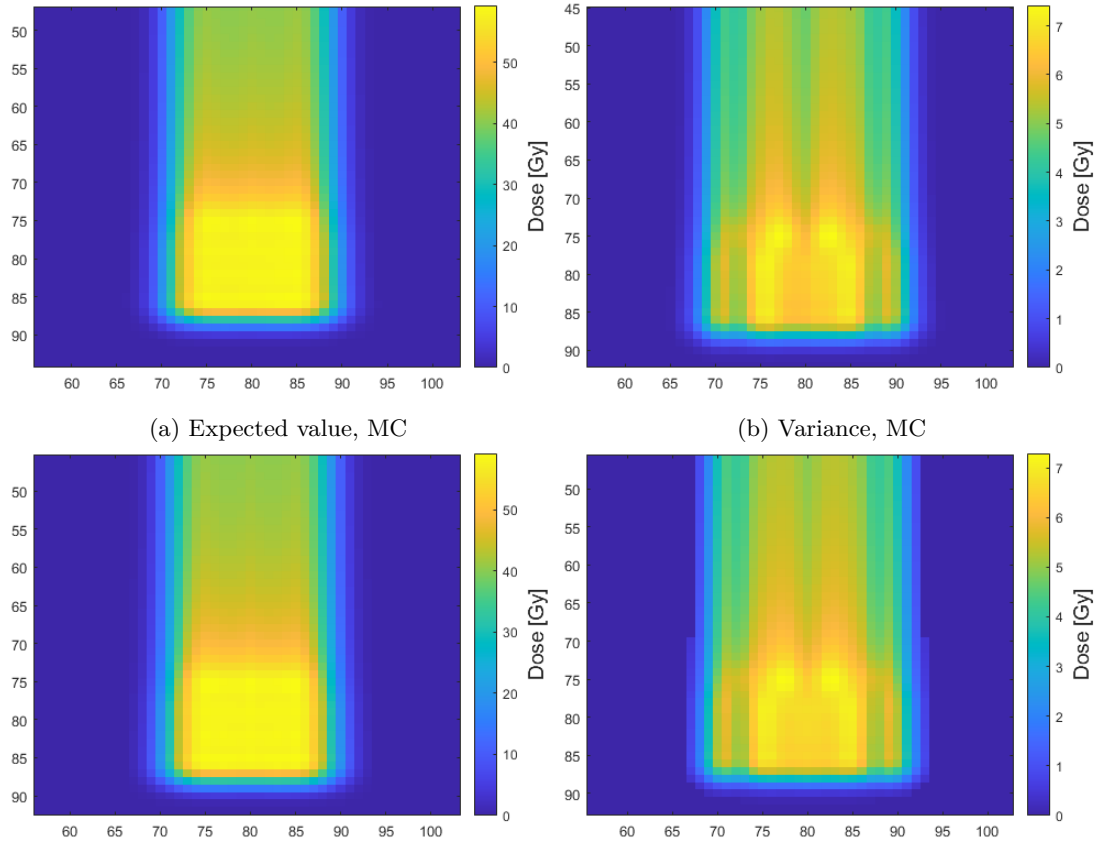
Further, we want to validate, that the derived multivariate Gaussian models with tailored covariance matrices are sufficient to represent the considered time-dependent processes. First, we consider a simple one-dimensional spatial uncertainty following an AR(1) process as described in section 11.4. Figure 11.15 shows that a step-by-step simulation of the time series using a Monte Carlo method for the uncertain variable U_t yields equivalent results for the dose expected value and variance in a water phantom as inputting the covariance matrix into the uncertainty propagation method introduced in chapter 13.

Figure 11.16 shows that this also holds for a three-dimensional uncertainty following a Gaussian process with a periodic squared exponential kernel with parameters $\sigma = 1$, $p = 3$, $l_1 = 10$ and $l_2 = 2$. Note, that this level of accuracy can only be expected when using the exact time points at which beamlets are applied for the simulation of the time series and when considering either homogeneous materials or shifts of the complete region of interest. Future work could investigate the possibility of an application for uncertain shifts of the target within a heterogeneous material.



(a) Expected value (b) Variance

Figure 11.15: Comparison of expected value and variance for a Monte Carlo simulation of the AR(1) process and a direct APM computation based on the covariance matrix.



(a) Expected value, MC (b) Variance, MC

(c) Expected value, APM (d) Variance, APM

Figure 11.16: Comparison of a Monte Carlo simulation of the AR(1) process using eq. (11.11) and the approximation using a multivariate Gaussian with a covariance structure according to the AR(1)-process.

CHAPTER 12

Background: Uncertainty propagation methods

Given an uncertainty model for the input parameters of a dose calculation algorithm, as defined in the previous chapter, we want to gain insight into the uncertainty of the output, i.e., the dose. Typically, this uncertainty is quantified in terms of a probability distribution or moments of this distribution, such as the expected value or variance.

Different approaches for uncertainty propagation can be grouped into *intrusive* and *non-intrusive* methods according to the extent of their interference with the deterministic solution method (here dose calculation). The boundaries between both are blurred, here we therefore define non-intrusive to mean that a method does not change the system of equations of the deterministic problem. Note, that this also includes methods where slight changes to the solver would be necessary, e.g., in the accumulation of quantities of interest or the sampling of random parameters. Approaches that fall into this marginal category will sometimes be referred to as *minimally intrusive*. In the following, we will briefly review a number of well-known uncertainty propagation methods for a general class of problems. Afterward, we will put a more in-depth focus on three methods aimed at specific types of dose calculation algorithms, which are developed and/or extended within the novel work presented in this thesis.

12.1 Intrusive methods

12.1.1 Stochastic Galerkin

Generalized polynomial chaos

The most frequently used intrusive uncertainty propagation method is based on an expansion of the random variable or function in orthogonal polynomials. While the initial polynomial chaos approach [65, 231] used only Hermite polynomials as the orthogonal basis to represent random processes, *generalized*

polynomial chaos uses different kinds of orthogonal polynomials based on the probability density function of the random quantity [237]. Table 12.1 gives an overview of common distributions and the associated orthogonal polynomials.

Following the Cameron-Martin theorem [30], any function from $L_2(C)$ can be approximated using orthogonal polynomials and the approximation converges in an $L_2(C)$-sense. Thus, we can e.g. rewrite the uncertain particle density in eq. (10.1)

$$\psi(\boldsymbol{Z};\boldsymbol{\Delta}) = \sum_{p=0}^{\infty} c_p(\boldsymbol{Z})\Phi_p(\boldsymbol{\Delta}) \tag{12.1}$$

with Φ_p orthogonal polynomials, i.e.

$$\langle \Phi_i, \Phi_j \rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_i(\boldsymbol{\Delta})\Phi_j(\boldsymbol{\Delta})p_{\Delta}(\boldsymbol{\Delta})\mathrm{d}\boldsymbol{\Delta} = \delta_{ij}\langle \Phi_i^2 \rangle = \delta_{ij}h_j^2. \tag{12.2}$$

In practice, the sums in eq. (12.1) are truncated to a finite order P, resulting in the approximation

$$\psi^P(\boldsymbol{Z};\boldsymbol{\Delta}) = \sum_{p=0}^{P} c_p(\boldsymbol{Z})\Phi_p(\boldsymbol{\Delta}). \tag{12.3}$$

The goal is to compute the coefficients c_p. This would give us not only a functional representation of the particle density depending on the uncertain parameter but also the expected value and variance which can be determined using the orthogonality of the polynomials as follows:

$$\begin{aligned}
\mu_\psi &= \mathbb{E}_{p_\Delta}[\psi(\boldsymbol{Z};\boldsymbol{\Delta})] = \mathbb{E}_{p_\Delta}\left[\sum_{p=0}^{\infty} c_p(\boldsymbol{Z})\Phi_p(\boldsymbol{\Delta})\right] \\
&= \sum_{p=0}^{\infty} c_p(\boldsymbol{Z})\mathbb{E}_{p_\Delta}[1\cdot\Phi_p(\boldsymbol{\Delta})] \\
&= \sum_{p=0}^{\infty} c_p(\boldsymbol{Z})\underbrace{\mathbb{E}_{p_\Delta}[\Phi_0(\boldsymbol{\Delta})\cdot\Phi_p(\boldsymbol{\Delta})]}_{=\delta_{0p}} \\
&= c_0(\boldsymbol{Z}) \qquad (12.4)\\
\sigma_\psi^2 &= Var\left(\sum_{p=0}^{\infty} c_p(\boldsymbol{Z})\Phi_p(\boldsymbol{\Delta})\right) \qquad (12.5)\\
&= \mathbb{E}_{p_\Delta}\left[\left(\sum_{p=0}^{\infty} c_p(\boldsymbol{Z})\Phi_p(\boldsymbol{\Delta}) - c_0(\boldsymbol{Z})\right)^2\right] \\
&= \mathbb{E}_{p_\Delta}\left[\left(\sum_{p=1}^{\infty} c_p(\boldsymbol{Z})\Phi_p(\boldsymbol{\Delta})\right)^2\right] \\
&= \sum_{p=1}^{\infty} c_p(\boldsymbol{Z})^2 h_p^2 \approx \sum_{p=1}^{P} c_p(\boldsymbol{Z})^2 h_p^2. \qquad (12.6)
\end{aligned}$$

Table 12.1: Common continuous probability distributions and the associated types of gPC basis polynomials [237, see e.g.]

Distribution	Orthogonal polynomial
Gaussian	Hermite
Uniform	Legendre
Exponential	Laguerre
Gamma	Generalized Laguerre
Beta	Jacobi

Galerkin projection

In order to find the coefficients of the truncated polynomial expansion eq. (12.3), a Galerkin projection is applied. If $f(\psi(\boldsymbol{Z}, \boldsymbol{\Delta}), \boldsymbol{\Delta}) = 0$ describes the transport equation for $\psi^{(P)}$, plugging in the truncated expansion from eq. (12.3) creates a non-zero residual $r^{(P)}(\boldsymbol{Z}, \boldsymbol{\Delta}) = f(\psi^{(P)}(\boldsymbol{Z}, \boldsymbol{\Delta}), \boldsymbol{\Delta})$. In the stochastic Galerkin method, we demand that the residual is orthogonal to the ansatz space spanned by the polynomial basis functions, i.e.,

$$\mathbb{E}[f(\psi^{(P)}(\boldsymbol{Z}; \boldsymbol{\Delta}), \boldsymbol{\Delta})\Phi_j] \stackrel{!}{=} 0 \quad j = 0, ..., P. \tag{12.7}$$

This yields a new system of $P+1$ equations. Depending on the structure of the equations and the required order P of the expansion, the stochastic Galerkin method can be very efficient compared to non-intrusive methods. However, it also requires the sometimes non-trivial derivation of a new system of equations and subsequently changes in the solution algorithm. Especially in clinical practice, intrusion into the solver code is often infeasible due to the use of proprietary software for dose computations. For this reason, we focus mostly on non- or minimally intrusive strategies when considering Monte Carlo dose calculations, which are typically carried out with professional software. Note, that there is also a non-intrusive method based on generalized polynomial chaos, which will be discussed in section 12.2.4.

12.2 Non-intrusive methods

Non-intrusive strategies typically involve computing the dose at different points in the uncertain parameter space. When considering non-intrusive UQ methods, the dose calculation algorithm can thus be treated as a black box returning the dose $D(\boldsymbol{r}; \boldsymbol{\delta}_j)$ at position $\boldsymbol{r}$ for a point $\boldsymbol{\delta}_j$ in the space of the uncertain variable $\boldsymbol{\Delta} \sim p_\Delta$.

12.2.1 Monte Carlo

Remark 12.1. *Note, that for the reader's convenience, we again recapitulate the classical Monte Carlo method, previously introduced in section 5.1, in the following section.*

The Monte Carlo method is a numerical integration method, which is based on evaluating the integrand at random positions sampled from the probability distribution of the integration variable. For realizations

$\boldsymbol{\delta}_i\,, i = 1, ..., N$ of the random variable $\boldsymbol{\Delta} \in \mathbb{R}^d$ with probability density function $p : \mathbb{R}^d \to \mathbb{R}_+$, the estimator is defined as follows:

$$I(f) = \mathbb{E}_p[f(\boldsymbol{\Delta})] = \int_{\mathbb{R}^d} f(\boldsymbol{\Delta}) p(\boldsymbol{\Delta}) \mathrm{d}\boldsymbol{\Delta} \approx \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{\delta}_i) := I^N(f)\,. \tag{12.8}$$

Note, that with the *Strong Law of Large Numbers*, one can show

$$\mathbb{P}\left(\lim_{N\to\infty} I^N(f) = I(f)\right) = 1\,, \tag{12.9}$$

i.e., the Monte Carlo estimator converges to the true value in probability. Further, the independence of the random variables $\boldsymbol{\delta}_i$ and the *Central Limit Theorem* provide a formula for the variance of the estimator:

$$Var(I^N(f)) = \frac{v_f^2}{N}\,, \tag{12.10}$$

which depends solely upon the variance v_f^2 of the integrand and the number of samples N. The estimator thus converges with $O(\frac{1}{\sqrt{N}})$ and its accuracy can be controlled through the number of computed realizations.

12.2.2 (Randomized) quasi-Monte Carlo

Due to the slow convergence and therefore often large sample sizes required for Monte Carlo estimates, numerous variations of the method exist, which reduce the variance of the estimator [28]. The Monte Carlo method relies on randomly sampled points, which are in practice implemented using pseudo-random numbers [29]. These can be generated to be uniformly distributed using a random number generator and transformed to the required probability distribution [210, 48, 21]. The quasi-Monte Carlo method replaces the pseudo-random points, which imitate a random uniform distribution, by a set of deterministic *low discrepancy* points $P_N := \{\boldsymbol{u_1}, ..., \boldsymbol{u_N}\} \in [0,1)^d$. These points aim to cover the domain more uniformly [28]:

$$I^N_{qMC}(f) := \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{u_i})\,. \tag{12.11}$$

Here, the discrepancy of a point set P_N is defined as

$$D_N(P_N) = \sup_{\mathcal{J}\subseteq[0,1]^d} \left| \frac{\#\text{points from } P_N \text{ in } \mathcal{J}}{N} - \text{vol}(\mathcal{J}) \right| \tag{12.12}$$

where $\text{vol}(\mathcal{J})$ is the volume of the d-dimensional subset $\mathcal{J}$ of the unit cube. Further, the Koksma-Hlawka inequality states that for functions with finite variation, the integration error is bounded by the product of the variation and the discrepancy of the sequence used for numerical integration. Thus, quasi-Monte Carlo methods aim at minimizing the discrepancy.

Theorem 12.1 (Koksma-Hlawka)**.** *Let $\{u_i\}_{i=1,..,N} \in [0,1)^d$ a sequence of N points with discrepancy D_N and f a real-valued function on $[0,1)^d$ with bounded Hardy-Krause variation $V(f)$, then*

$$\left| \int_{[0,1)^d} f(\boldsymbol{X}) \mathrm{d}\boldsymbol{X} - \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{u_i}) \right| \leq V(f) \cdot D_N$$

This can for example be achieved using grid points or number theoretic sequences such as the Halton [70] or Sobol sequence [198], which will be used in chapter 14. An upper bound to the discrepancy is derived by [155]:

$$D_N \leq c(d)\frac{\log(N)^d}{N}. \tag{12.13}$$

When considering a non-uniform distribution p, the low-discrepancy points have to be transformed to represent this probability distribution, i.e., apply τ_p such that $\tau_p(\boldsymbol{u}) \sim p$ for $\boldsymbol{u} \sim U[0,1]^d$ [29]. In order to analyze the properties of the quasi-MC approximation for non-uniform distributions, one can either use eq. (12.11) and theorem 12.1 with the original low-discrepancy point-set P_N and the composed function $f_{\tau_p} = f \circ \tau_p$ or use a generalized framework that applies to measures apart from the Lebesgue measure [29], such as the generalized Koskma–Hlawka inequalities discussed in [78, 4].

Typically, a faster convergence rate can be achieved with quasi-MC compared to standard MC, especially in lower dimensions. However, due to the deterministic nature of quasi-MC, properties such as the unbiasedness of the standard Monte Carlo estimator are lost and no general error formula is known [116].

Randomized quasi-Monte Carlo therefore reintroduces randomness to the method, while retaining the convergence properties of quasi-MC. The resulting point set is a low discrepancy sequence, in which each individual point is uniformly distributed [116]. In the case of a grid rule, this can, for example, be achieved by rotating the lattice or randomly drawing points within the grid cells from a uniform distribution. For number theoretic sequences, different random shuffling and scrambling[1] strategies have been developed [163, 143, 117, 105]. We use a Sobol sequence with linear scrambling and a random digit shift according to [143]. For square-integrable functions, the resulting randomized quasi-Monte Carlo approach preserves the properties of Sobol sequences, ensuring a variance of $O\left(\frac{\log(N)^d}{N}\right)$ for N d-dimensional sample points, while also generating an unbiased estimate and allowing for statistical estimation of the integration error [163, 164].

12.2.3 Stochastic collocation

Similar to quasi-Monte Carlo, the *stochastic collocation* method is also a non-intrusive method based on a deterministic set of points, at which the model (here dose calculation algorithm) is evaluated. However, collocation methods are based on the idea of finding a low-dimensional, continuous representation of the solution (often a polynomial). This representation should satisfy the underlying equations exactly at the so-called collocation points and serves as an interpolation function otherwise. Note, that *stochastic collocation* is sometimes also used as an overarching term to denote non-intrusive methods which evaluate a deterministic solver at different collocation points. In this definition, Monte Carlo as well as quasi-Monte Carlo methods would also be included.

An example of interpolation polynomials used in stochastic collocation are Lagrange polynomials, described e.g. in [56] for the interpolation of one-dimensional points:

[1] Shuffling refers to a randomization of the sample index while scrambling refers to the randomization of the sample value.

$$L_j(\delta) = \prod_{\substack{k=1 \\ k \neq j}}^{N} \frac{\delta - \delta_k}{\delta_j - \delta_k}. \tag{12.14}$$

The dose response would then be approximated using N collocation points by

$$D(\boldsymbol{r}; \delta) \approx \sum_{j=1}^{N} D(\boldsymbol{r}; \delta_j) L_j(\delta). \tag{12.15}$$

From eq. (12.14), it is apparent, that $D(\boldsymbol{r}; \delta)$ returns exactly the dose response values at points δ_j, $j = 1, ..., N$ and interpolates smoothly inbetween. The choice of $\delta_1, ..., \delta_N$ is crucial for the efficiency of the approach, especially in higher dimensions. Typically the point sets are then constructed from one-dimensional quadratures (e.g. Clenshaw-Curtis or Gaussian quadrature rules [41, 207]). In the following, we will briefly review two common ways of constructing higher-dimensional grids.

Tensorized grids

The simplest way to obtain a d-dimensional quadrature rule $Q^{(d)}$ from a given one-dimensional quadrature rule

$$Q^{(1)} f := \sum_{i=1}^{N} f(\delta_i) w_i \approx \int_{[0,1]} f(\delta) p(\delta) \mathrm{d}\delta, \quad \delta \sim p(\delta)$$

is tensorization, i.e. applying the one-dimensional quadrature in each dimension. Then

$$Q^{(d)} f := \sum_{i_1=1}^{N_1} ... \sum_{i_d=1}^{N_d} f(\delta_{i_1}^{(1)}, ..., \delta_{i_d}^{(d)}) \cdot w_{i_1}^{(1)} \cdot ... \cdot w_{i_d}^{(d)} \approx \int_{[0,1]} ... \int_{[0,1]} f(\delta_1, ..., \delta_d) p_1(\delta_1) \cdot ... \cdot p(\delta_d) \mathrm{d}\delta_1 ... \mathrm{d}\delta_d,$$

$$\text{with } \boldsymbol{\delta} = (\delta_1, ..., \delta_d) \sim p(\boldsymbol{\delta}) = \prod_{k=1}^{d} p_k(\delta_k)$$

is the d-dimensional quadrature. The tensorization of a one-dimensional quadrature rule with N points requires N^d points to maintain the same error in d dimensions. Thus, the convergence rate decreases with the number of dimensions. This issue is referred to as the *curse of dimensionality* and prohibits the use of grid-based collocation methods for applications with high-dimensional phase spaces (such as radiation therapy). An effort to mitigate the *curse of dimensionality* is made by sparse grids.

Sparse grids

In contrast to tensorized grids, which aim at integrating polynomials of *maximal* degree N exactly, sparse grids reduce the number of required quadrature points and focus on the integration of polynomials with *total*[2] degree N. To achieve this, sparse grids use *nested* quadratures, where each level l reuses the points of the previous level $l - 1$. A one-dimensional quadrature at level l is defined by

$$Q_l^{(1)} f := \sum_{i=1}^{N_l} f(\delta_i^{(l)}) w_i^{(l)}.$$

[2]Note, that for a multivariate polynomial with d variables, the maximal degree refers to the highest occurring power of any one variable with non-zero coefficient, while the total degree refers to the maximum sum of such powers in a product of the d variables

We further define the difference between two consecutive levels of a one-dimensional quadrature as

$$\mathrm{Diff}_l^{(1)} f := \left(Q_l^{(1)} - Q_{l-1}^{(1)}\right) f,$$

where $Q_0^{(1)} = 0$.

These operators also constitute a quadrature with the nodes of $Q_l^{(1)}$ and weights which are given by the difference of the weights in $Q_l^{(1)}$ at levels l and $l-1$. We can tensorize this new quadrature to construct a $d-$dimensional sparse grid as follows

$$Q_l^{(d)} f = \sum_{|l'| \leq l+d-1} \left(\mathrm{Diff}_{l_1}^{(1)} \otimes ... \otimes \mathrm{Diff}_{l_d}^{(1)}\right) f \ ,$$

where $|l'|$ is the total degree of the multi-index $l' \in \mathbb{N}^d$.

According to [213], sparse grids require $O(N \log_2(N)^{d-1})$ nodes to integrate polynomials of total degree N exactly. While this is an improvement over tensorized grids, the number of points is still higher than the order in a gPC expansion with the same total degree [108]. Therefore, we now lastly consider a non-intrusive variation of the gPC approach.

12.2.4 Non-intrusive polynomial chaos

From section 12.1.1, we recall the truncated polynomial chaos expansion, here however we directly expand the dose itself

$$D(\boldsymbol{r}; \boldsymbol{\Delta}) \approx \sum_{p=1}^{P} c_p(\boldsymbol{r})\, \Phi_p(\boldsymbol{\Delta}).$$

For a d-dimensional random variable, univariate polynomials can be chosen according to the distribution in each dimension based on the Wiener-Askey scheme (see table 12.1). Multivariate orthogonal polynomials can then be derived similarly to the construction of higher-dimensional quadrature rules through tensorization.

Construction of a polynomial basis

Let $\{\gamma_j\}_{j=0,..,N} \subseteq \mathbb{N}_0^d$ denote the set of multi-indices and let $\gamma_j = (\gamma_j^{(1)}, ..., \gamma_j^{(d)})$ be a multi-index from the set. A multivariate polynomial Φ_j is constructed from univariate polynomials as follows

$$\Phi_j(\boldsymbol{\delta}) = \prod_{k=1}^{d} \Phi_{\gamma_j^{(k)}}(\boldsymbol{\delta}_k).$$

The full O^{th}-order set of basis polynomials is then usually considered to include all $N+1 = \frac{(d+O)!}{d!O!}$ polynomials of total order $\leq O$:

$$\left\{ \prod_{k=1}^{d} \Phi_{\gamma_j^{(k)}}(\boldsymbol{\delta}_k) : \sum_{k=1}^{d} \gamma_j^{(k)} \leq O \right\} = \left\{ \Phi_j(\boldsymbol{\delta}) : \sum_{k=1}^{d} \gamma_j^{(k)} \leq O \right\}.$$

Table 12.2: Two-dimensional example for the construction of multi-indices and the full polynomial basis set of order $O = 2$. Here $\Phi^1_{\gamma_j^{(k)}}, \Phi^2_{\gamma_j^{(k)}}$ could be two different types of polynomials chosen according to the distributions of δ_1, δ_2.

order	γ_j	Φ_j
$O = 0$	$\gamma_0 = (0,0)$	$\Phi_0(\boldsymbol{\delta}) = \Phi^1_0(\delta_1) \cdot \Phi^2_0(\delta_2)$
1	$\gamma_1 = (1,0)$	$\Phi_1(\boldsymbol{\delta}) = \Phi^1_1(\delta_1) \cdot \Phi^2_0(\delta_2)$
	$\gamma_2 = (0,1)$	$\Phi_2(\boldsymbol{\delta}) = \Phi^1_0(\delta_1) \cdot \Phi^2_1(\delta_2)$
2	$\gamma_3 = (1,1)$	$\Phi_3(\boldsymbol{\delta}) = \Phi^1_1(\delta_1) \cdot \Phi^2_1(\delta_2)$
	$\gamma_4 = (2,0)$	$\Phi_4(\boldsymbol{\delta}) = \Phi^1_2(\delta_1) \cdot \Phi^2_0(\delta_2)$
	$\gamma_5 = (0,2)$	$\Phi_5(\boldsymbol{\delta}) = \Phi^1_0(\delta_1) \cdot \Phi^2_2(\delta_2)$

Table 12.2 shows the construction of a polynomial basis for $d = 2$ and maximal order $O = 2$. It is apparent that the number of polynomials in the basis set grows rapidly with increasing order and dimension. For this reason, approaches to generate sparse polynomial bases have been developed [22, 23, 173]. A simple method introduced e.g. in [22] excludes high-order interactions between parameters by truncating the multi-index set. Here only multivariate polynomials whose associated multi-index has a q-quasi norm $\leq$ order O are included. The q-quasi norm is defined as [23]

$$\|\gamma_j\|_q = \left(\sum_{k=1}^{d} \gamma_j^{(k)\,q} \right)^q .$$

Note that by choosing $q = 1$ we recover the full basis set.

Non-intrusive spectral projection

Using the orthogonality and a spectral projection, the coefficients of the polynomial expansion are given by

$$c_p = \frac{\langle D(\boldsymbol{r};\boldsymbol{\Delta}), \Phi_p(\boldsymbol{\Delta})\rangle}{\langle \Phi_p(\boldsymbol{\Delta}), \Phi_p(\boldsymbol{\Delta})\rangle} = \frac{\int \dots \int D(\boldsymbol{r};\boldsymbol{\Delta}) \cdot \Phi_p(\boldsymbol{\Delta}) \cdot p_\Delta(\boldsymbol{\Delta}) \mathrm{d}\Delta_1 \dots \mathrm{d}\Delta_d}{h_p^2}. \tag{12.16}$$

Thus, to determine the coefficients, we need to solve a multi-dimensional integral depending on the unknown dose response. This can be done using any of the methods introduced in this chapter, e.g. for a grid rule with nodes $\boldsymbol{\delta}^{(1)}, \dots, \boldsymbol{\delta}^{(N_q)} \in \mathbb{R}^d$, weights $w^{(1)}, \dots, w^{(N_q)}$ and dose responses $D(\boldsymbol{r}, \boldsymbol{\delta}^{(1)}), \dots, D(\boldsymbol{r}, \boldsymbol{\delta}^{(N_q)})$ the coefficients can be estimated by

$$c_p = \frac{1}{h_p^2} \sum_{i=1}^{N_q} D(\boldsymbol{r}, \boldsymbol{\delta}^{(i)})\, \Phi_p(\boldsymbol{\delta}^{(i)}) w^{(i)}. \tag{12.17}$$

Point collocation

Instead of deriving the expansion coefficients from a spectral projection, the *point collocation* method solves a linear system of equations

$$\begin{pmatrix} \Phi_0(\boldsymbol{\delta}^{(1)}) & \cdots & \Phi_N(\boldsymbol{\delta}^{(1)}) \\ \vdots & & \vdots \\ \Phi_0(\boldsymbol{\delta}^{(N+1)}) & \cdots & \Phi_N(\boldsymbol{\delta}^{(N+1)}) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} = \begin{pmatrix} D(\boldsymbol{r}, \boldsymbol{\delta}^{(1)}) \\ \vdots \\ D(\boldsymbol{r}, \boldsymbol{\delta}^{(N+1)}) \end{pmatrix} \tag{12.18}$$

which requires at least $N+1$ dose response values at collocation points chosen e.g. according to a multi-dimensional quadrature as described in section 12.2.3. For more than $N+1$ collocation points, the system is overdetermined and can be solved using a least squares approach. In [85] oversampling by at least factor 2 is recommended, to increase the accuracy of the polynomial coefficients. For higher dimensions, oversampling is even found to be more computationally efficient than increasing the number of polynomials to achieve the same accuracy [85].

CHAPTER 13

Analytical probabilistic modeling for pencil beam methods

The first more specialized uncertainty propagation method we want to introduce here was first presented in [10]. Analytical probabilistic modeling (APM) exploits the simple functional form of the dose in pencil beam computations (section 5.2), which is based on a Gaussian kernel convolved with the incident fluence. This model makes it possible to analytically include Gaussian uncertainty models. In this thesis, we merely extend the applicability of APM to a more generalized set of uncertainty models (see chapter 11) and use the results to validate other, novel methods.

The goal of APM is to compute moments of the dose (or some other related quantity of interest [e.g. 232]) given by

$$\mathbb{E}_{p_{\boldsymbol{\Delta}}}[D(\boldsymbol{r},\boldsymbol{\Delta})^m] = \int p_{\boldsymbol{\Delta}}(\boldsymbol{\Delta}) D(\boldsymbol{r},\boldsymbol{\Delta})^m \mathrm{d}\boldsymbol{\Delta}.$$

The simple functional approximation of the pencil beam algorithm is used for the dose and uncertainties are assumed to have a multivariate normal distribution $\boldsymbol{\Delta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\Delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\Delta}})$. While this seems like a very restricting assumption, it is a common one [233, 217, 174, 10, 61, 226]. Also, we have shown in chapter 11 that even complex time-dependent uncertainties can be fit into a Gaussian framework and how this can be achieved. We now consider the deterministic dose approximation in the pencil beam approach:

$$\begin{aligned} D(\boldsymbol{r}) &= \sum_j w_j L_j^x(\boldsymbol{r}) L_j^y(\boldsymbol{r}) R_j(\boldsymbol{r}) \\ &= \sum_j w_j L_j^x(r_x) L_j^y(r_y) R_j(r_z) \\ &= \sum_j w_j \mathcal{N}(r_x; \mu_x^{(j)}, \Sigma_x^{(j)}) \cdot \mathcal{N}(r_y; \mu_y^{(j)}, \Sigma_y^{(j)}) \cdot R_j(r_z) \end{aligned}$$

The pencil beam approximation already employs a conveniently Gaussian form for lateral components L_j^x and L_j^y [200, 68]. The depth component $R_j(r_z)$ on the other hand cannot be modeled with an individual Gaussian [25], however under the assumption of smoothness, it can be approximated arbitrarily well using a Gaussian mixture [191]:

$$R_j(r_z) = \sum_{k=1}^{K} w_j^{(k)} \mathcal{N}(r_z; \mu_z^{(k)}, \Sigma_z^{(k)}).$$

In [10] the parameters $\mu_z^{(k)}, \Sigma_z^{(k)}$ are found by fitting a Gaussian mixture model with ten components, each with distinct weight, mean and variance to integrated depth dose data from TOPAS MC [175] using an expectation maximization algorithm. Then, the integrals for expected value and variance as well as higher moments can be solved in closed form. Let $\boldsymbol{\delta} = (\delta_x, \delta_y, \delta_z)$ be a realisation of the lateral offset in x and y direction as well as range uncertainties affecting the radiological depth z. We assume the errors to be independent, i.e.

$$p(\delta_x, \delta_y, \delta_z) = p_x(\delta_x) \cdot p_y(\delta_y) \cdot p_z(\delta_z) = \mathcal{N}(\delta_x; 0, \Sigma_{\Delta_x}) \cdot \mathcal{N}(\delta_y; 0, \Sigma_{\Delta_y}) \cdot \mathcal{N}(\delta_z; 0, \Sigma_{\Delta_z})$$

The dose for specific realization $\boldsymbol{\delta}$ of the uncertain parameters is given by

$$\begin{aligned} D(\boldsymbol{r}, \boldsymbol{\delta}) &= \sum_j w_j L_j^x(r_x + \delta_x) \cdot L_j^y(r_y + \delta_y) \cdot R_j(r_z + \delta_z) \\ &= \sum_j w_j \mathcal{N}(r_x; \mu_x^{(j)} + \delta_x, \Sigma_x^{(j)}) \cdot \mathcal{N}(r_y; \mu_y^{(j)} + \delta_y, \Sigma_y^{(j)}) \cdot \sum_{k=1}^{K} w_j^{(k)} \mathcal{N}(r_z; \mu_z^{(j,k)} + \delta_z, \Sigma_z^{(j,k)}). \end{aligned}$$

The expected value can then be computed by solving

$$\begin{aligned} \mathbb{E}_{p_\Delta}[D(\boldsymbol{r}, \boldsymbol{\Delta})] &= \int D(\boldsymbol{r}, \boldsymbol{\Delta}) \cdot p(\Delta_x, \Delta_y, \Delta_z) \mathrm{d}\boldsymbol{\Delta} \\ &= \int\int\int D(\boldsymbol{r}, (\Delta_x, \Delta_y, \Delta_z)) \cdot p_x(\Delta_x) \cdot p_y(\Delta_y) \cdot p_z(\Delta_z) \mathrm{d}\Delta_x \mathrm{d}\Delta_y \mathrm{d}\Delta_z \\ &= \sum_j w_j \left\{ \int L_j^x(r_x + \Delta_x) p_x(\Delta_x) \mathrm{d}\Delta_x \right\} \left\{ \int L_j^y(r_y + \Delta_y) p_y(\Delta_y) \mathrm{d}\Delta_y \right\} \left\{ \int R_j(r_z + \Delta_z) p_z(\Delta_z) \mathrm{d}\Delta_z \right\} \end{aligned}$$

and the covariance by plugging $\mathbb{E}_{p_z}[D(\boldsymbol{r}, \boldsymbol{\Delta})]$ and

$$\begin{aligned} &\mathbb{E}_{p_\Delta}[D(\boldsymbol{r}, \boldsymbol{\Delta}) D(\boldsymbol{r'}, \boldsymbol{\Delta})] \\ &= \int\int\int D(\boldsymbol{r}, (\Delta_x, \Delta_y, \Delta_z)) D(\boldsymbol{r'}, (\Delta_x, \Delta_y, \Delta_z)) \cdot p_x(\Delta_x) \cdot p_y(\Delta_y) \cdot p_z(\Delta_z) \mathrm{d}\Delta_x \mathrm{d}\Delta_y \mathrm{d}\Delta_z \\ &= \sum_{i,j} w_j w_m \cdot \left\{ \int L_j^x(r_x + \Delta_x) L_m^x(r'_x + \Delta_x) p_x(\Delta_x) \mathrm{d}\Delta_x \right\} \cdot \left\{ \int L_j^y(r_y + \Delta_y) L_m^y(r'_y + \Delta_y) p_y(\Delta_y) \mathrm{d}\Delta_y \right\} \\ &\qquad \cdot \left\{ \int R_j(r_z + \Delta_z) R_m(r'_z + \Delta_z) p_z(\Delta_z) \mathrm{d}\Delta_z \right\} \end{aligned}$$

into

$$Cov(D(\boldsymbol{r}, \boldsymbol{\Delta}), D(\boldsymbol{r'}, \boldsymbol{\Delta})) = \mathbb{E}_{p_\Delta}[D(\boldsymbol{r}, \boldsymbol{\Delta})D(\boldsymbol{r'}, \boldsymbol{\Delta})] - \mathbb{E}_{p_\Delta}[D(\boldsymbol{r}, \boldsymbol{\Delta})]\mathbb{E}_{p_\Delta}[D(\boldsymbol{r'}, \boldsymbol{\Delta})].$$

Since all terms are Gaussian, the integrals can be solved analytically and the expected value and covariance of the dose can be computed directly. While this is very efficient, it relies on strong assumptions about both the dose deposition function and the uncertainties. The underlying pencil beam method is known to have difficulties with heterogeneities, near which the variance due to uncertainties is however especially high. While this could e.g. be alleviated by combining the approach with methods like fine-sampling, this is outside the scope of this thesis. For this reason, APM will only be used for validation in homogeneous water test cases.

In the following two chapters, we introduce two novel approaches developed within the context of this thesis to improve the efficiency of uncertainty propagation in Monte Carlo as well as grid-based dose computations.

CHAPTER 14

Importance (re-)weighting for Monte Carlo dose simulations

First, we introduce a (re-)weighting approach to speed up non-intrusive uncertainty propagation in Monte Carlo dose calculations following [203, 206].

The proposed method generates estimates of the expected value and variance of radiation therapy doses from a single MC simulation run. It exploits the probabilistic nature of regular dose calculations to frame uncertainties in the phase space parameters as a change of measure. The approach then uses the concept of importance sampling [95, 71] and mimics the computation of scenarios in the parameter space by (re-)weighting the results of a single realization. The computational overhead of a complex simulation is thus reduced to that of the (re-)weighting step, effectively making the uncertainty quantification problem a scoring problem independent of the underlying physical simulation.

Similar strategies have been previously employed, e.g., to estimate the average of samples from different distributions [15, 214], for more efficient updates in optimization [196] and reinforcement learning [176, 69] or to artificially increase the sample size in costly simulations [208]. However, to the best of our knowledge, there is no previous research on computing a conditional variance by (re-)weighting a single original sample. Also, in the field of medical physics and in particular dose calculation, we are not aware of any previous use of importance (re-)weighting for the reuse of samples, neither for dose estimates (which are averages) nor for the variance with respect to input uncertainties.

In the following, we first introduce the problem of particle transport with uncertain input in the context of Monte Carlo dose calculations. We then show how to use importance sampling to quantify uncertainties (section 14.3), discuss mathematical properties of this approach (section 14.6) and lastly show results for a water phantom as well as several patient cases and different uncertainty models (section 14.9).

Note, that the proposed method falls into the aforementioned category of minimally intrusive methods. Although the approach is based on non-intrusive methods, it is not completely non-intrusive as it

requires access to the probability distributions of input parameters as well as the accumulation of quantities of interest such as the dose. The underlying equations are however not affected and even proprietory code frameworks typically allow access to superficial parts of the code like input configuration and scoring functions. Therefore, our approach does not suffer from the typical drawbacks of a non-intrusive method.

14.1 Solution of the uncertain problem using Monte Carlo

When considering non-intrusive UQ methods, the Monte Carlo simulation of a particle history can be treated as a black box $BB(\boldsymbol{r}; \boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)$, returning the dose $D(\boldsymbol{r})$ at position $\boldsymbol{r}$ for the random initial phase space $\boldsymbol{Z} = (\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0) \sim S_0$ on $\mathcal{D}_{S_0}$ (compare section 5.1).

$$D(\boldsymbol{r}) = \mathbb{E}_{S_0}[BB(\boldsymbol{r}; \boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)] = \int_{\mathcal{D}_{S_0}} BB(\boldsymbol{r}; \boldsymbol{Z}) S_0(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} \approx \frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{r}; \boldsymbol{z_p}) \,, \tag{14.1}$$

where the realizations $\boldsymbol{z}_1, ..., \boldsymbol{z}_H$ are sampled from $S_0(\boldsymbol{Z})$.

Recall, that the black box for fixed realizations of the input parameters still describes a stochastic process, due to subsequent sampling of the path and interactions of the particles. The input sampled from the source distribution merely fixes the initial properties of released particles. Each response $BB(\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0)$ however also implicitly includes realizations of the probability distributions governing the interactions of these particles. Thus, the Monte Carlo method is applied to the complete random particle trajectory, we however omit this, since we are only interested in uncertainties affecting the phase space at a superficial level in the source distribution.

Most non-intrusive strategies involve computing the dose at different points in the parameter space. In order to understand how non-intrusive methods are implemented when using MC dose calculation algorithms, we now consider the Monte Carlo method, also for uncertainties. Here, these points are realizations randomly drawn from the probability distribution p_Δ of the uncertain variable $\boldsymbol{\Delta}$.

Note, that the introduction of an uncertain parameter adds a second stochastic process around the black box transport solver. On the one hand, the dose distribution depends on random input parameters following a distribution defined by the physical radiation source. On the other hand, this distribution now depends on the random error vector $\boldsymbol{\Delta}$. Thus, if we want to compute the expected value $\mathcal{D}$ of the dose with respect to both processes, we are looking for

$$\mathcal{D}(\boldsymbol{r}) = \mathbb{E}_{p_\Delta}\left[\mathbb{E}_{S_{\boldsymbol{Z}|\boldsymbol{\Delta}}}\left[BB(\boldsymbol{r}; \boldsymbol{Z}(\boldsymbol{\Delta})) \mid \boldsymbol{\Delta} = \boldsymbol{\delta}\right]\right].$$

The Monte Carlo method can now be applied to both random processes. Starting with the error variable, for a fixed realization, the black box MC solver can be applied analogously to the original problem. Thus, the dose estimate $D_i(\boldsymbol{r})$ for realization $\boldsymbol{\delta}_i$ and the corresponding conditional distribution

$$S_i := \mathbb{P}(\boldsymbol{Z}(\boldsymbol{\Delta}) \mid \boldsymbol{\Delta} = \boldsymbol{\delta}_i)$$

is given by

$$D_i(\boldsymbol{r}) = \mathbb{E}_{S_i}[BB(\boldsymbol{r}; \boldsymbol{Z}(\boldsymbol{\Delta})) \mid \boldsymbol{\Delta} = \boldsymbol{\delta}_i] = \int_{\mathcal{D}_{S_i}} BB(\boldsymbol{r}; \boldsymbol{Z}) S_i(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} \approx \frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{r}; \boldsymbol{z}_p^i) \tag{14.2}$$

with $\boldsymbol{z}_1^i, ..., \boldsymbol{z}_H^i$ sampled from S_i. The Monte Carlo estimate of the expected value with respect to both processes is then computed as the mean over $D_1(\boldsymbol{r}), ..., D_N(\boldsymbol{r})$ for $\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_N$ sampled from p_Δ

$$\mathcal{D}(\boldsymbol{r}) \approx \frac{1}{N} \sum_{i=1}^{N} D_i(\boldsymbol{r}) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{r}; \boldsymbol{z}_p^i) \,, \tag{14.3}$$

and the variance as the sample variance, defined as

$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})^2 \;\text{ where } \bar{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x_i} \,, \tag{14.4}$$

for $\boldsymbol{x_i},\ i = 1, ..., N$ that are independent and identically distributed. This is an unbiased estimator of the variance of the random variable $\boldsymbol{X}$. Applied to the problem at hand, this results in the following formula for the dose variance:

$$\begin{aligned} Var_{p_{\boldsymbol{\Delta}}}(D(\boldsymbol{r}, \boldsymbol{\Delta})) &\approx \frac{1}{N-1} \sum_{i=1}^{N} \left(D_i(\boldsymbol{r}) - \mathcal{D}(\boldsymbol{r})\right)^2 \\ &\approx \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{z}_p^i) - \frac{1}{N} \sum_{j=1}^{N} \frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{z}_p^j) \right)^2 . \end{aligned} \tag{14.5}$$

14.2 Direct computation of the expected value

When the conditional and marginal distributions of $\boldsymbol{Z}$ and $\boldsymbol{\Delta}$ are known, according to Bayes' theorem their joint distribution $\mathcal{S}(\boldsymbol{Z}(\boldsymbol{\Delta}))$ can also be determined. It is then possible to compute the expected dose directly

$$\mathbb{E}_{p_{\boldsymbol{\Delta}}}[D(\boldsymbol{r}, \boldsymbol{\Delta})] = \mathbb{E}_{\mathcal{S}}[BB(\boldsymbol{Z}(\boldsymbol{\Delta}))] \approx \frac{1}{H} \sum_{p=1}^{H} BB(\boldsymbol{z}_p) \,, \tag{14.6}$$

with $\boldsymbol{z}_p,\ p = 1, .., H$ sampled from $\mathcal{S}$.

14.3 (Re-)weighting Monte Carlo samples for uncertainty quantification

In the following section, we introduce the proposed (re-)weighting approach. We exploit the fact, that the input vector $\boldsymbol{Z}$ is non-deterministic, i.e., has the probability density $S_0(\boldsymbol{Z})$ regardless of uncertainties. Uncertainties in the phase space parameters usually manifest in a way that merely changes this distribution. E.g., when using multivariate normal distributions the mean is shifted or the standard deviation increased in the dimensions corresponding to an uncertain phase space parameter. The concept

of importance sampling can be used to gain insight into dose realizations for different input distributions $S_i(\boldsymbol{Z})$ corresponding to error realizations $\boldsymbol{\delta}_i$, $i = 1, ..., N$ from p_Δ, from one initial simulation.

We therefore now recall the basic principles of importance sampling.

14.4 Standard importance sampling

Importance sampling is a method originally introduced by [95] and later made popular in a more general context by [71].

Let's assume we are interested in $g(Z)$. Let Z have the probability distribution function $p(Z)$ with support $\mathcal{D}_p$, i.e., $z \sim p(Z)$. Then

$$\int_{\mathcal{D}_p} g(Z)p(Z)\mathrm{d}Z = \int_{\mathcal{D}_p} g(Z)\frac{p(Z)}{q(Z)}q(Z)\mathrm{d}Z \,. \tag{14.7}$$

So the integral can be approximated as

$$\int_{\mathcal{Q}} g(Z)\frac{p(Z)}{q(Z)}q(Z)\mathrm{d}Z \approx \sum_{i=1}^{N} g(z_i)\frac{p(z_i)}{q(z_i)} \,. \tag{14.8}$$

Where z_i, $i = 1, ..., N$ are realisations of $Z \sim q(Z)$ and q is a distribution with support $\mathcal{Q} \supseteq \mathcal{D}_p$.

Thus, the computation of $g(Z)$ for realisations of $p(Z)$ can be replaced by computations of $g(Z)$ with realisations z_i, $i = 1, ..., N$ from $q(Z)$, using additional weights $\frac{p(z_i)}{q(z_i)}$ in the aggregation of the results. This can be convenient if it is easier to sample from $q(Z)$ than $p(Z)$ or because choosing $q(Z)$ cleverly can reduce the variance of the estimate. It can also be helpful when evaluating g is much more expensive than evaluating $\frac{p(z_i)}{q(z_i)}$, for example, because g is an expensive simulation. In this case, if we are interested in $g(Z)$ for different distributions of Z, we can reuse the realizations $g(z_1), ..., g(z_N)$.

14.5 Importance (re-)weighting

For our purpose, this can be applied to mimic dose computations at different points in the parameter space from a previously computed sample of particle trajectories. In the following, we will assume random samples, the method can however be analogously applied to quadratures or quasi-random numbers. We further assume that the uncertain parameter $\boldsymbol{\Delta}$ is independent of the initial phase space parameters $\boldsymbol{Z}$.

Instead of solving the transport problem N times with slightly different input distributions $S_1, ..., S_N$ (compare eq. (14.2)), it is solved once and the results for other error realizations are determined by (re-)weighting the deposited dose of this one sample of particles as explained in section 14.4. For an error realization $\boldsymbol{\delta}_i$ from p_Δ, initial parameter realizations $\boldsymbol{z}_1, ..., \boldsymbol{z}_H$ from the sampling distribution $q(\boldsymbol{Z})$ and

target distributions $S_i(\boldsymbol{Z})$, the dose is given by

$$D_i(\boldsymbol{r}) = \int_{\mathcal{D}_{S_i}} BB(\boldsymbol{r};\boldsymbol{Z}) S_i(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} = \int_{\mathcal{D}_{S_i}} \frac{BB(\boldsymbol{r};\boldsymbol{Z})\ S_i(\boldsymbol{Z})}{q(\boldsymbol{Z})}\ \cdot q(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}$$
$$\approx \frac{1}{H}\sum_{p=1}^{H} \frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\ S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}\ . \tag{14.9}$$

This yields the following procedure:

Algorithm 14.1 Calculate doses D_i

(a) On-the-fly

for particles p=1:H **do**
  Sample $\boldsymbol{z}_p \leftarrow q(\boldsymbol{Z})$
  Compute $BB(\boldsymbol{r};\boldsymbol{z}_p)$
  Sample $\boldsymbol{\delta}_i \leftarrow p_\Delta(\boldsymbol{\Delta})$
  for target distributions $S_i(\boldsymbol{Z})$,
  $i = 1, ..., N$ **do**
    D_i += $\frac{1}{H}\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\ S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}$
  end for
end for

(b) In post-processing

for particles p=1:H **do**
  Sample $\boldsymbol{z}_p \leftarrow q(\boldsymbol{Z})$
  Compute and store $BB(\boldsymbol{r};\boldsymbol{z}_p)$
end for
Sample $\boldsymbol{\delta}_i \leftarrow p_\Delta(\boldsymbol{\Delta})$
for target distributions $S_i(\boldsymbol{Z})$,
$i = 1, ..., N$ **do**
  $D_i = \frac{1}{H}\sum_{p=1}^{H} \frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\ S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}$
end for

Here, the on-the-fly computation in algorithm 14.1 (a) has the advantage of not requiring the storage of particle histories. However, in this case the target distributions have to be defined a-priori, whereas the post-processing routine (algorithm 14.1 (b)) can be applied for different uncertainty models and quantities of interests after the initial simulation.

The importance and target distributions are chosen according to the quantities of interest. Here, we show the derivation for the nominal dose without uncertainty, the expected dose and the dose variance with respect to uncertainties.

14.5.1 Nominal dose

The nominal dose is the solution to the problem introduced in section 4.1 without uncertainty, i.e., for the input distribution $S_0(\boldsymbol{Z})$. Let $q(\boldsymbol{Z})$ be the simulated particle distribution, then the nominal dose can be computed by replacing $p(\boldsymbol{Z})$ with $S_0(\boldsymbol{Z})$ in section 14.4.

$$D(\boldsymbol{r}) = \mathbb{E}_{S_0}[BB(\boldsymbol{r};\boldsymbol{Z})] = \int_{\mathcal{D}_{S_0}} BB(\boldsymbol{r};\boldsymbol{Z}) S_0(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} = \int_{\mathcal{D}_{S_0}} \frac{BB(\boldsymbol{r};\boldsymbol{Z}) S_0(\boldsymbol{Z})}{q(\boldsymbol{Z})}\ \cdot q(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} \tag{14.10}$$
$$\approx \frac{1}{H}\sum_{p=1}^{H} \frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\ S_0(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}$$

This reduces to the regular dose computation with Monte Carlo described in eq. (5.6), when q is chosen to be S_0.

14.5.2 Expected value

Using eq. (14.9) to compute the dose for individual error realizations, we can mimic the Monte Carlo method for uncertainties and compute the expected dose as

$$\mathcal{D}(\boldsymbol{r}) \approx \frac{1}{N}\sum_{i=1}^{N} D_i(\boldsymbol{r}) \approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{H}\sum_{p=1}^{H}\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\, S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}\ , \tag{14.11}$$

with $\boldsymbol{\delta}_i$, $i = 1, ..., N$ sampled from $S_i = \mathbb{P}(\boldsymbol{Z}(\boldsymbol{\Delta}) \mid \boldsymbol{\Delta} = \boldsymbol{\delta}_i)$ and $BB(\boldsymbol{r};\boldsymbol{z}_p)$ again the pre-computed particle trajectories with input parameters $\boldsymbol{z}_p$, $p = 1, ..., H$ sampled from q.

Using the joint distribution of the uncertain input and initial phase space parameters, it is possible to directly compute the expected value, without sampling from p_Δ, by adapting eq. (14.6)

$$\mathcal{D}(\boldsymbol{r}) = \mathbb{E}_{\mathcal{S}}[BB(\boldsymbol{r};\boldsymbol{Z}(\boldsymbol{\Delta}))] \approx \frac{1}{H}\sum_{p=1}^{H}\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\, \mathcal{S}(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}\ , \tag{14.12}$$

where $\boldsymbol{z}_p,\ p = 1, ..., H$ sampled from $\mathcal{S}$.

14.5.3 Variance

The variance can also be computed for error realizations $\boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_N$ from p_Δ, using dose realizations from eq. (14.11) with the sample variance formula

$$\sigma^2 = Var_{p_{\boldsymbol{\Delta}}}(D(\boldsymbol{r},\boldsymbol{\Delta})) \approx \frac{1}{N-1}\sum_{i=1}^{N}\left(D_i(\boldsymbol{r}) - \mathcal{D}(\boldsymbol{r})\right)^2\ . \tag{14.13}$$

14.5.4 Choosing the importance distribution

The optimal choice of $q(\boldsymbol{Z})$ depends largely on the target quantities. It follows from sections 14.5.1 to 14.5.3 that the nominal dose, expected dose and the error scenarios required for the variance computation correspond to $N + 2$ different distributions of the input parameters. Thus, when all of these quantities are computed from the same set of realizations, $q(\boldsymbol{Z})$ needs to be suitable for not just one but $N+2$ different target distributions. Therefore, it will most likely not be possible to choose a function that is optimal for each individual computation and the choice depends on how the accuracy of the different target quantities is prioritized.

For a given target distribution $p(\boldsymbol{Z})$, there are two criteria by which $q(\boldsymbol{Z})$ should be chosen (see e.g. [165]):

- Let $Q = \{\boldsymbol{Z} \mid q(\boldsymbol{Z}) > 0\}$, if $p(\boldsymbol{Z}) \cdot BB(\boldsymbol{r};\boldsymbol{Z}) \neq 0$ then $\boldsymbol{Z} \in Q$

- To minimize the standard error of the approximation $\sigma_q^2 = \frac{1}{H}\sum_{p=1}^{H}\left[\frac{p(\boldsymbol{z}_p)\cdot BB(\boldsymbol{r};\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)} - \mu\right]^2$ with $\mu = \mathbb{E}_{S_0}[BB(\boldsymbol{r};\boldsymbol{Z})]$, $q(\boldsymbol{Z})$ should be chosen to be proportional to $p(\boldsymbol{Z}) \cdot BB(\boldsymbol{r};\boldsymbol{Z})$

The first point is jointly attainable for all target values, for example, if the support of $q(\boldsymbol{Z})$ includes the supports of $S_1(\boldsymbol{Z}), ..., S_N(\boldsymbol{Z})$ as well as those of S_0 and $\mathcal{S}$.

The second point creates a trade-off between the different targets. Choosing $q(\boldsymbol{Z}) = \mathcal{S}$ or $q(\boldsymbol{Z}) = S_0$, leads to accurate estimates of the expected dose and nominal dose respectively. It can be expected, that $q(\boldsymbol{Z}) = \mathcal{S}$ is a better choice with respect to the dose variance since it has more density in the outer regions, which is necessary for an acceptable estimate of the shifted densities S_i for larger errors $\boldsymbol{\delta}_i$.

When prioritizing the computation of the variance estimate, a more fat-tailed choice of distribution for $q(\boldsymbol{Z})$ is beneficial. This ensures a larger number of realizations and therefore higher accuracy in the outer regions of the densities, where the variance is typically high. In general, it is not advisable to choose a q which is more light-tailed than the target distribution, as the variance can explode if q is too small in regions where the target distribution is not zero [165]. One option is using a mixture distribution of the individual shifted distributions we want to reconstruct [75]. Another common rule of thumb in case of a normal target distribution is to use a Student's t distribution, as they are similar but the latter has heavier tails [165]. The accuracy of estimates for some exemplary choices of q will be investigated numerically in section 14.9.4.

14.6 Mathematical properties

14.6.1 Accuracy & bias

An advantage of the proposed method is that it inherits some of the mathematical properties of importance sampling. Namely, for the nominal dose, expected dose and dose for the single error realizations, the estimates are unbiased [165].

Proposition 14.1. *For $q(\boldsymbol{Z}) : \mathcal{D}_Z \to \mathbb{R}_+$ with $Q = \{\boldsymbol{Z} \mid q(\boldsymbol{Z}) > 0\}$, if $p(\boldsymbol{Z}) \cdot BB(\boldsymbol{r};\boldsymbol{Z}) \neq 0$ let $\boldsymbol{Z} \in Q$. Then the importance reweighting estimator $\hat{D}(\boldsymbol{r})$ is unbiased, i.e. $\mathbb{E}_q[\hat{D}(\boldsymbol{r})] = D(\boldsymbol{r})$.*

Proof: For $\boldsymbol{Z} \sim q(\boldsymbol{Z})$ it follows by construction of the importance sampling estimator that

$$
\begin{aligned}
&\mathbb{E}_q\left[\frac{BB(\boldsymbol{r};\boldsymbol{Z})\, p(\boldsymbol{Z})}{q(\boldsymbol{Z})}\right] \\
&= \int_{\mathcal{Q}} \frac{BB(\boldsymbol{r};\boldsymbol{Z})\, p(\boldsymbol{Z})}{q(\boldsymbol{Z})} \cdot q(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} = \int_{\mathcal{Q}} BB(\boldsymbol{r};\boldsymbol{Z}) \cdot p(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} \\
&= \int_{\mathcal{D}_p} BB(\boldsymbol{r};\boldsymbol{Z}) \cdot p(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} + \int_{\mathcal{Q}\cap\mathcal{D}_p^c} \underbrace{BB(\boldsymbol{r};\boldsymbol{Z})}_{=0} \cdot p(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} + \int_{\mathcal{Q}^c\cap\mathcal{D}_p} BB(\boldsymbol{r};\boldsymbol{Z}) \cdot \underbrace{p(\boldsymbol{Z})}_{=0}\mathrm{d}\boldsymbol{Z} \\
&= \int_{\mathcal{Q}} BB(\boldsymbol{r};\boldsymbol{Z}) \cdot p(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} = \mathbb{E}_p[BB(\boldsymbol{Z})]
\end{aligned}
$$

Then the importance reweighting estimator

$$
\hat{D}(\boldsymbol{r}) = \frac{1}{H}\sum_{p=1}^{H} \frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\, p(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)} \ ,
$$

where $\boldsymbol{z}_p \sim q,\ p = 1, ..., H$ are independent and identically distributed random variables, is obtained by Monte Carlo integration and the unbiasedness can be shown analogously

$$\begin{aligned}\mathbb{E}_q[\hat{D}(\boldsymbol{r})] &= \mathbb{E}_q\left[\frac{1}{H}\sum_{p=1}^{H}\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\,p(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}\right] = \frac{1}{H}\sum_{p=1}^{H}\mathbb{E}_q\left[\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\,p(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}\right] \\ &= \frac{1}{H}\sum_{p=1}^{H}\int_{\mathcal{Q}}\frac{BB(\boldsymbol{r};\boldsymbol{Z})\,p(\boldsymbol{Z})}{q(\boldsymbol{Z})}\cdot q(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} \\ &= \mathbb{E}_p[BB(\boldsymbol{r};\boldsymbol{Z})] = D(\boldsymbol{r}).\end{aligned}$$

□

Also, the standard error ϵ_q of these estimators can be computed similarly to that of Monte Carlo estimates [165]. It depends on the number of realizations as well as the similarity of the importance distribution and the distribution of interest:

$$\epsilon_q = \frac{\sigma_q^2}{H}\text{, where } \sigma_q^2 \approx \hat{\sigma}_q^2 = \frac{1}{H-1}\sum_{p=1}^{H}\left[\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)p(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)} - \hat{D}(\boldsymbol{r})\right]^2. \tag{14.14}$$

To lower the standard error, the importance distribution should resemble the target distribution. In the variance computation, the estimates for more extreme error realizations will thus be less accurate. For the type and extent of errors in radiation therapy, the distributions are however mostly similar enough, as we will demonstrate further in section 14.9.

In contrast to the dose estimates, the variance computed as in section 14.5.3 is not generally unbiased. This is due to its computation from estimates of the dose for shifted beams, which rely on the same single sample of particles. What we consider as the realizations in the sample variance computation, while individually unbiased, are thus no longer independent.

Proposition 14.2. *For $\boldsymbol{z}_p$ from $S_i = \mathbb{P}(\boldsymbol{Z}(\boldsymbol{\Delta}) \mid \boldsymbol{\Delta} = \boldsymbol{\delta}_i)$ and $\boldsymbol{\delta}_i \sim p_\Delta$, let $w_i(\boldsymbol{z}_p) = \frac{S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}$ be conditionally independent of $BB(\boldsymbol{r};\boldsymbol{z}_p)$ given $\boldsymbol{z}_p$. Then the importance reweighting estimator for the dose variance is not generally unbiased and the bias is given by*

$$b_{S_N^2} = \left(\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\cdot\mathbb{E}_{p_\Delta}[w_i(\boldsymbol{z}_p)]\right)^2 - \frac{1}{H^2}\sum_{p=1}^{H}\sum_{h=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)BB(\boldsymbol{r};\boldsymbol{z}_h)\cdot\mathbb{E}_{p_\Delta}[w_i(\boldsymbol{z}_p)w_j(\boldsymbol{z}_h)]$$

Proof: In [16], general formulas for the moments of the sample variance S_n^2 are derived. The first moment $\mathbb{E}[S_n^2]$ for general non independent or identically distributed (iid) samples $X_1, ..., X_N$ is given by:

$$\mathbb{E}[S_n^2] = \frac{\sum_{i=1}^{n}\mathbb{E}[X_i^2]}{n} - \frac{\sum_{i\neq j}\mathbb{E}[X_iX_j]}{n(n-1)}.$$

Applied to our context, the realizations are $D_i = \frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)w_i(\boldsymbol{z}_p)$ for $i = 1, ..., N$, with $w_i(\boldsymbol{z}_p) = \frac{S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}$. Since all realizations use the same sample of computed particle trajectories $BB(\boldsymbol{r};\boldsymbol{z}_p)$ and initial particle properties $\boldsymbol{z}_p,\ p = 1, ..., H$, they cannot be assumed to be independent, in particular $D_i = D_j + \frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\,(w_i(\boldsymbol{z}_p) - w_j(\boldsymbol{z}_p))$. Thus,

$$\mathbb{E}[S_n^2] = \frac{\sum_{i=1}^{N} \mathbb{E}[D_i^2]}{N} - \frac{\sum_{i \neq j} \mathbb{E}[D_i D_j]}{N(N-1)} \,.$$

Since we assume that the D_i are identically distributed, this reduces to

$$\begin{aligned}\mathbb{E}[S_n^2] &= \mathbb{E}_{p_\Delta}[D_i^2] - \mathbb{E}_{p_\Delta}[D_i D_j] \\ &= \underbrace{\mathbb{E}_{p_\Delta}[D_i^2] - \mathbb{E}_{p_\Delta}[D_i]\mathbb{E}_{p_\Delta}[D_j]}_{=Var(D)} \underbrace{- \left(\mathbb{E}_{p_\Delta}[D_i D_j] - \mathbb{E}_{p_\Delta}[D_i]\mathbb{E}[D_j]\right)}_{=b_{S_N^2} \neq 0 \text{ , for } i \neq j} \,.\end{aligned}$$

Thus the estimator is not necessarily unbiased and the bias is

$$\begin{aligned}b_{S_N^2} &= \underbrace{\mathbb{E}_{p_\Delta}[D_i]\mathbb{E}_{p_\Delta}[D_j]}_{=\mathbb{E}_{p_\Delta}[D_i]\mathbb{E}_{p_\Delta}[D_i]} - \mathbb{E}_{p_\Delta}[D_i D_j] \\ &= \mathbb{E}_{p_\Delta}\left[\frac{1}{H}\sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p) w_i(\boldsymbol{z}_p)\right]^2 - \mathbb{E}_{p_\Delta}\left[\left(\frac{1}{H}\sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p) w_i(\boldsymbol{z}_p)\right)\left(\frac{1}{H}\sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p) w_j(\boldsymbol{z}_p)\right)\right] \\ &= \left(\frac{1}{H}\sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p)\mathbb{E}_{p_\Delta}\left[w_i(\boldsymbol{z}_p)\right]\right)^2 - \mathbb{E}_{p_\Delta}\left[\frac{1}{H^2}\sum_{p=1}^{H}\sum_{h=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p) BB(\boldsymbol{r};\boldsymbol{z}_h) w_i(\boldsymbol{z}_p) w_j(\boldsymbol{z}_h)\right] \\ &= \left(\frac{1}{H}\sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p)\mathbb{E}_{p_\Delta}\left[w_i(\boldsymbol{z}_p)\right]\right)^2 - \frac{1}{H^2}\sum_{p=1}^{H}\sum_{h=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p) BB(\boldsymbol{r};\boldsymbol{z}_h)\mathbb{E}_{p_\Delta}\left[w_i(\boldsymbol{z}_p) w_j(\boldsymbol{z}_h)\right] \,.\end{aligned}$$

□

Since this bias term is costly to compute, it is generally not feasible to score this in addition to the quantities of interest. Computing the bias once for a given problem class could however be of interest to judge its order of magnitude. Its potential impact on our results will be assessed in greater detail in section 14.9.

14.6.2 Bounds on the variance

While the discussed estimators can be expected to significantly lower the expense of scenario computations, the variance still requires N scoring operations, either accumulated on the fly after each simulated particle history, or in post-processing as (re-)weighting of stored histories. For a large number of simulated particles, run times can therefore exceed the required time frame, especially in the context of robust optimization, where the variance has to be computed in each of numerous iterations [178, 130, 38].

Proposition 14.3. *An upper bound to the variance estimator $\hat{\sigma}^2$ is given by*

$$\hat{\sigma}^2 \leq \left(\frac{1}{H}\sum_{p=1}^{H} BB(\boldsymbol{r};\boldsymbol{z}_p)^2\right)\left(\frac{1}{H}\sum_{p=1}^{H} S_N^2(w(\boldsymbol{z}_p))\right)$$

Proof: Using the definition of the importance reweighing estimator for the dose variance, the dose for a single error realization and the expected dose (see eqs. (14.9), (14.12) and (14.13)), we get

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{1}{H}\sum_{p=1}^{H}\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\,S_i(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)} - \frac{1}{N}\sum_{j=1}^{N}\frac{1}{H}\sum_{p=1}^{H}\frac{BB(\boldsymbol{r};\boldsymbol{z}_p)\,S_j(\boldsymbol{z}_p)}{q(\boldsymbol{z}_p)}\right]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\,w_i(\boldsymbol{z}_p) - \frac{1}{N}\sum_{j=1}^{N}\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\,w_j(\boldsymbol{z}_p)\right]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\,w_i(\boldsymbol{z}_p) - \frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\frac{1}{N}\sum_{j=1}^{N}w_j(\boldsymbol{z}_p)\right]^2 \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)\left(w_i(\boldsymbol{z}_p) - \frac{1}{N}\sum_{j=1}^{N}w_j(\boldsymbol{z}_p)\right)\right]^2 \\
&\leq \left(\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)^2\right)\left(\frac{1}{H}\sum_{p=1}^{H}\underbrace{\frac{1}{N-1}\sum_{i=1}^{N}\left[w_i(\boldsymbol{z}_p) - \sum_{j=1}^{N}w_j(\boldsymbol{z}_p)\right]^2}_{=S_N^2(w(\boldsymbol{z}_p))\approx Var(w(\boldsymbol{z}_p))}\right) \\
&= \left(\frac{1}{H}\sum_{p=1}^{H}BB(\boldsymbol{r};\boldsymbol{z}_p)^2\right)\left(\frac{1}{H}\sum_{p=1}^{H}S_N^2(w(\boldsymbol{z}_p))\right)\,.
\end{aligned}
$$

□

The advantage of this bound is, that the squared simulation response $BB(\boldsymbol{r};\boldsymbol{z}_p)^2$ is aggregated independently from the error realizations. This includes the implicit assumption, that the variance in the likelihood ratio between the simulated and target distribution, is enough to describe the variance of the output. While this might hold for relatively homogeneous background media, it is unclear whether it is sufficient in heterogeneous media as present in many patient anatomies.

14.7 Uncertainty models

In the following, we consider different uncertainty models introduced in chapter 11. We again use Gaussian beam and uncertainty assumptions. Then for set-up uncertainties, the required probability distributions $S_0(\mathbf{Z})$, $S_i(\mathbf{Z})$ and $\mathcal{S}$ corresponding to the nominal dose, individual error realizations and the joint phase space and error distribution, respectively, are derived in section 11.1. For range uncertainties, we use the approximation based on the initial beam energy introduced in section 11.2. Different correlation assumptions as well as time-dependent spatial uncertainties are included using the covariance matrix of the Gaussian beamlet distribution as discussed in sections 11.3 and 11.4.

14.8 Implementation and test cases

The described method was implemented as post-processing in Matlab. Radiation plans were generated for proton dose calculations using the treatment planning software matRad [234] and dose calculations were performed with the Monte Carlo engine TOPAS [175]. For quicker convergence, we use randomized

quasi-Monte Carlo with scrambled Sobol numbers when sampling the error scenarios for the dose variance estimate (see section 12.2).

In the following, we present the results for a 3D water phantom as well as a prostate and liver patient obtained from the open CORT (common optimization for radiation therapy) data set [42]. Overall, we consider three different treatment plans, combined with several error models discussed in chapter 11. For the Gaussian set-up error, we assume a standard deviation of 3 mm and mean of $\mu_\delta = 0$, i.e., the errors do not have a systematic component [comp. 174]. The range error is modeled using the approximation introduced in section 11.2, with a standard deviation of 3 % and mean of 0. For three of the treatment plans, intensity-modulated proton therapy (IMPT) [96, 131] was employed. In these cases, each source consists of a large number of narrow pencil beams to enable better and more flexible coverage of the target volume. More details concerning the treatment plans and the respective error models applied to each of them can be found in table 14.1.

Table 14.1: Overview of simulated plans and error models per test case.

Test case	Water phantom		Prostate	Liver
Size	60x25x25 voxels		183x183x90 voxels	217x217x168 voxels
Irradiation angles	(0°,0°)		(0°,90°)/(0°,270°)	(0°,315°)
Pencil beams	1	175	1 375/1 383	1408
Simulated particles	100 000	2 566 453	16 992 193/16 748 034	13 528 430
Error model	global	global/AR(1)/GP	global, simple models (fig. 11.1)	global/AR(1)/GP
Sample size	500	500	100	100

Note, that in this chapter, we only use a Gaussian distribution for the spatial and energy variable of the phase space. The initial beam direction $\mathbf{\Omega}_0$ is assumed to be deterministic. The presented concepts can however be used analogously for a fully random initial phase space. A brief discussion of this case and additional results can be found in the appendix (A4).

14.8.1 Evaluation criteria

To evaluate the quality of our estimates, we compute a reference for each test case using the regular randomized quasi-Monte Carlo method for the uncertainties. The number of realizations was chosen to be equal to that used for the importance (re-)weighting estimates, which is 500 for the water phantom and 100 for the prostate patient.

We compare the results according to the γ criterion (see section 3.3), as well as a difference map (see section 3.4).

14.9 Results

14.9.1 Fully correlated pencil beams

Figure 14.1 presents the expected dose and standard deviation for a single pencil beam in a water phantom. Figure 14.2 shows results for the prostate patient with fully correlated pencil beams, i.e., one global error. The importance distribution was chosen to be the nominal distribution S_0. For this reason, the nominal dose estimate is just a regular Monte Carlo estimate and is therefore omitted in the following.

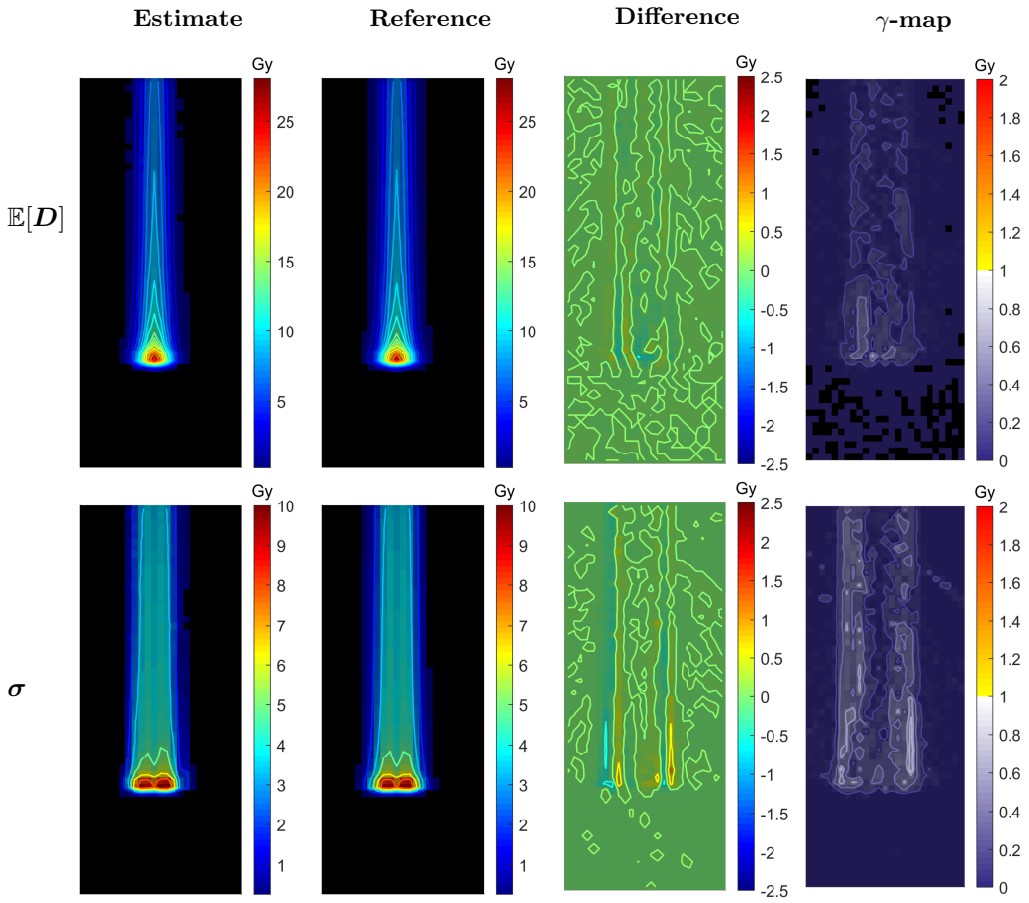


Figure 14.1: Expected dose and standard deviation w.r.t. set-up uncertainties with 3 mm standard deviation for one pencil beam in a water box. The columns from left to right show the estimate computed with the proposed (re-)weighting approach, a reference computed with randomized quasi-MC, a difference plot and a plot of the $\gamma_{3\,\mathrm{mm}/3\,\%}$-indices.

The results using the importance (re-)weighting estimate are in good agreement with the reference, with a $\gamma_{3\,\mathrm{mm}/3\,\%}$-passrate of 100 % for the expected value and over 99.9 % with respect to the standard deviation in both waterbox and prostate patient. Additional results for the liver case with a full Gaussian phase space distribution, i.e. a distribution also in the angular variable, can be found in the appendix (A1).

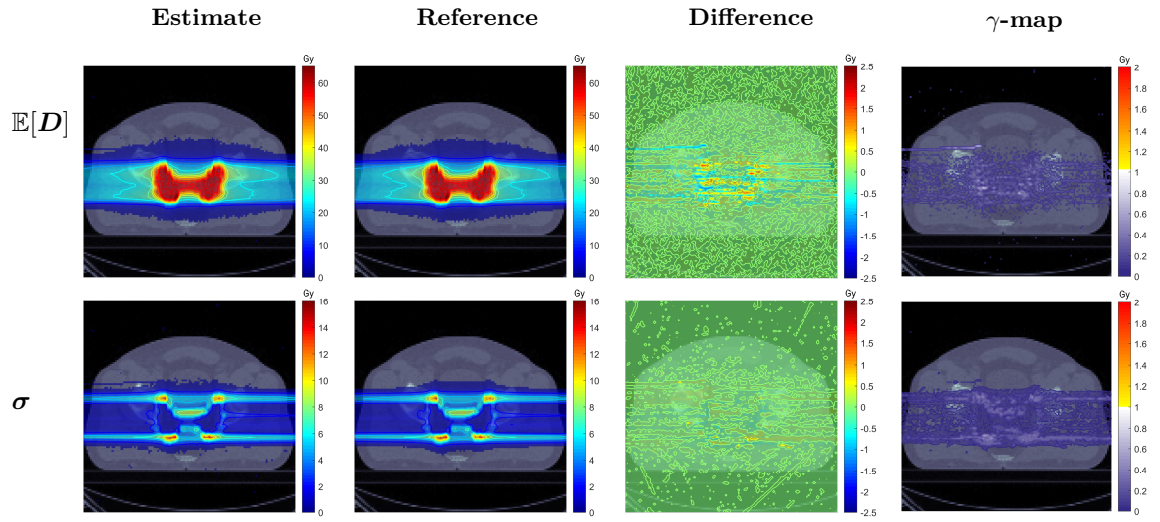


Figure 14.2: Expected dose and standard deviation w.r.t. set-up uncertainties with 3 mm standard deviation for two beams (gantry angles 90° and 270°, couch angle 0°) in a prostate patient. The columns from left to right show the estimate computed with the proposed (re-)weighting approach, a reference computed with randomized quasi-MC, a difference plot and a plot of the $\gamma_{3\,\mathrm{mm}/3\,\%}$-indices.

14.9.2 Range errors

In contrast to the set-up errors, for which dose estimates can also be shown to be mathematically accurate, range errors can only be modeled through an approximation introduced in section 11.2. Figure 14.3 displays results for range errors as well as the combination of range and set-up errors in the water phantom.

The difference maps for both expected value and standard deviation show that the deviations when including range errors are expectedly higher. We observe a systematic bias primarily at the distal edge, where our method seems to consistently underestimate the variance. The standard deviation estimate using our importance weighting method also expresses strong local artifacts, as evident in the difference maps (compare figure 14.3). This is an indicator of too little statistical mass, i.e., computed particle trajectories, in the original simulation. For more extreme error realizations, relatively high weights are assigned to a small number of particles, thereby amplifying single realizations or errors. Especially in the case of relatively small beam energy spread in the original simulation (here 1 %), compared to the range error of 3 %, such artifacts are likely to appear. In order to prevent this, one could either compute a larger number of particle histories in the simulation or sample the particles from a different distribution that has more density mass in its outer regions or tails.

To underline the explanation for the appearance of the artifacts above, we recomputed the estimates using the (re-)weighting method based on a direct computation of the expected value, which can be calculated using the convolution $\mathcal{S}$ of the Gaussian error kernel with the nominal phase space parameter distribution (compare 14.2). Figure 14.3 shows that this alleviates the discrepancy from the references, causing artifacts to disappear and also reducing the overall amount of deviation displayed in the difference maps.

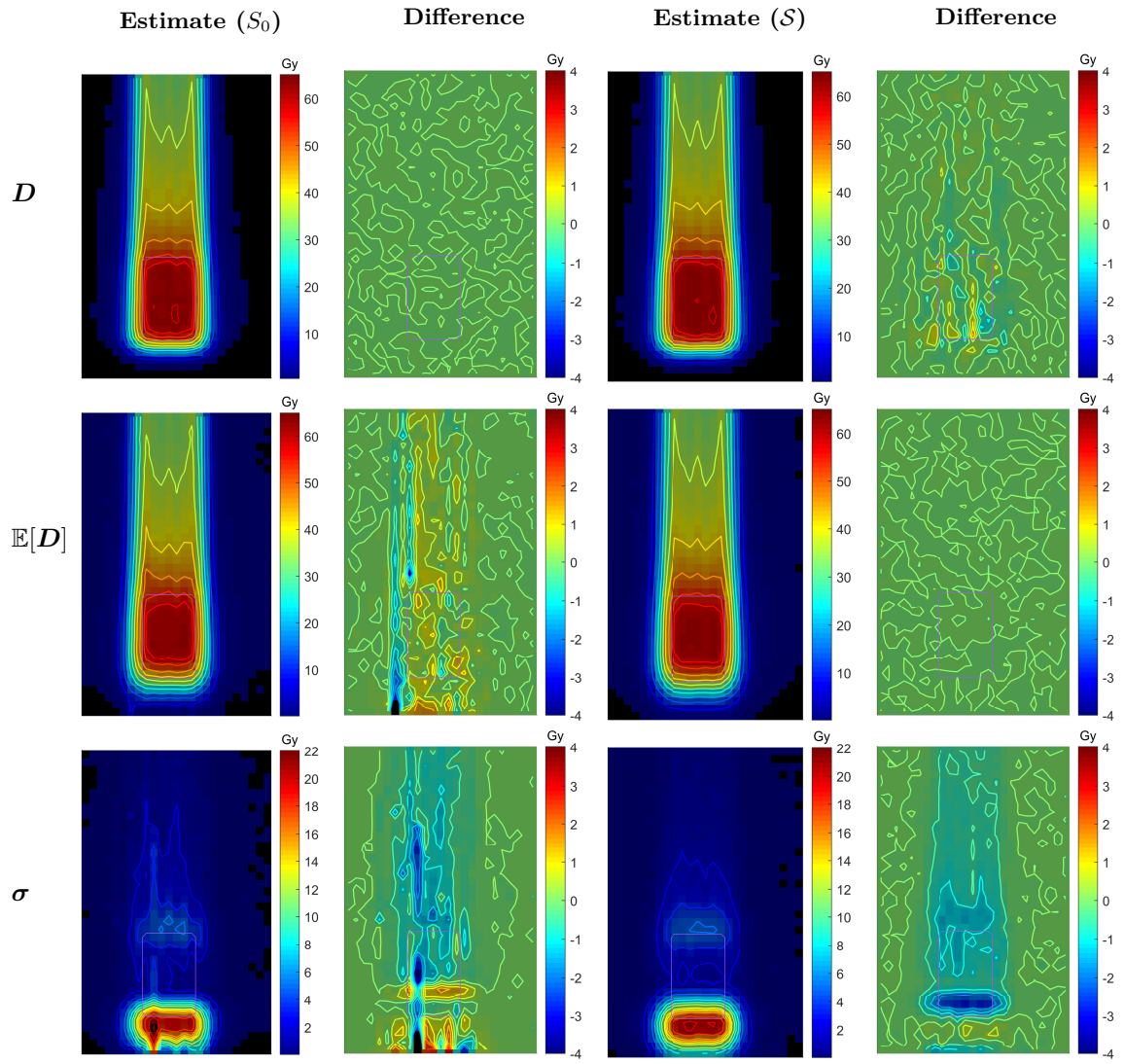


Figure 14.3: Nominal dose D, expected dose $\mathbb{E}[D]$ and standard deviation σ w.r.t. range uncertainties with a 3 % standard deviation for a spread out Bragg peak in a water phantom. The first and third column shows the estimate computed with the proposed (re-)weighting approach, reconstructed either from the nominal distribution S_0 or its convolution $\mathcal{S}$ with the error kernel. The second and fourth column shows the difference to the corresponding references.

Thereby, we can conclude that the irregularities in the solution can be attributed to the lack of statistical support in certain areas. Contrary to this, parts of the systematic differences remain and are thus most likely a result of the model approximations.

Figures 14.5 and 14.6 validate these observations for a liver patient. The difference maps for estimates computed based on the expected distribution $\mathcal{S}$, have less severe artifacts and systematic deviations. The $\gamma_{2\,\%}^{2\,\mathrm{mm}}$-pass rate also consistently increases for both the liver patient and water phantom (see table A1 (b), (c)).

Also, it has to be noted that using $\mathcal{S}$ to sample the initial particles leads to an expected dose estimate

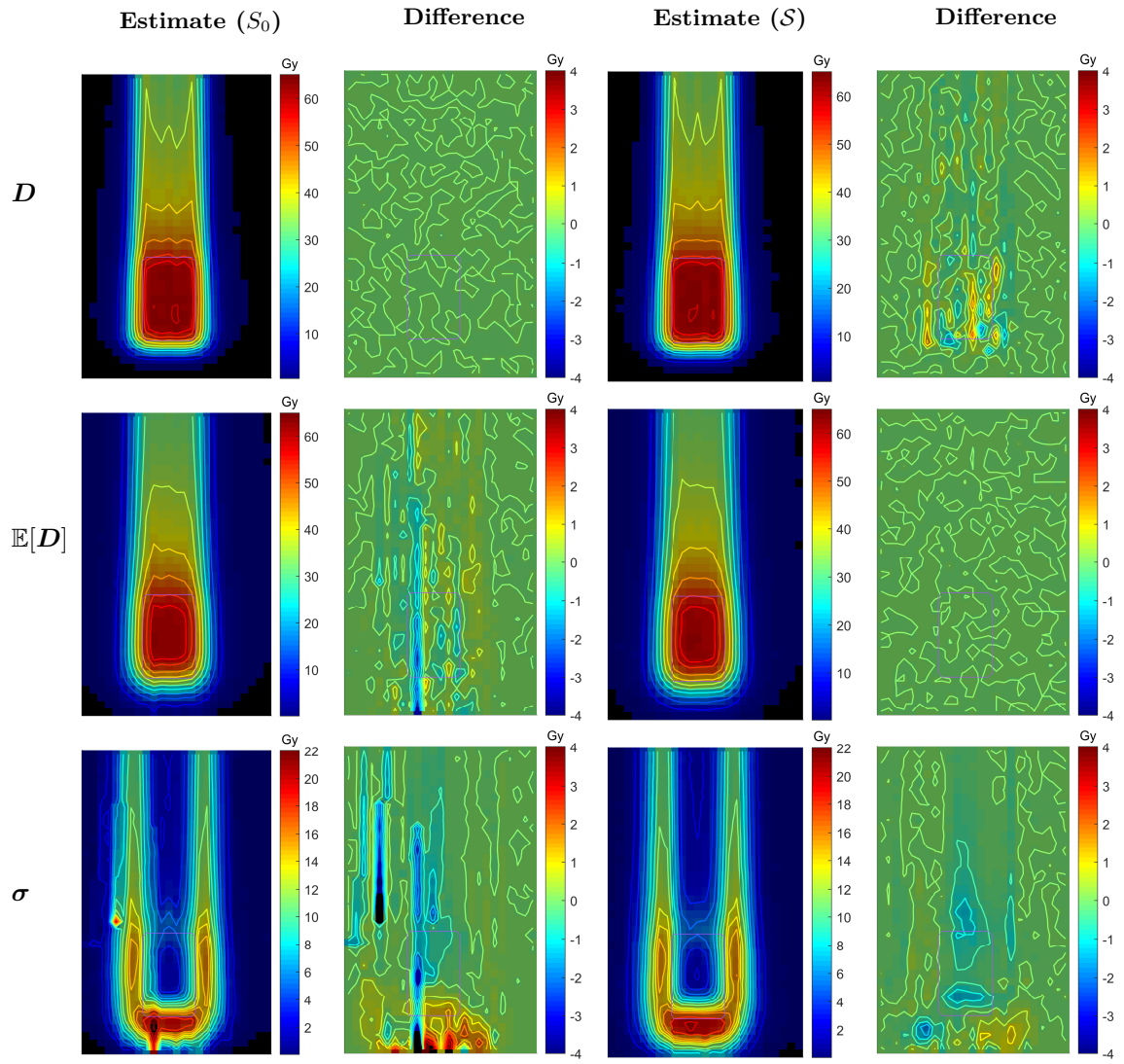


Figure 14.4: Nominal dose D, expected dose $\mathbb{E}[D]$ and standard deviation σ w.r.t. range uncertainties with 3 % and set-up errors with 3 mm standard deviation for a spread out Bragg peak in a water phantom. The first and third column shows the estimate computed with the reweighting approach, reconstructed either from the nominal distribution S_0 or its convolution $\mathcal{S}$ with the error kernel. The second and fourth column shows the difference to the corresponding references.

which is equivalent to the reference computations (compare eq. (14.6)), but a nominal dose estimate which now shows deviations from a nominal standard Monte Carlo reference computation in the order of magnitude that we could previously observe for the expected dose (see table A1 (b), (c)). This is due to the fact that the importance sampling error depends on the similarity of the sampling and target distribution. One can however expect a slightly lower error when constructing the nominal dose from a simulation of $\mathcal{S}$, since here the deviations of the target distribution occur in a region with a higher probability mass.

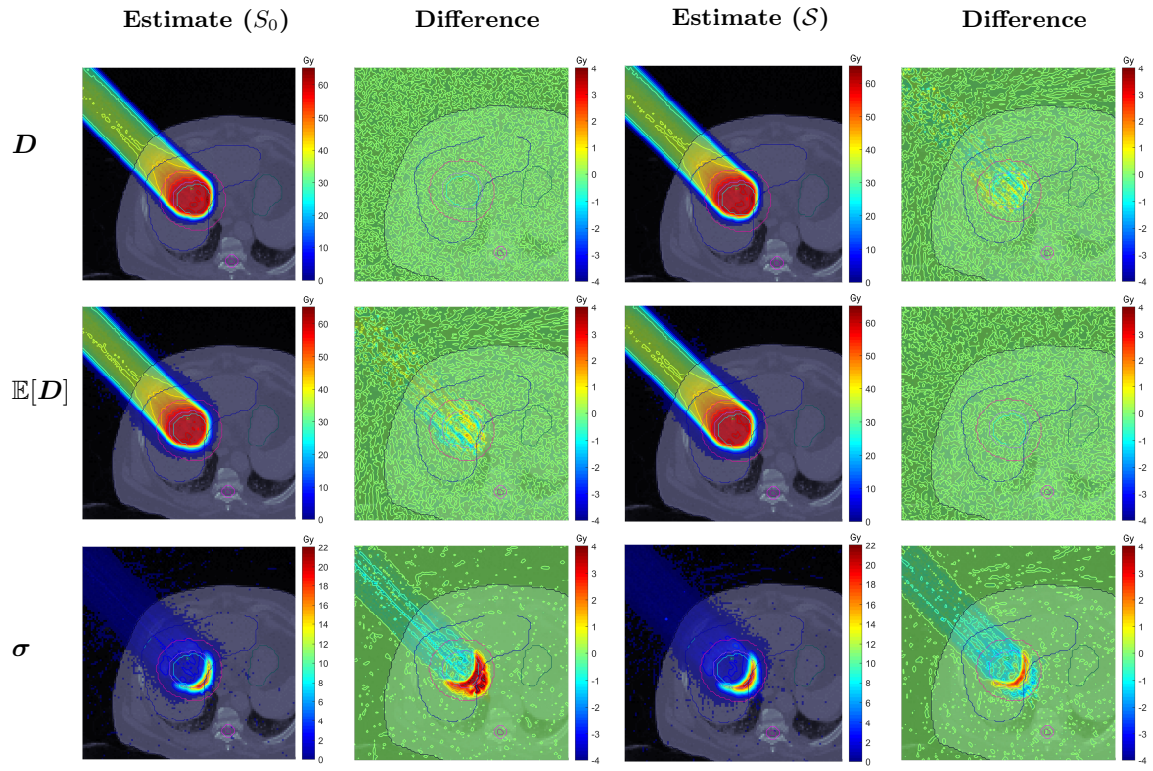


Figure 14.5: Nominal dose D, expected dose $\mathbb{E}[D]$ and standard deviation σ w.r.t. range uncertainties with a 3 % standard error in a liver patient (couch angle 0°, gantry angle 315°). The first and third column shows the estimate computed with the reweighting approach, reconstructed either from the nominal distribution S_0 or its convolution $\mathcal{S}$ with the error kernel. The second and fourth column shows the difference to the corresponding references.

14.9.3 Other correlation models

In chapter 11, we derived correlation matrices and error distributions for different uncertainty models. In figure 14.7 we first present the standard deviation estimate for four examples of simple error correlation models discussed in section 11.3. The results indicate that different correlation assumptions have a crucial impact on the standard deviation of dose distributions. While it is in principle possible to define arbitrary correlations within the proposed framework, estimates can be prone to artifacts due to a lack of statistical information. This is especially noticable for the ray-wise correlation model since a ray comprises the smallest number of beamlets. When sampling error realizations independently for smaller beam components, the reconstruction depends solely on the particle histories associated with these components. For rays with small weights, only very few histories are computed, therefore we observe similar artifacts as encountered in above range uncertainty computations (section 14.9.2).

We further investigate the more complex time-dependent uncertainty models presented in section 11.4.1. Our implementation was validated against Monte Carlo results for individual realizations of shifts over time. For a single ray with 6 energy levels in a water box, shifted by ±2 mm every 10 ms, the dose results in each time step reached 100 % agreement in a γ-analysis with tolerances of 2 mm/2 % in dose and location, respectively. Figure 14.8 compares the dose standard deviation for three such models in a water

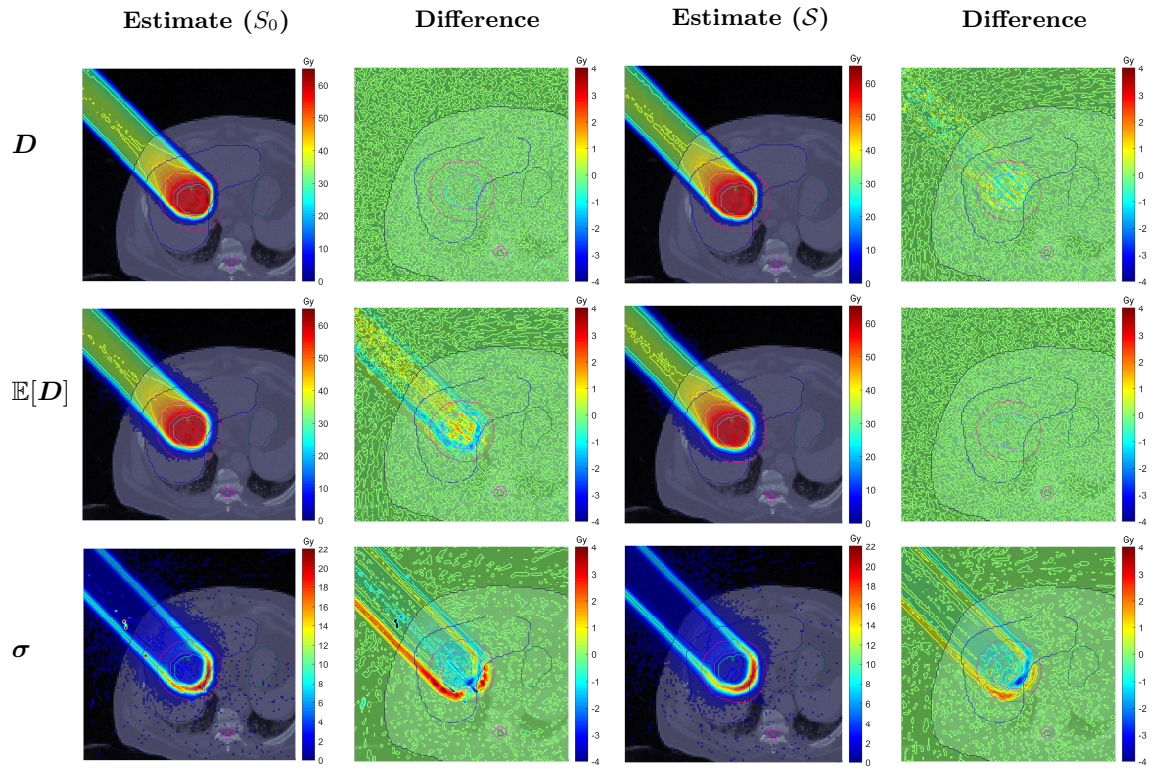
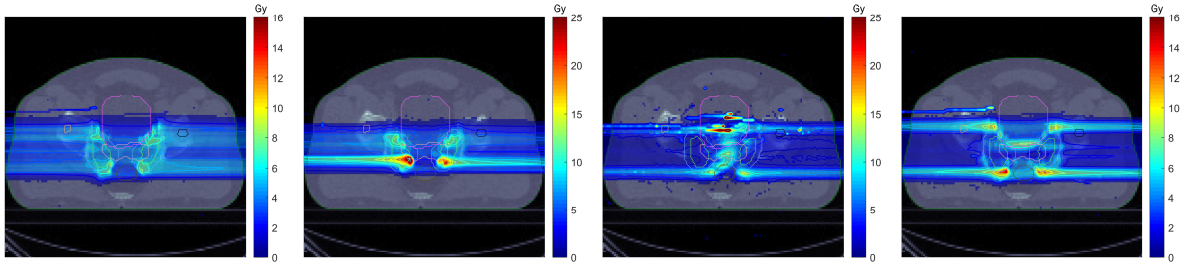


Figure 14.6: Nominal dose D, expected dose $\mathbb{E}[D]$ and standard deviation σ w.r.t. range uncertainties with 3 % and set-up errors with 3 mm standard deviation in a liver patient (couch angle 0°, gantry angle 315°). The first and third column shows the estimate computed with the reweighting approach, reconstructed either from the nominal distribution S_0 or its convolution $\mathcal{S}$ with the error kernel. The second and fourth column shows the difference to the corresponding references.

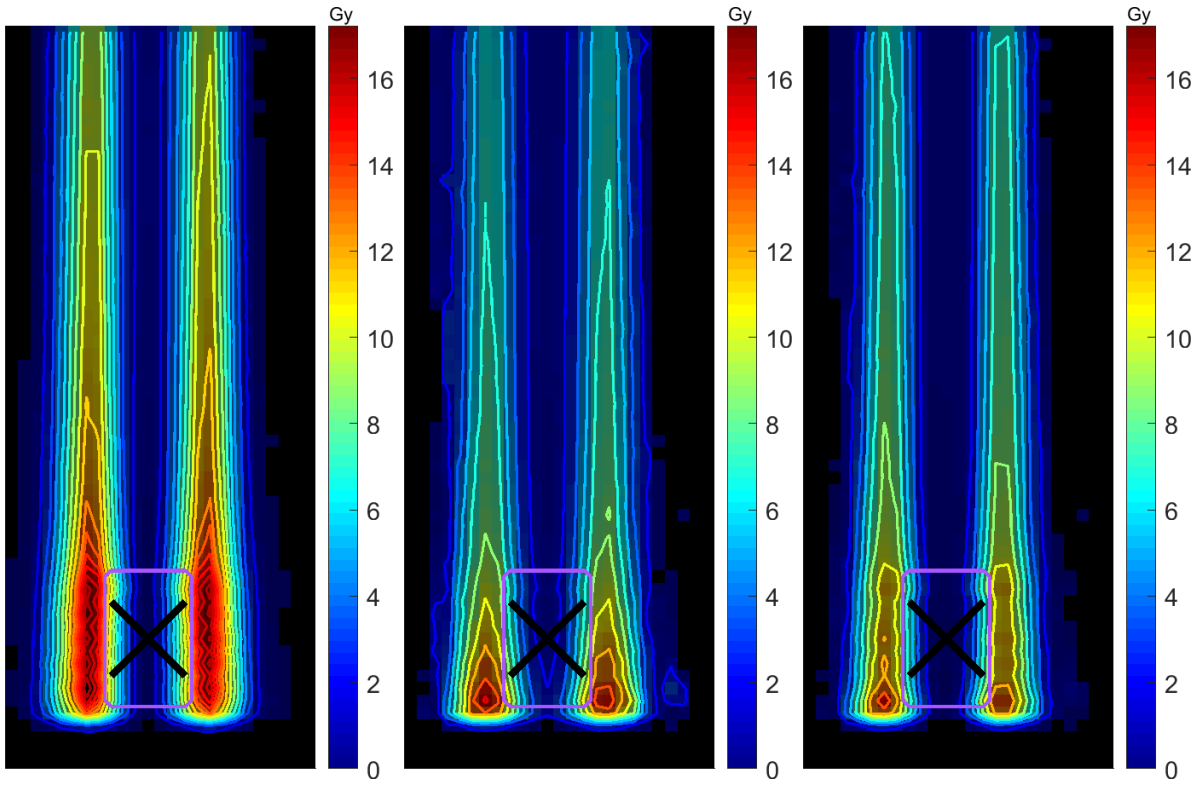


(a) (b) (c) (d)

Figure 14.7: Standard deviation of dose in a prostate patient for (a) no correlation (b) correlation between pencil beams in the same energy level (c) "ray-wise" correlation (i.e., between pencil beams with the same lateral position) and (d) correlation between pencil beams with the same irradiation angle (compare figure 11.1), w.r.t set-up errors and in case (c) also range errors.

phantom with a beam consisting of several superimposed pencil beams of seven different energy groups. The results illustrate that already in this simple case the uncertainty assumption has a significant impact on both the amount and shape of the dose standard deviation. However, due to the low amount of energy

levels and homogeneous tissue, the characteristics of the different input models are not fully visible. This



(a) (b) (c)

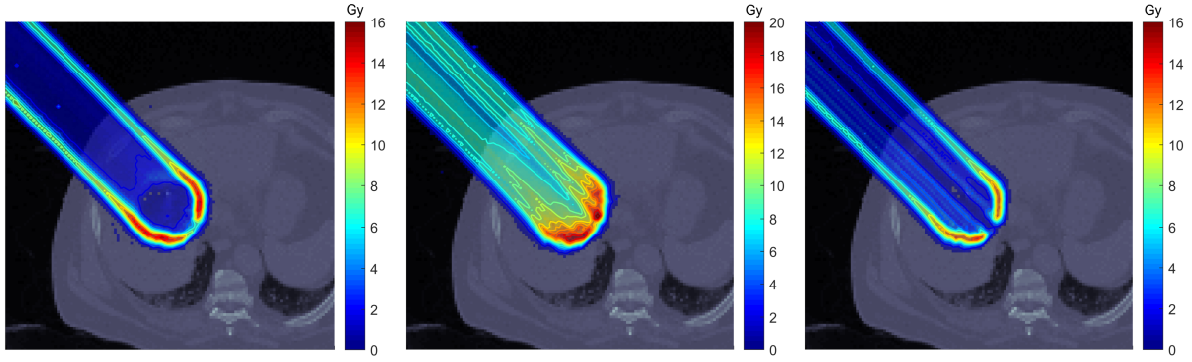
Figure 14.8: Standard deviation of dose from a beam made up of 175 pencil beams in a 3D waterbox for different error models: (a) one global error (b) AR(1) model for movement between energy levels (c) local periodic model for movement between energy levels

becomes apparent in fig. 14.9, which compares the standard deviation for the three error models in a realistic liver patient, with 21 energy levels and 166 rays. While the region of high standard deviation is still at the gradients at the edge of the beam in the fully correlated and periodic model, several local peaks can additionally be observed in the latter. In this case, considering the accumulated dose over all time steps/energy levels results in a lower magnitude of standard deviation, due to averaging of the errors. In the AR(1) model, which mimics a random movement process over time, the standard deviation is less concentrated to the beam edges. Here, pencil beams are frequently shifted against instead of with each other, due to the larger random component of the error in each time step/energy level.

14.9.4 Practical investigation of mathematical properties

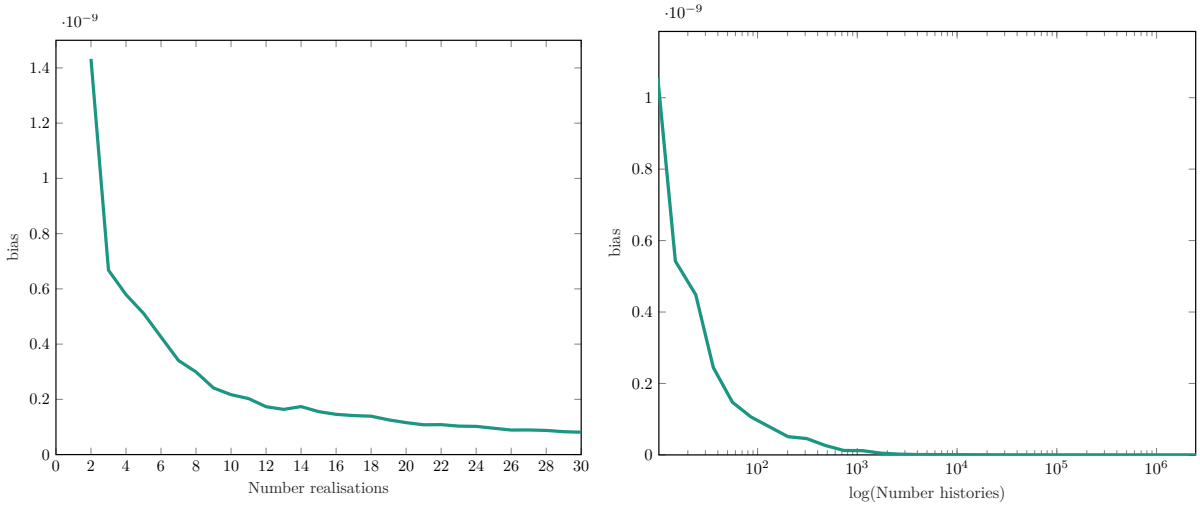
Bias

The above findings indicate that the bias in the dose variance estimate, introduced in proposition 14.2, does not significantly affect the accuracy of the results, as the agreement with the reference solutions is comparable for the unbiased expected value estimator and the biased variance estimator for set-up errors. Further, figure 14.10 shows that the bias reduces for an increasing number of particle histories.



(a) (b) (c)

Figure 14.9: Standard deviation of dose from a beam with irradiation angle (gantry angle 315°, couch angle 0°) made up of 1378 pencil beams in a liver patient for different error models: (a) one global error (b) AR(1) model for movement between energy levels (c) local periodic model for movement between energy levels

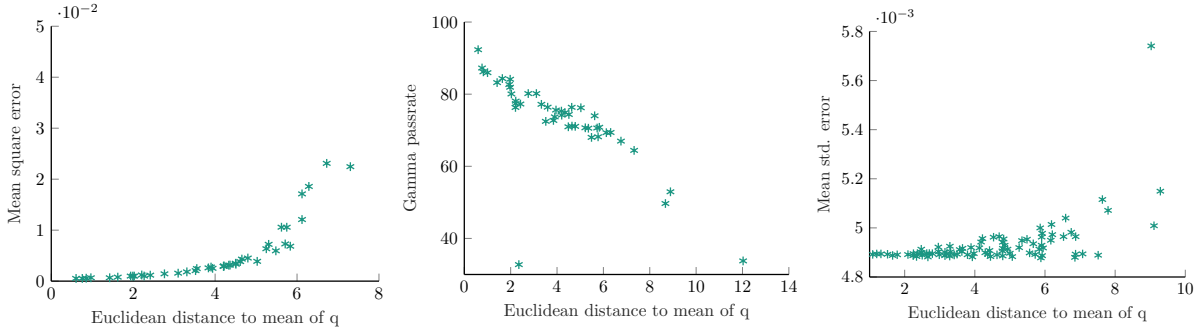


(a) (b)

Figure 14.10: Convergence of the variance estimator's bias in (a) number of error realizations used to compute the bias and (b) the number of particle histories used for the dose reconstructions for a single pencil beam in a water box.

Standard error

So far, the dose estimates were reconstructed mainly from samples of the distribution S_0. However, we have seen in eq. (14.14), that the standard error of the estimates depends on the relation of distribution from which the original sample was drawn to the target distribution. When computing the dose standard deviation, each error scenario corresponds to a Gaussian input distribution with a shifted mean value (see chapter 11).



(a) Mean square error (b) $\gamma_{2\,\mathrm{mm}/2\,\%}$-passrate (c) Mean theoretical error

Figure 14.11: Accuracy of dose estimates for increasing distance between importance distribution q and target distribution S_i. (a) Measured by mean square error (MSE), (b) by γ-passrate and (c) the mean of the theoretical error (see eq. (14.14)) in voxels with non-zero dose.

Figure 14.11 shows that the accuracy of the estimates decreases with growing distance of the mean of the target distributions from that of the sampling distribution. This development can be observed not only for the theoretical standard error but also for the mean square error and γ-passrates. This indicates that the more extreme realizations of the uncertain parameters introduce relatively high errors to the overall standard deviation estimate. Section 14.9.2 has already shown that it can be beneficial for the variance estimation to use a wider Gaussian distribution. We want to investigate more systematically, whether the use of a better-suited sampling distribution, such as a wider Gaussian or a Gaussian mixture of some of the target distributions, can improve the accuracy of the estimate. For this, we sample input parameters once from the joint distribution $\mathcal{S}$ and once from a mixture of the narrower Gaussian S_0, shifted by $\pm\sigma$ and $\pm 2\sigma$ in all coordinate directions. To better detect differences in the results, we use the stricter criteria $2\,\mathrm{mm}/2\,\%$ for the γ-analysis.

Figure 14.12 illustrates that the γ-passrate increases when using both the wider Gaussian and the Gaussian mixture. Thus, especially when one has an uncertainty model involving errors with a high variance or if the expected value and standard deviation of the dose are more important than the nominal dose, it can be worthwhile to not use the regular parameter distribution to generate the initial sample. For medical purposes, the high accuracy of the nominal dose is however an integral part of the quality assurance of a method and could often be prioritized over the quantification of uncertainties.

Variance upper bound

Lastly, in section 14.6.2 an upper bound for the dose variance estimate was derived, which is significantly less computationally expensive than the variance estimate itself. Figure 14.13 shows a comparison of this upper bound estimate to the dose variance. While the variance values are expectedly higher in the upper bound estimate, both exhibit a similar structure with two distinct variance peaks at positions with high dose gradients. This indicates a potential use for optimization purposes. However, further analyses in more complex test cases, in particular such with several irradiation angles and in heterogeneous materials, are necessary.

(a) S_0, $\gamma_{2\,\mathrm{mm}/2\,\%}$-passrate = 94.64 % **(b)** $\mathcal{S}$, $\gamma_{2\,\mathrm{mm}/2\,\%}$-passrate = 99.86 % **(c)** Gaussian mixture, $\gamma_{2\,\mathrm{mm}/2\,\%}$-passrate = 99.98 %

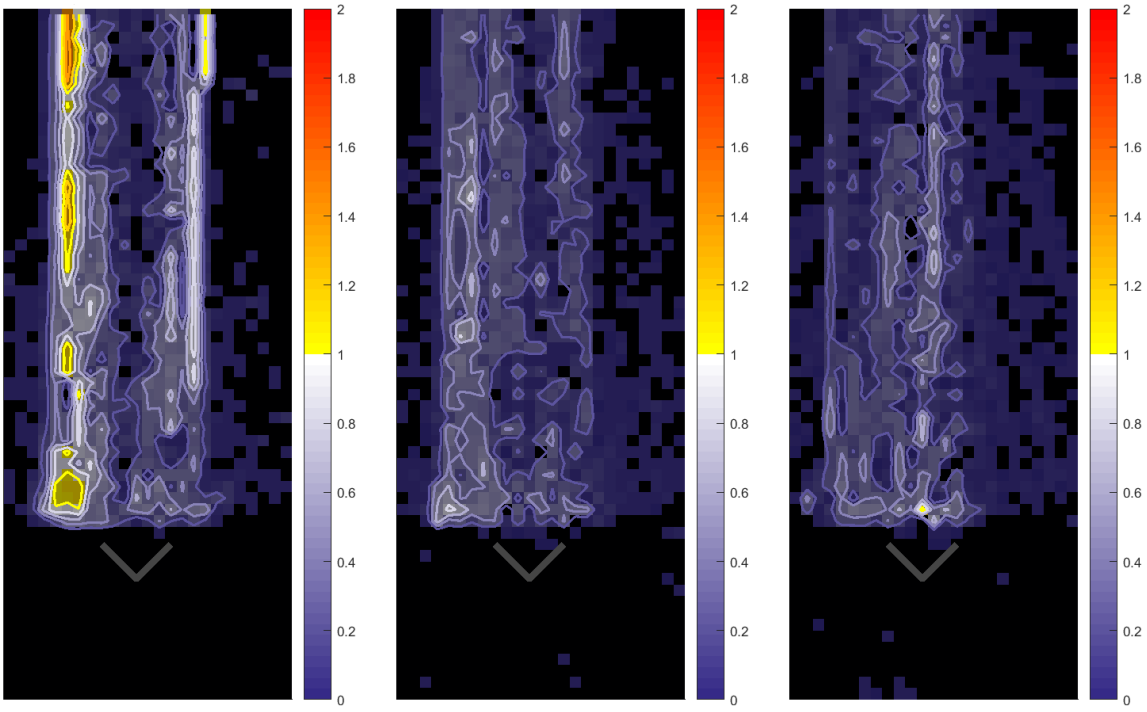


Figure 14.12: Estimates of the dose standard deviation in a water box using different sampling distributions: (a) the nominal distribution S_0, (b) the distribution $\mathcal{S}$ corresponding to the expected value and (c) a Gaussian mixture distribution

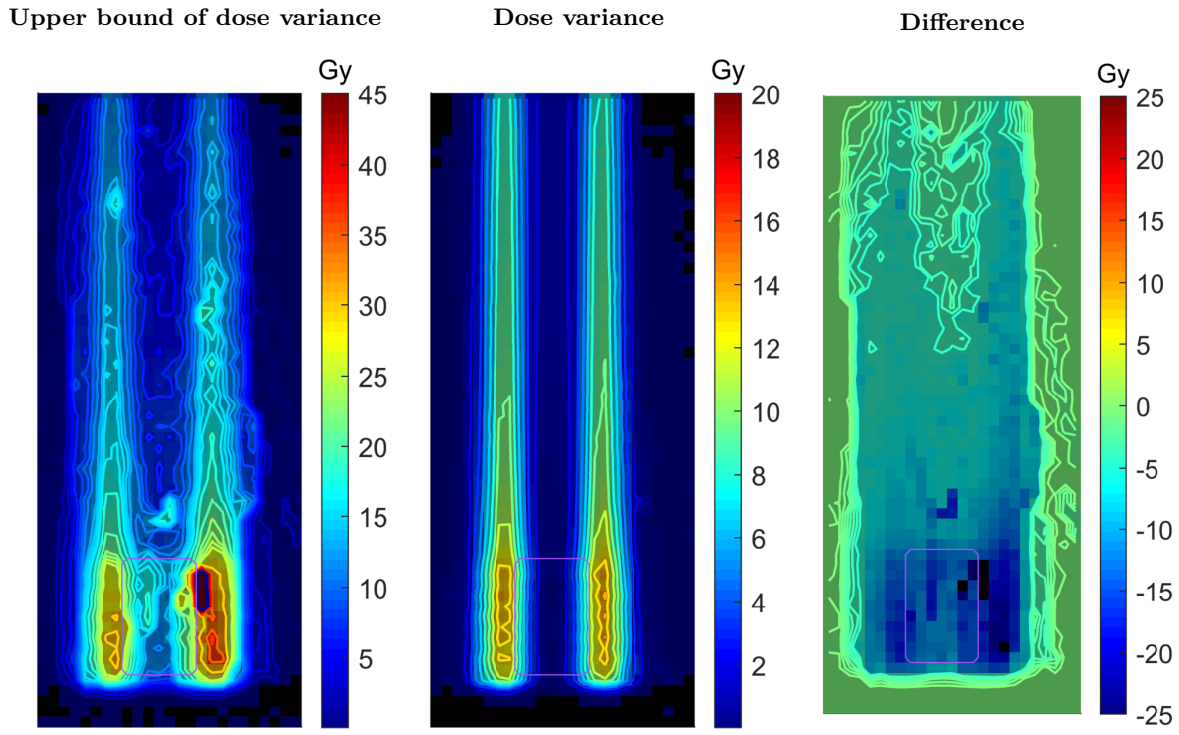
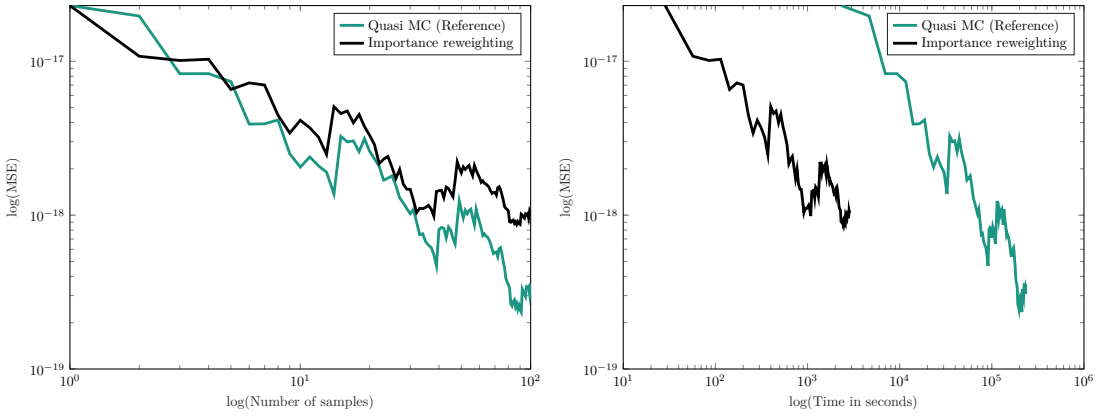


Figure 14.13: Comparison of the upper bound to the dose variance estimate for a beam made up of 175 pencil beams in a 3D water box.

14.9.5 Convergence

The proposed approach mimics a sampling-based uncertainty quantification method. The solution of the modified problem for each error realization using the black box solver is replaced by (re-)weighting the dose corresponding to particle trajectories from a previous simulation. Thus, the convergence of the method is equivalent to that of the underlying UQ method, which in our case is randomized quasi-Monte Carlo. Figure 14.14 illustrates this for the dose standard deviation of the single beam in a water box. While the convergence of reference and estimate is very similar per computed realization, the speed-up is achieved by the quicker computation of each iteration, that is, only performing MC scoring operations compared to physical simulations. Therefore, the amount of time that can be saved also depends on the specific patient case, treatment plan and software, which determine how much overhead is produced for example by the initialization and simulation of physical processes.



(a) (b)

Figure 14.14: Mean squared error of different methods compared (a) per iteration and (b) per time for dose computation in the water box with 175 pencil beams (see table 14.1).

For run-time comparisons, reference computations and the (re-)weighting approach were run on the same virtual machine[1] and timed during the computation of the dose standard deviation using the global error model. As table 14.2 shows, we observe reduced CPU times by a factor of approximately 183 and 80 for the two treatment plans using the water phantom and factors of 32 and 23 for the liver and prostate patient. Note, that the large difference in initialization times between the IMPT plans and the plan involving just one beam is due to the size of the sparse matrix containing the doses for all particle histories. For the larger plans, this exceeded the available RAM such that the use of less efficient tall arrays was necessary. Note, that these conclusions refer solely to a post-processing implementation of the method. When implemented as on-the-fly scoring, no storage of histories is required, instead, one sparse dose cube per considered scenario has to be stored. Also, the possibility to use existing efficient structures of the Monte Carlo code might enable performance improvements compared to the post-processing routine.

[1] Virtual machine including 64 CPUs with 1.995 GHz and 200GB RAM

Table 14.2: CPU time comparison for the reference vs. (re-)weighting approach applied to different test cases and computed on the same machine. All values are given in seconds. Note that the times for 100 realizations include the initialization times, while the time for a single realization only refers to the dose computation time.

		reference	(re-)weighting
Water phantom (1 beam)	Initialization	2.27	1.34
	One realization	72.21	0.38
	100 realizations	7221.64	39.34
Water phantom (IMPT)	Initialization	2.35	61.53
	One realization	2331.30	28.51
	100 realizations	233126.93	2912.53
Liver	Initialization	2.44	2038.75
	One realization	39066.44	1198.74
	100 realizations	3906650.90	121912.75
Prostate	Initialization	4.26	4867.75
	One realization	58762.40	2479.07
	100 realizations	5876253.86	252774.75

14.10 Extension to other non-intrusive methods

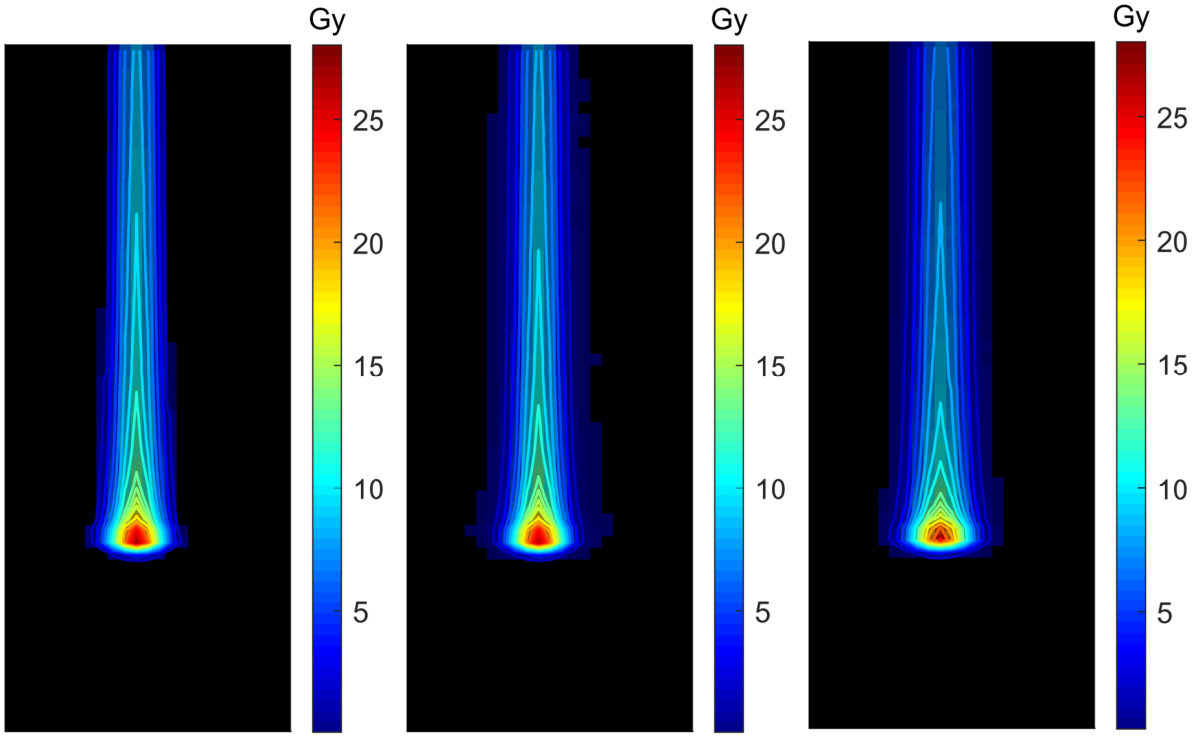
The proposed method can be used analogously to speed up uncertainty quantification using other non-intrusive methods, simply by replacing the required function evaluations by reweighting instead of several runs of the Monte Carlo dose calculation algorithm.

In the case of the standard Monte Carlo method, we merely have to change the choice of sample points for the uncertain parameter to random sampling instead of a number theoretic sequence. The results are more or less equivalent to the randomized quasi-MC method but require slightly higher sample sizes.

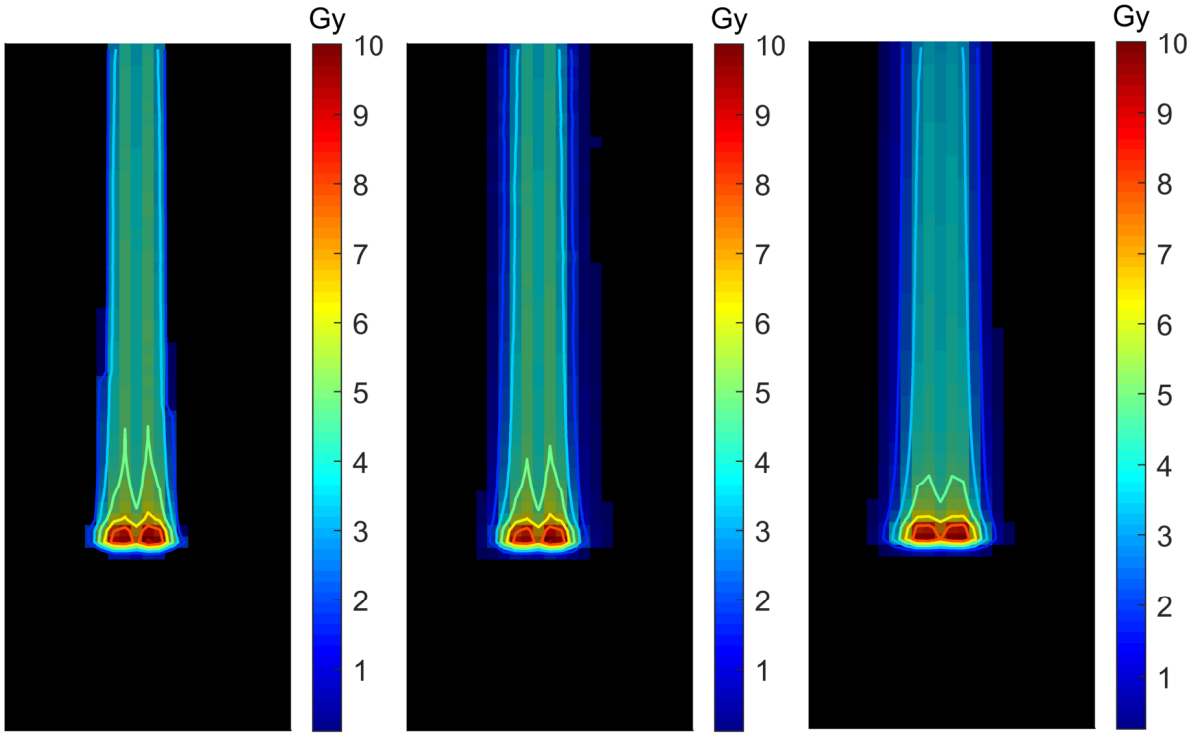
14.10.1 NISP

A more interesting application is for the response calculation in a non-intrusive spectral projection. Here, we follow the method described in [174] and replace the deterministic pencil beam method for dose calculations with MC importance reweighting. Our implementation is an extension of the open-source package OpenPC. We further implement a regression-based fitting of the coefficients besides the included spectral projection.

Figure 14.15 compares the expected dose and variance for both NISP approaches combined with importance reweighting and the reference from fig. 14.1. All methods are in good agreement visually. The regression-based NISP approach however takes on average around 10 times longer for the simple water box test case (209.36 CPU seconds) than the projection-based approach. For a polynomial model with 35 coefficients, the projection approach is approximately two times quicker (19.52 CPU seconds) than the quasi-Monte Carlo importance reweighting approach with 100 sample points (comp. table 14.2). The



(a) Expected value: Left to right: Estimate with regression-based NISP, estimate with projection-based NISP and a randomized quasi-MC reference.

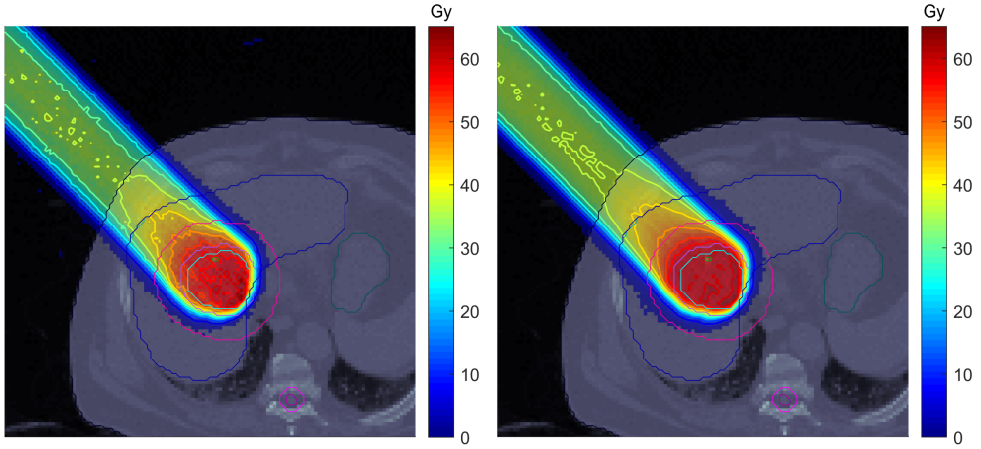


(b) Variance: Left to right: Estimate with regression-based NISP, estimate with projection-based NISP and a randomized quasi-MC reference.

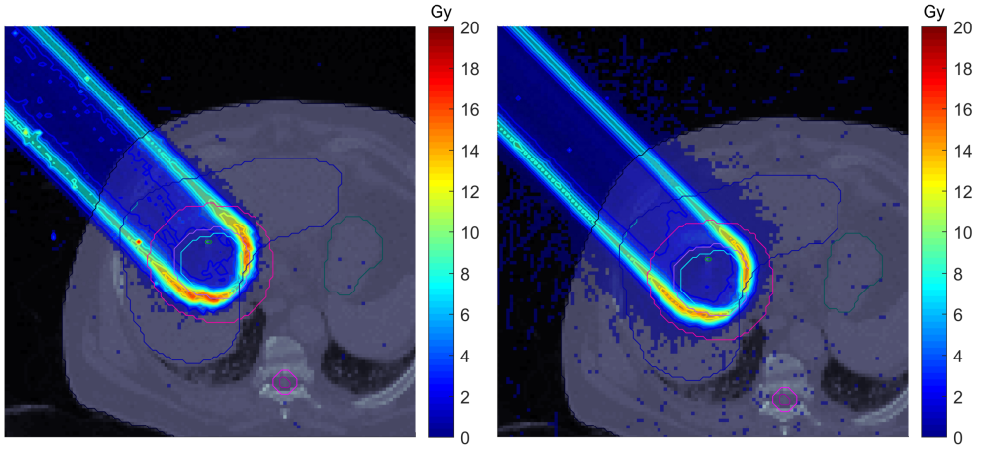
Figure 14.15: (a) Expected dose and (b) dose variance computed for a three-dimensional, Gaussian set-up error with 3mm standard deviation and a single pencil beam in a water box using a non-intrusive spectral projection with 35 coefficients/polynomials compared to a randomized quasi-MC reference where full dose computations are run for each sample point.

number of required coefficients and therefore dose evaluations however depends on the complexity of the test case and will likely be higher for a heterogeneous CT scan.

Figure 14.16 shows the dose expected value and variance for a three-dimensional, Gaussian set-up error with 3mm standard deviation in a liver patient. Even using sparse grids and a truncated polynomial basis, the method requires at least $50 - 100$ dose response calculations to achieve good variance results in these heterogeneous cases. We however not only receive a dose expected value and variance estimate but also a functional representation of the dose depending on the uncertain parameters, which could be used for sensitivity analyses or to generate dose samples cheaply on demand.



(a) Expected value: Left estimate with NISP and importance reweighting, right a randomized quasi-MC reference.



(b) Variance: Left estimate with NISP and importance reweighting, right a randomized quasi-MC reference.

Figure 14.16: (a) Expected dose and (b) dose variance computed for a three-dimensional, Gaussian set-up error with 3mm standard deviation in a liver patient using a non-intrusive spectral projection with 84 coefficients/polynomials compared to a randomized quasi-MC reference where full dose computations are run for each sample point.

CHAPTER 15

Dynamical low-rank approximation for uncertainty quantification

In the previous chapter, we have seen an approach that allows a more efficient consideration of uncertainties when using Monte Carlo dose calculations. Even in the deterministic case and with more efficient uncertainty propagation methods, MC simulations are however relatively computationally expensive and the storage requirements quickly exceed the limitations of regular personal computers as well as smaller workstations.

The dynamical low-rank approximation has been introduced in chapter 7 to increase the cost and storage efficiency of deterministic Boltzmann solvers. Several ways of making use of this framework in a UQ context are imaginable. First, due to the reduced computational cost, larger sample sizes and thus a higher accuracy can be afforded. Thus, non-intrusive propagation methods from section 12.2 that were deemed infeasible for a use with standard dose calculation methods can now be applied using DLRA. Further, the choice of rank could be utilized as the lever for different levels of accuracy and cost in a mutlilevel Monte Carlo approach [73, 66]. While these options are straightforward to implement, they draw their increased efficiency solely from the improved deterministic dose calculation rather than a specific uncertainty quantification approach. We therefore want to consider an intrusive approach, making use of the fact that the dynamical low-rank approximation for matrix (partial) differential equations can be extended to tensors [158, 103, 139]. Recently, a use of the additional dimensions for uncertain parameters has been proposed and applied for different simple problems in [110, 222]. In the following, we first recall the DLRA for tensors and subsequently demonstrate an application for uncertainty quantification in radiation therapy including initial numerical results compared to a collocation approach.

15.1 Tensor DLRA

We follow the description in [158] and extend the dynamical low-rank approximation for matrix differential equations $\dot{\mathbf{u}}(t) = \mathbf{F}(t, \mathbf{u}(t))$ to tensor differential equations $\dot{\mathcal{U}}(t) = \mathcal{F}(\mathcal{U}(t))$, where $\mathcal{U}(t) \in \mathbb{R}^{N_1 \times N_2 \times N_3}$. Note, that an extension to higher dimensional tensors is possible and straightforward. The tensor could for example denote the time (or energy) dependent solution of the transport equation for a discretization in space, angle as well as a one-dimensional uncertainty.

Let $\mathcal{M}_{\mathcal{R}} = \mathcal{M}_{\mathcal{R}}^{(N_1,N_2,N_3)}$ be the manifold of rank $\boldsymbol{\mathcal{R}} = (\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3)$ tensors. A low-rank representation is given by

$$\mathcal{U}_{hij}(t) \approx \sum_{k=1}^{\mathcal{R}_1} \sum_{l=1}^{\mathcal{R}_2} \sum_{m=1}^{\mathcal{R}_3} c_{klm} q_{hk} \otimes v_{il} \otimes w_{jm} \ \in \mathcal{M}_{\mathcal{R}}, \tag{15.1}$$

where $(q_{hk}) \in \mathbb{R}^{N_1 \times \mathcal{R}_1}, (v_{il}) \in \mathbb{R}^{N_2 \times \mathcal{R}_2}$ and $(w_{jm}) \in \mathbb{R}^{N_3 \times \mathcal{R}_3}$. Here, $C = (c_{klm}) \in \mathbb{R}^{\mathcal{R}_1 \times \mathcal{R}_2 \times \mathcal{R}_3}$ is an arbitrary tensor of full-rank also referred to as core tensor. Using the notation from [158, 46], the entries of the n-mode tensor product $\mathcal{A} \times_n B$ are given by

$$(\mathcal{A} \times_n B)_{i_1 i_2 \cdots i_{n-1} j_n i_{n+1} \cdots i_d} = \sum_{i_n} a_{i_1 i_2 \cdots i_{n-1} i_n i_{n+1} \cdots i_d} b_{j_n i_n},$$

for $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_d}, B \in \mathbb{R}^{J_n \times I_n}$. Then every tensor $\mathcal{U} \in \mathcal{M}_{\mathcal{R}}$ can be written as

$$\mathcal{U} = C \times_1 Q \times_2 V \times_3 W$$

and $Q = (q_1 | \cdots | q_{\mathcal{R}_1}) \in \mathbb{R}^{N_1 \times \mathcal{R}_1}, V = (v_1 | \cdots | v_{\mathcal{R}_2}) \in \mathbb{R}^{N_2 \times \mathcal{R}_2}, W = (w_1 | \cdots | w_{\mathcal{R}_3}) \in \mathbb{R}^{N_3 \times \mathcal{R}_3}$ have orthonormal columns. Further, the tangent tensor can be uniquely decomposed as

$$\partial\mathcal{U} = \partial C \times_1 Q \times_2 V \times_3 W + C \times_1 \partial Q \times_2 V \times_3 W + C \times_1 Q \times_2 \partial V \times_3 W + C \times_1 Q \times_2 V \times_3 \partial W,$$

where

$$Q^T \partial Q = 0,\ V^T \partial V = 0,\ W^T \partial W = 0, \quad \partial Q \in \mathbb{R}^{N_1 \times \mathcal{R}_1},\ \partial V \in \mathbb{R}^{N_2 \times \mathcal{R}_2},\ \partial W \in \mathbb{R}^{N_3 \times \mathcal{R}_3}$$

Analogous to chapter 7, time evolution equations for the factors can be defined by imposing

$$\dot{\mathcal{U}}(t) \in T_{\mathcal{U}(t)} \mathcal{M}_{\mathcal{R}} \qquad \text{such that} \qquad \left\| \dot{\mathcal{U}}(t) - \mathcal{F}(\mathcal{U}(t)) \right\| = \min .$$

We now want to apply the tensor DLRA to a CSD equation with uncertainties and derive equations for the unconventional integrator [36].

15.2 Application to radiation transport

We start with the original collided-uncollided split

$$\partial_t \psi_u = -\boldsymbol{\Omega} \cdot \nabla_{\mathbf{x}} \frac{\psi_u}{\rho} - \Sigma_t \psi_u := R_u(t, \psi_u) , \tag{15.2a}$$

$$\partial_t \psi_c = -\boldsymbol{\Omega} \cdot \nabla_{\mathbf{x}} \frac{\psi_c}{\rho} - \Sigma_t \psi_c + \mathcal{Q}(\psi_u + \psi_c) := F(t, \psi_u, \psi_c) . \tag{15.2b}$$

Note that for simplicity, we will only consider a one level collided-uncollided split and the equation in one spatial dimension, the derivations for higher dimensions and levels are however analogous.
Let us thus simplify eq. (15.2) to slab geometry, i.e., we have

$$\partial_t \psi_c = -\mu \partial_x \frac{\psi_c}{\rho} - \Sigma_t \psi_c + \mathcal{Q}\left(\psi_u + \psi_c\right) := F(t, \psi_u, \psi_c) \tag{15.3}$$

and extend the phase space to include the uncertain parameter Δ: $\psi_c(t, x, \mu, \Delta)$. A Tucker tensor approximation of this gives

$$\psi_c(t, x, \mu, \Delta) = C(t) \times_1 U^x(t, x) \times_2 U^\mu(t, \mu) \times_3 U^\Delta(t, \Delta). \tag{15.4}$$

Here, $U^x \in \mathbb{R}^{N_1 \times \mathcal{R}_1}$, $U^\mu \in \mathbb{R}^{N_2 \times \mathcal{R}_2}$, $U^\Delta \in \mathbb{R}^{N_3 \times \mathcal{R}_3}$ and $C \in \mathbb{R}^{\mathcal{R}_1 \times \mathcal{R}_2 \times \mathcal{R}_3}$ is also referred to as the core tensor. For the sake of brevity, we will often omit the phase space dependencies in the following. To define the tensor unconventional integrator, we now want to iteratively matricize the tensor equations and apply the matrix unconventional integrator (UI) [36] following the derivation in [36, 34]. More details on the derivation for general tensor differential equations and properties of the resulting integrator can also be found in [36, 34, 138, 228]. To be able to work with the more convenient matrix operations, we now introduce i-mode matricization of tensor $\mathcal{A} \in \mathbb{R}^{N_1 \times \cdots N_{i-1} \times N_i \times Ni+1 \times \cdots \times N_d}$ as

$$\mathbf{Mat}_i(\mathcal{A}) \in \mathbb{R}^{N_i \times N_1 \cdots N_{i-1} N_{i+1} \cdots N_d},$$

where the $i-th$ mode of the original tensor is fixed and all other indices are folded together. In the resulting matrix, the k-th row then contains all entries of $\mathcal{A}$ which have index k in the i-th mode, ordered colexicographically. The reverse operation of i-mode tensorization is defined by

$$\mathbf{Ten}_i(\mathbf{Mat}_i(\mathcal{A})) = \mathcal{A}.$$

Then according to [228] eq. (15.4) is equivalent to

$$\mathbf{Mat}_1(\psi_c) = U^x \mathbf{Mat}_1(C) \left(U^{\mu,T} \otimes U^{\Delta,T}\right) \tag{15.5}$$

$$\mathbf{Mat}_2(\psi_c) = U^\mu \mathbf{Mat}_2(C) \left(U^{x,T} \otimes U^{\Delta,T}\right) \tag{15.6}$$

$$\mathbf{Mat}_3(\psi_c) = U^\Delta \mathbf{Mat}_3(C) \left(U^{x,T} \otimes U^{\mu,T}\right). \tag{15.7}$$

Let's consider one time step from t_0 to $t_1 = t_0 + h$ with initial decomposition $\psi_{c,0} = C_0 \times_1 U_0^x \times_2 U_0^\mu \times_3 U_0^\Delta$. We start with the first, i.e., spatial dimension and factorize the matricized core tensor using a QR-decomposition:

$$\mathbf{Mat}_1(C_0)^T = R_0^x S_0^{x,T} \in \mathbb{R}^{\mathcal{R}_2 \mathcal{R}_3 \times \mathcal{R}_1},$$

where $R_0^x \in \mathbb{R}^{\mathcal{R}_2 \mathcal{R}_3 \times \mathcal{R}_1}$ and $S_0^x \in \mathbb{R}^{\mathcal{R}_1 \times \mathcal{R}_1}$. With

$$V_0^{x,T} := R_0^{x,T} \left(U_0^{\mu,T} \otimes U_0^{\Delta,T}\right) \in \mathbb{R}^{\mathcal{R}_1 \times N_2 N_3}$$

the solution can be rewritten as

$$\begin{aligned} \mathbf{Mat}_1(\psi_{c,0}) &= U_0^x \mathbf{Mat}_1(C_0) \left(U_0^{\mu,T} \otimes U_0^{\Delta,T}\right) \\ &= U_0^x S_0^x V_0^{x,T}. \end{aligned} \tag{15.8}$$

Note, that here V_0^x is constant in time. Equation (15.8) is now an SVD-like matrix representation to which the UI can be applied. With $K^x(t_0, x) := U_0^x S_0^x$, this yields:

$$\partial_t K^x(t,x) = \mathbf{Mat}_1 \left(F\left(t, \psi_u, \mathbf{Ten}_1(K^x(t) V_0^{x,T})\right)\right) V_0^x. \tag{15.9}$$

Integrating eq. (15.9) from t_0 to t_1, the updated factor U_1^x can be directly determined using a QR-decomposition

$$K^x(t_1) = U_1^x R_1^x \tag{15.10}$$

and we store $M^x = U_1^{x,T} U_0^x$. To update the core tensor, the updated factors corresponding to the angular and uncertain parameter dimensions are required. We can proceed analogously and now matricize according to the second, i.e., angular, mode and again factorize the core tensor into $\mathbf{Mat}_2(C_0)^T = R_0^\mu S_0^{\mu,T} \in \mathbb{R}^{\mathcal{R}_1 \mathcal{R}_3 \times \mathcal{R}_2}$. Moreover, we define $K^\mu(t_0, \mu) := U_0^\mu S_0^\mu$ and $V_0^{\mu,T}(t, \Delta) := R_0^{\mu,T} \left(U_0^{x,T} \otimes U_0^{\Delta,T} \right)$. Then the equations of the UI read

$$\partial_t K^\mu(t,\mu) = \mathbf{Mat}_2 \left(F\left(t, \psi_u, \mathbf{Ten}_2(K^\mu V_0^{\mu,T})\right)\right) V_0^\mu \tag{15.11}$$

$$K^\mu(t_1) = U_1^\mu R_1^\mu \tag{15.12}$$

$$M^\mu = U_1^{\mu,T} U_0^\mu. \tag{15.13}$$

Lastly, we consider the third mode, corresponding to the uncertain parameter Δ and factorize the core tensor into $\mathbf{Mat}_3(C_0)^T = R_0^\Delta S_0^{\Delta,T} \in \mathbb{R}^{\mathcal{R}_1 \mathcal{R}_2 \times \mathcal{R}_3}$ with $K^\Delta(t_0, \Delta) := U_0^\Delta S_0^\Delta$ and $V_0^{\Delta,T}(t, \Delta) := R_0^{\Delta,T} \left(U_0^{x,T} \otimes U_0^{\mu,T} \right)$. We again integrate

$$\partial_t K^\Delta(t,\Delta) = \mathbf{Mat}_3 \left(F\left(t, \psi_u, \mathbf{Ten}_3(K^\Delta V_0^{\Delta,T})\right)\right) V_0^\Delta \tag{15.14}$$

from t_0 to t_1 and set

$$K^\mu(t_1) = U_1^\mu R_1^\mu \tag{15.15}$$

$$M^\mu = U_1^{\mu,T} U_0^\mu. \tag{15.16}$$

Note, that these updates of factors U^x, U^μ, U^Δ can all be performed in parallel. Finally, the core tensor can be updated using the updated factors and stored quantities by integrating the following tensor differential equations from t_0 to t_1

$$\begin{aligned} C(t_0) &= C_0 \times_1 M^x \times_2 M^\mu \times_3 M^\Delta \\ \dot{C(t)} &= F\left(t, \psi_u, C(t) \times_1 U_1^x \times_2 U_1^\mu \times_3 U_1^\Delta\right) \times_1 U_1^{x,T} \times_2 U_1^{\mu,T} \times_3 U_1^{\Delta,T} \end{aligned} \tag{15.17}$$

and setting $C_1 = C(t_1)$.

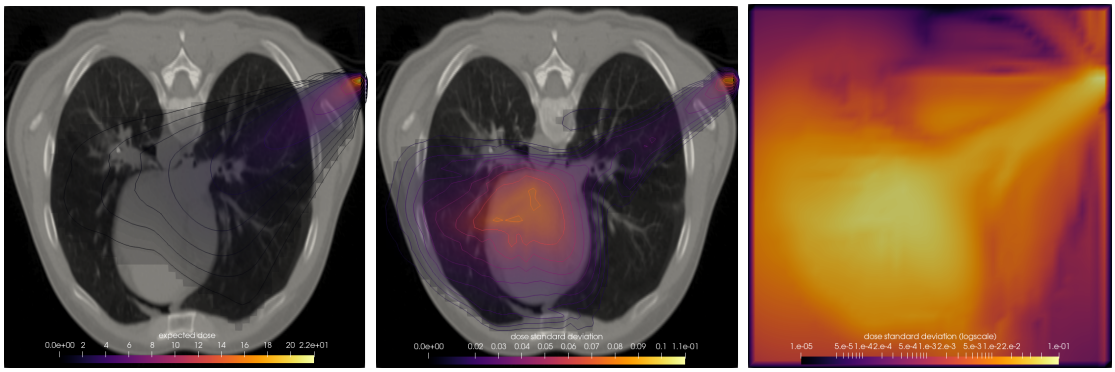
The explicit form of the equations can now be obtained by plugging in the right-hand side function F from eq. (15.3). Moreover, we assume in the following that $\rho(x, \Delta) = \frac{c\rho_0(x)}{c + \rho_1(x)\Delta}$, i.e. the uncertainty manifests in the tissue density.

15.3 Results

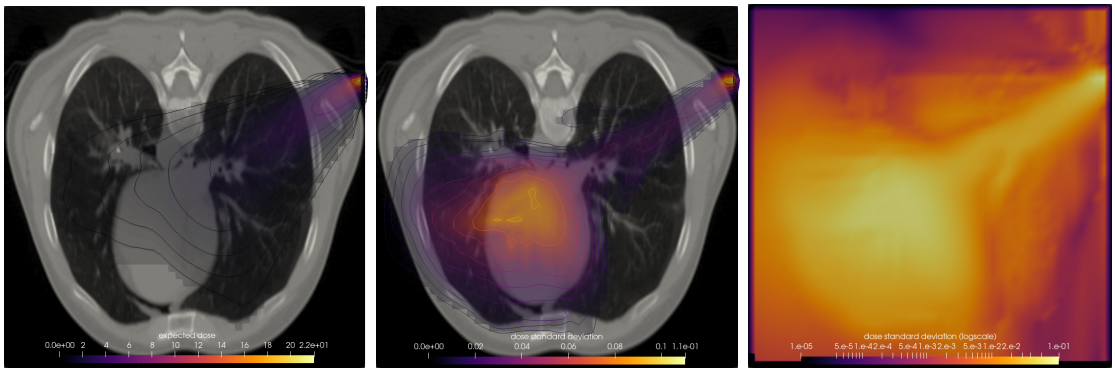
Quantifying the effect of arbitrary time-dependent variations of the density is typically difficult to achieve in standard dose calculation algorithms combined with non-intrusive uncertainty propagation, due to the

high dimensionality of the parameter space when sampling the density at each spatial position and time-point explicitly. Using the time-dependent CT data from [88], we want to demonstrate that tensor DLRA is suitable to capture realistic uncertainties due to respiratory motion, including anatomical deformations.

As a proof-of-concept, we interpret a time series of 10 CT images from the given data set as samples from a stochastic process. The CT images are interpolated to obtain a continuous process and facilitate sampling at arbitrary time points. Figure 15.1 shows a comparison between a computationally expensive non-intrusive collocation approach with 100 nodes, which runs a full dynamical low-rank dose calculation for each sample point and the tensor DLRA approach, which incorporates the samples directly as a discretization of the uncertain parameter dimension. For both approaches, we use a rank of 50 and



(a) Collocation, 100 nodes



(b) DLRA, rank 50

Figure 15.1: Dose expected value (left), standard deviation (middle) and log-scale plot of the standard deviation (right) for a reference collocation approach (a) compared to tensor DLRA for uncertainty quantification (b).

moments up to degree 17 for the spherical harmonics angular discretization. The CT scans are scaled to a 14.5cm×14.5cm spatial domain, which is divided into a 50×50 cell angular grid. One can see that the variance due to density changes is highest at the beam entry, where dose values are high as well as near the end of the beam, where the largest deformations occur in the time-dependent CT scan. At first glance, all results shown in fig. 15.1 agree well between the tensor DLRA and the collocation approach. In the log plots of the standard deviation, it is however visible, that the DLRA estimate is slightly less smooth

and exhibits some artifacts in the low-dose regions. However, the standard deviation and expected value show the same structure and agree well in the regions where higher doses are delivered. Further studies with different types of uncertainties and higher temporal resolution are necessary to validate these results and investigate whether small artifacts are a common effect of the DLRA or merely a result of the low sample size of the original data set (only 10 points in the time dimension). In the initial implementation, the tensor approach is approximately 2 times faster and allocates overall 3.1 times less memory than the collocation method[1]. This performance can likely be improved through optimized implementation and possibly by using a different tensor format. Since classical tensor formats such as the Tucker tensor used in this work can exhibit structural weaknesses in higher dimensions [82], the use of different, more sophisticated tensors formats such as the hierarchical Tucker tensor or Tensor train could be beneficial [139]. Further, [139] show that these formats can be used to compute updates within an optimization algorithm. Future work could also investigate an adaptation of this approach for (robust) optimization in treatment planning.

[1]Performance measured on the same workstation using TimerOutputs.jl. Note, that the memory allocations refer to the total allocations made during the computation, thus in the collocation approach, the allocations per sample are much lower. Depending on whether the computations are done in parallel for several samples or sequentially, either approach could have an advantage memory-wise.

CHAPTER 16

Discussion

Chapters 10 to 15 discussed the uncertain transport problem, starting with mathematical models of uncertainties over common uncertainty propagation methods to novel and efficient approaches tailored to different dose calculation methods.

Uncertainty modeling

Initially, we recall relevant uncertainties in radiation therapy and the simplified mathematical models typically used in radiation therapy planning. We then proceed to introduce several novel modeling approaches which model uncertainties in terms of the multivariate Gaussian phase space parameter distribution. Range errors can be approximated using the distribution of the initial particle energy. Time-dependent movements are translated into beamlet correlations according to the beam application pattern. We further demonstrate how to adjust these time-dependent models to realistic patient data and validate our approximations. The range approximation preserves the correct solution structure but introduces a small systematic error. For our model, which assumes all materials as water-equivalent with different densities, these errors are not amplified by heterogeneous vs. homogeneous materials. Future work could investigate whether the approximation error increases for more detailed material calibration models. The time series modeling produces accurate results when using the beam application times as time steps and considering whole-body motion or homogeneous materials.

As an extension to this work, the time-dependent error modeling introduced in chapter 11 might be extended to simulate spatial motion for 4D treatment planning or even deformations in geometry components [10]. The derived correlation models are of interest in their own right and can be transferred to any dose calculation and uncertainty quantification framework which supports multivariate error correlations [e.g. 10]. Further, the data-based models could be used outside the realm of uncertainty quantification

to predict patient or tumor motion in inline radiation therapy.

Uncertainty propagation

Next, we discuss different established and novel approaches for uncertainty propagation using the derived uncertainty models. Due to the high computational cost of dose calculations, uncertainty propagation can be infeasible with standard approaches. We introduce two methods aimed at the efficient computation of uncertainty statistics for Monte Carlo and deterministic dose calculations, respectively.

Importance (re-)weighting

In chapter 14 we demonstrate how importance sampling can be adapted to lower the computational costs of sampling-based uncertainty quantification in Monte Carlo dose calculations using the example of proton irradiation. We derive an unbiased estimator and the standard error for the expected value, as well as an estimator, bias and upper bound for the standard deviation. We apply the proposed method to Gaussian set-up uncertainties with 3 mm standard deviation in a homogeneous water phantom, a prostate and a liver patient. For a simple global uncertainty model, we observe high agreements of at least 99.9 % for both test cases (figs. 14.1 and 14.2). We show that these results can be further improved by adapting the distribution function used for the initial Monte Carlo simulation to the quantity which is of the highest interest. Here, the user can choose between the exactness of the nominal dose estimate and that of the expected dose or standard deviation.

The sampling of error realizations is completely independent of the solution of the transport problem, which allows us to incorporate complex correlation models including movement patterns and time dependence. This is a clear advantage compared to classic sampling-based methods, which would require an elaborate setup and a large number of realizations to account for the degrees of freedom in a multivariate model. Therefore, more complex correlation models are currently often neglected in uncertainty studies [10, 178, 130], although it is apparent in figs. 14.8 and 14.9, that different uncertainty models can have a significant impact on the dose standard deviation, already in a homogeneous material.

While an extension to other types of uncertainties in the phase space parameters is straightforward, the method is limited to parameters that have a non-degenerate density function. Uncertainties in other parameters of the simulation might require more complex, non-Gaussian error models which are able to reflect physical constraints as well as space- and time dependencies. Here, the derivation of the required conditional and joint probability distributions may not be straightforward. Future work could explore the adaptation and application of the model to uncertainties beyond Gaussian errors in the patient setup and particle range.

The main computational advantage of the proposed method is that it minimizes the overhead associated with complex MC simulations. Therefore, we cannot claim a faster convergence rate and the speed-up is dependent on the specific application and implementation. A combination with more sophisticated scenario-based UQ methods, such as quadratures or a non-intrusive polynomial chaos expansion, could be explored further, also with view to the use of the derived functional model for sensitivity analyses and probabilistic optimization.

Tensor dynamical low-rank approximation

Since the efficiency of the importance (re-)weighting method is still limited by the costly initial MC simulation, we then consider a second uncertainty propagation method based on the more time and memory-efficient dynamical low-rank approximation for deterministic dose calculations. Here, tensor DLRA is used to include additional dimensions for uncertain parameters in the transport problem. Due to the explicit time/energy evolution of the solution also in the uncertain dimensions, this approach allows direct modeling of temporal changes even in previously deterministic quantities, such as the tissue density. A proof-of-concept example with an electron beam in a lung CT and uncertainties according to 4D-CT density data shows that the results are comparable to those determined with a standard collocation approach. In future work, the method could be extended to higher-dimensional time-dependent uncertainties and different tensor formats could be investigated, also with view to a use within robust optimization algorithms. With view to a realistic clinical application, an extension to proton transport as well as further validations against Monte Carlo-based results and a more in-depth analysis of the convergence and run-time improvements would be of interest.

Preview

So far, we have discussed different numerical methods for dose calculations and established realistic uncertainty models as well as methods to efficiently propagate them through the dose calculations. In the next and last part of this work, we tackle the questions of how dose calculations as well as uncertainty estimates can be included in treatment plan optimization in order to determine more robust plans.

Part IV

Robust planning

Predictions of the delivered dose and the impact of uncertainties are used in robust treatment planning to adjust the beam set-up and achieve the optimal treatment outcome. As a proof-of-concept, we consider two common approaches for the design of a robust optimization problem, based on worst-case scenarios or the expected value, respectively. Further, the speed-up of optimization through precomputation of dose and variance influence matrices is discussed. We derive an algorithm to compute such influence terms efficiently for Monte Carlo dose calculations using importance (re-)weighting.

CHAPTER 17

Modeling robustness

In part III of this thesis, we have covered how uncertainties in radiation therapy can be modeled mathematically and propagated through the dose calculation algorithm to receive estimates of uncertainty statistics such as the expected value and variance. Ideally, this knowledge should be incorporated into the treatment planning process to reduce negative side effects and ensure tumor coverage and organ sparing. In the following chapters, we discuss how to quantify the robustness of a treatment plan and include this in the treatment plan optimization. Further, we demonstrate the use of the importance (re-)weighting method for uncertainty propagation (chapter 14) within a robust optimization framework.

17.1 Margin recipes

Treatment planning is based on an initial delineation of volumes of interest such as the tumor or organs at risk. A simple but common way of handling uncertainties is to define a planning margin around these structures according to the expected uncertainty model. Here, the majority of approaches rely on linearly expanding the margin size based on the aggregated standard deviation [218]. While recent works explore margin recipes aimed at incorporating more complex models of patient behavior and adapting margins in time according to time-dependent motion [e.g. 101, 26, 92], [26] show that a more expensive optimization-based consideration can still significantly reduce the dose delivered to organs at risk while ensuring within 1% equivalent dose in the tumor.

17.2 Robust and probabilistic optimization

Especially when considering intensity-modulated radiation therapy, it seems natural to incorporate knowledge about uncertainties into the optimization process which determines the treatment plan, i.e., beams

at which position and with which energy are irradiated with which intensity (duration). There are many approaches to how this can be done, including simple approximations which assume invariance of the dose [43] as well as more sophisticated approaches with probabilistic objective functions or constraints [199, 37, 7]. Here we will consider two examples from the generalized class of minimax stochastic programming [59]. First, we however recall the deterministic dose and optimization problem in IMRT from part I:

$$D(\mathbf{w}, \mathbf{r}) = \sum_{b=1}^{B} w_b \cdot D_b(\mathbf{r}), \tag{17.1}$$

where $D_b(\mathbf{r})$ is the dose delivered to spatial position $\mathbf{r}$ by the beamlet $b = 1, ..., B$ with weight/intensity w_b. The weights are determined through optimization

$$w^* = \arg\min_{w \geq 0} F(D(\mathbf{w}, \mathbf{r})) \tag{17.2}$$

and the objective function can be chosen to reflect different quality metrics, such as the tumor coverage or deviation from the prescribed dose. Different common choices for the objective function are described e.g. in [234]. Here, we will use a squared deviation function, i.e.,

$$F(D(\mathbf{w}, \mathbf{r})) = P(\mathbf{r}) \cdot \left(D(\mathbf{w}, \mathbf{r}) - D^*(\mathbf{r})\right)^2, \tag{17.3}$$

where $D^*(\mathbf{r})$ is the prescribed dose and $P(\mathbf{r})$ is a location-specific penalty, which can be used to especially enforce compliance with the prescription in regions of importance like the tumor or vital organs.

Note that in practice, the dose computation is often shifted out of the optimization routine by precomputing so-called *dose influence matrices*. For a voxelized geometry with $V = n_x \cdot n_y \cdot n_z$ cells and an IMRT dose as in eq. (17.1), the dose influence matrix $DI \in \mathbb{R}^{N \times B}$ at entry $DI_{vb}, v = 1, .., V, \ , b = 1, ..., B$ contains the dose delivered by beamlet b to voxel v. Then for every iteration of the optimization algorithm the dose in each voxel $DI_{v,*}(\mathbf{w}) = \sum_{b=1}^{B} w_b \cdot DI_{vb}$ can be aggregated using just the updated weights. While these matrices can easily have more than $10^9 - 10^{10}$ elements for common patient CTs and particle treatment plans, they are extremely sparse ($\sim$1% of nonzeros).

17.2.1 Worst-case optimization

Worst-case or minimax optimization refers to a group of conservative robust optimization approaches which incorporate uncertainties only in terms of a discrete set of *worst case scenarios*, without considering their associated probabilities. There are several variations of worst-case optimization, described and compared for example in [60]. Here, we will focus on *composite worst-case optimization*, where the maximum over a weighted sum of the scenario objective functions is minimized:

$$\min_{w \geq 0} \quad \max_{s \in \mathcal{S}} \sum_{k=1}^{K} v_k F_k(D(\mathbf{w}, \mathbf{r}, s)). \tag{17.4}$$

Here $\mathcal{S}$ is the set of scenarios and $F_1, ..., F_K$ are different objective functions, e.g. representing different goals in the optimization.

When considering set-up and range uncertainties typically one positive and negative offset per parameter dimension is considered, i.e. $\pm\delta_{r_x}^{wc} = \delta_{r_y}^{wc} = \delta_{r_z}^{wc} \overset{e.g.}{=} 3\text{mm}$ for each spatial dimension of the set-up error and $\pm\delta_R^{wc} \overset{e.g.}{=} 3\%$ for the range error [174].

The advantage of this approach is that no knowledge of a probability distribution is required and only a small number of scenario doses needs to be computed. Further, [59] have shown that a worst-case approach can lead to sharper dose fall-off outside of the treated volume compared to an expected value approach. However, potential knowledge of the uncertainty model or probability distributions is ignored and a conservative choice of the "worst case" can lead to possibly unjustified pessimistic predictions of plan quality [33]. We therefore also consider a second option, which makes use of the uncertainty models and propagation techniques developed previously in chapter 11 and chapters 12 to 15.

17.2.2 Expected value optimization

As the name suggests, in expected value optimization we consider the expected value of the objective function with respect to the probability distribution of uncertain parameters. Thus, the following optimization problem has to be solved

$$w^* = \arg\min_{w \geq 0} \mathbb{E}_{p_\Delta} \left[F(D(\mathbf{w}, \mathbf{r}, \Delta)) \right]. \tag{17.5}$$

For a squared deviation objective function and dose influence matrix approach, it can be shown that this expected value is given by [227]:

$$\mathbb{E}_{p_\Delta} \left[F(DI(\mathbf{w}, \Delta)) \right] = \sum_{v=1}^{V} P_v \sum_{b,m=1}^{N} \mathcal{V}_{vbm} w_b w_m + F(\mathbb{E}_{p_\Delta} \left[DI(\mathbf{w}, \Delta) \right]), \tag{17.6}$$

where P_v is the penalty in voxel v. Thus, similarly to the dose influence matrix, we can precompute the variance influence tensor $\mathcal{V} \in \mathbb{R}_+^{N \times B \times B}$, which in each entry $\mathcal{V}_{vbm}$ contains the variance between the doses delivered by beamlets b and m to voxel i. Since penalties are not dependent on the beamlet weights and vice versa the weights do not depend on the voxel position, we can further exchange the summation such that the first term becomes

$$\sum_{v=1}^{V} P_v \sum_{b,m=1}^{N} \mathcal{V}_{vbm} w_b w_m = \sum_{b,m=1}^{N} w_b w_m \sum_{v=1}^{V} P_v \mathcal{V}_{vbm} := \sum_{b,m=1}^{N} w_b w_m \Omega_{bm}, \tag{17.7}$$

$$\text{with } \Omega_{bm} = \sum_{v=1}^{V} P_v \mathcal{V}_{vbm} \tag{17.8}$$

and precompute and store the smaller matrix $\Omega \in \mathbb{R}_+^{B \times B}$.

CHAPTER 18

Robust optimization with importance (re-)weighting

Robust optimization entails the computation of several dose distributions in each iteration of the optimization routine. While precomputing the dose and variance influence matrices saves time during the optimization, it is still necessary to determine and store the dose influence for B individual beamlets and several scenarios. Especially when using Monte Carlo simulations for dose calculation, this can be very computationally expensive. As beamlet weights are not known in advance and the effects of each beamlet need to be separated, it is also not possible to reduce the costs by using fewer particles for beamlets with low intensities or reusing the particles simulated for neighboring beams.

For this reason, we want to incorporate the more efficient importance reweighting approach from chapter 14 to reconstruct the individual beamlet influences for different scenarios from one joint Monte Carlo simulation. This not only reduces the number of necessary simulations but could also allow for a more efficient initial simulation using cheaply precomputed weights or the synergies of neighboring beamlets by eliminating the need for separate computations.

18.1 Scenario and expected value computation

The dose distribution for worst-case scenarios as required for the worst-case optimization approach can be determined analogously to that of individual realizations in eq. (14.9). Similarly, the dose expected value and variance can be determined using eqs. (14.12) and (14.13). However, even using a more efficient approach, computing the variance in each iteration of the optimization is infeasible in a typical treatment planning process.

Therefore, we would like to precompute influence matrices for worst-case scenarios, expected dose and variance as efficiently as possible.

18.2 Dose and variance influence matrix

The computation of worst-case and expected dose influence matrices using importance reweighting is straightforward. Assuming we have simulated particles from the initial phase space distribution $q(\boldsymbol{Z})$, the target distributions in each beamlet $b = 1, ..., B$ are simply the individual Gaussian beamlet distributions $S_0^b = \mathcal{N}(\boldsymbol{\mu}_Z^b, \boldsymbol{\Lambda}^b)$ shifted by the worst-case offsets for each scenario $\pm\delta_{r_x}^{wc}, \pm\delta_{r_y}^{wc}, \pm\delta_{r_z}^{wc}, \pm\delta_R^{wc}$ or convolved with the (Gaussian) distribution of the uncertain parameter $p_\Delta = \mathcal{N}(\boldsymbol{\mu}_\Delta, \boldsymbol{\Sigma}_\Delta)$. Then the formulas in chapter 14 apply analogously and the B dose results for N voxels are collected in the matrix $DI \in \mathbb{R}^{N\times B}$.

For the variance influence, we need to compute

$$\Omega_{bm} = \sum_{v=1}^{V} P_v \cdot \mathcal{V}_{vbm} = \sum_{v=1}^{V} P_v \cdot Cov[DI_{vb}, DI_{vm}] = \mathbb{E}_{p_\Delta}[(DI_{vb} - \mathbb{E}_{p_\Delta}[DI_{vb}])(DI_{vm} - \mathbb{E}_{p_\Delta}[DI_{vm}])].$$

Using a (quasi-)MC approach with N points to solve the expected values numerically and applying the importance reweighting method, the summation order can be changed to yield the following algorithm, where $BB_I(\boldsymbol{z}_p)$ is the vector of voxelized black box responses for sampled phase space parameters $\boldsymbol{z}_p$.

Algorithm 18.1 Calculate variance influence matrix Ω

for particles p=1:H **do**
  Sample $\boldsymbol{z}_p \leftarrow q(\boldsymbol{Z})$
  Compute $BB_I(\boldsymbol{z}_p)$
end for
for i=1:N **do**
  Sample $\boldsymbol{\delta}_i \leftarrow p_\Delta(\boldsymbol{\Delta})$
  for beamlet distributions $S_i^b(\boldsymbol{Z})$,
  $b = 1, ..., B$ **do**
    $\mathcal{C}_{b,*} = \frac{1}{H}\sum_{p=1}^{H} BB_I(\boldsymbol{z}_p) \cdot \frac{\left(S_i^b(\boldsymbol{z}_p) - \mathcal{S}(\boldsymbol{z}_p)\right)}{q(\boldsymbol{z}_p)}$
  end for
  Ω += $\frac{1}{N}\sum_{v=1}^{V} P_v \cdot \mathcal{C}_{*,v} \cdot \mathcal{C}_{*,v}^T$
end for

Note, that the matrix Ω can also be aggregated per volume of interest by only summing over voxels within this volume in the last line of algorithm 18.1:

$$\Omega_{VOI} \mathrel{+}= \frac{1}{N} P_{VOI} \cdot \sum_{v\in VOI} \mathcal{C}_{*,v} \cdot \mathcal{C}_{*,v}^T.$$

Since each volume of interest typically shares the same penalty value for all voxels, the penalty multiplication can then also be pulled out of the summation. This strategy further has the advantage of allowing an individual analysis of the variance in each component.

It is further important for the convergence of the optimization algorithm, to ensure that the produced variance influence is always positive semidefinite. This can be easily shown by construction of the algorithm:

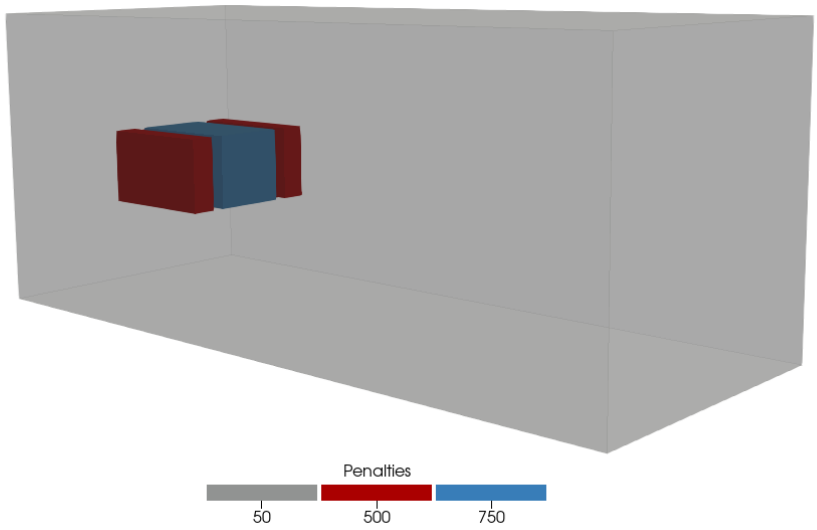


Figure 18.1: Phantom geometry with associated penalty values: Gray is the body, blue the tumor/target and red the organ at risk.

Proof. Since for any real matrix A and nonzero real column vector u:

$$u^T A A^T u = (A^T u)^T (A^T u) = |A^T u|_2^2 \geq 0$$

and $\mathcal{C}_{*,v} \cdot \mathcal{C}_{*,v}^T$ is the product of a real matrix by its transpose, this term is positive semi-definite. Further, the sum of positive semi-definite matrices multiplied by a positive factor is again positive semi-definite, therefore $\sum_{v=1}^{V} P_v \cdot \mathcal{C}_{*,v} \cdot \mathcal{C}_{*,v}^T$ is positive semi-definite and lastly for the same reason

$$\Omega \mathrel{+}= \frac{1}{N} \sum_{v=1}^{V} P_v \cdot \mathcal{C}_{*,v} \cdot \mathcal{C}_{*,v}^T$$

is always positive semi-definite.

Note, that Ω is also symmetric by construction. □

18.3 Results

We now consider a simple water phantom including a target surrounded by an organ at risk to illustrate the effects of the two previously discussed robust optimization strategies. The geometry and penalties used for the different volumes of interest are shown in fig. 18.1. The intensities or beamlet weights are optimized using the radiation therapy planning software matRad [234]. The matRad optimizer is based on IPOPT, which implements a primal-dual interior point method [224].

The target dose for the tumor is set to 60 Gy, while the upper bound for the organ at risk is 10 Gy and 20 Gy for the rest of the phantom body.

Note, that in the following, results from a robust optimization based on the pencil beam algorithm will be compared with such based on Monte Carlo dose calculations. Since both dose calculation methods themselves yield different results already in a deterministic setting, slight differences between the scenarios, variance influences or optimized dose are not necessarily due to the uncertainty propagation methods. Therefore, the focus here lies more on the structure and changes in the solution than the exact values.

18.3.1 Worst-case optimization

In the first optimization model, we consider 7 worst-case scenarios: ±3mm shift of the position in each spatial direction as well as a nominal scenario without shifts. The objective functions for the composite worst-case model are chosen to be a squared deviation function for the target volume and squared overdosing for the body and organ at risk. We compare a completely analytical computation using the pencil beam algorithm implemented in matRad with scenarios reconstructed from a nominal Monte Carlo dose calculation using importance reweighting. Through the use of the reweighting approach, we do not require the accurate simulation of individual beamlets to construct a dose influence matrix, as the dose for each beamlet can be derived from a joint simulation with arbitrary initial phase space distribution.

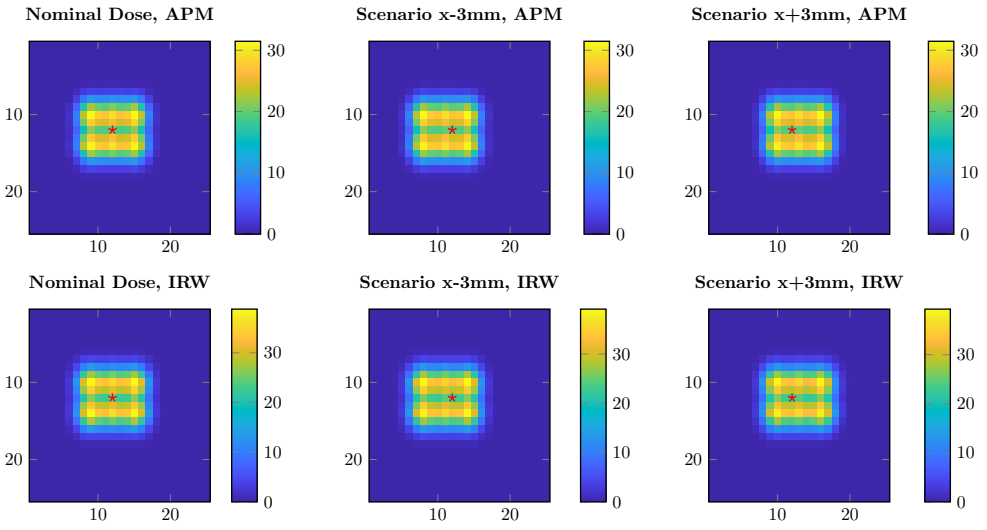


Figure 18.2: Comparison of nominal dose at surface depth and two error scenarios computed with a pencil beam method and APM vs. Monte Carlo dose calculation with importance reweighting (IRW). The red star marks the beam center in the nominal scenario without shift.

Figure 18.2 shows the surface level dose for the nominal scenario as well as shifts along one lateral dimension. It is apparent that both methods yield the same effects and beams are shifted analogously in each scenario. It is however notable, that the dose in the Monte Carlo-based computations is consistently

higher than that using the pencil beam method. Note that for this reason also slightly different ranges are used for the colormaps of the two methods in fig. 18.2. This serves to emphasize the effect of the scenario shifts rather than the differences due to the dose calculation algorithms.

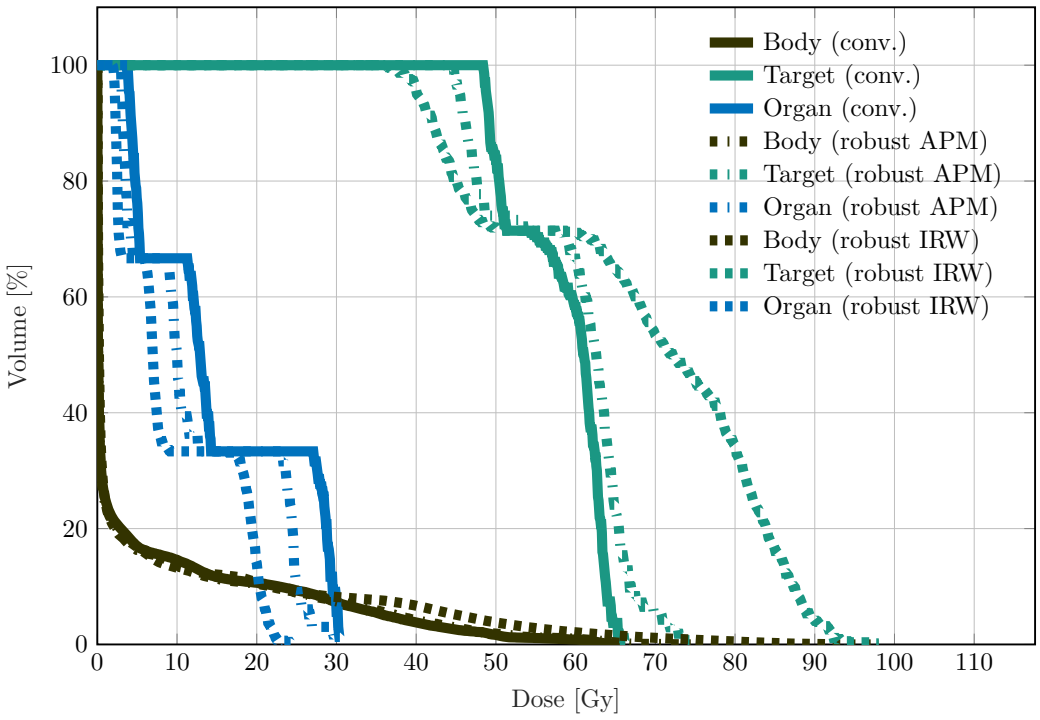


Figure 18.3: Comparison of conventional with robust DVHs computed with worst case optimization based on a pencil beam method and APM vs. Monte Carlo dose calculation with importance reweighting (IRW), respectively.

Figure 18.3 shows the dose volume histograms for a conventional (non-robust) optimization as well as the robust optimizations based on APM and Monte Carlo reweighting, respectively. Both methods exhibit the effect of reducing the dose delivered to the organ at risk and increasing the maximum dose delivered to the target. These effects are more pronounced in the Monte Carlo reweighting results, which further also slightly increase the dose delivered to the total body. Note, that the optimal outcome of robust optimization is to decrease the amount of dose delivered to organs at risk while maintaining at least the same level of dose in the tumor. The prioritizing of organs at risk is vs. target is then mainly steered by the chosen penalty values. Keeping this in mind, both methods have a positive effect on the expected treatment outcome and the MC reweighting approach performs better than APM. However, uniformity is also a criterion for the dose distribution in the target volume, which is not fulfilled well especially by MC reweighting. Since the organ at risk is placed to the sides of the target, a decrease of dose in the OAR can only be achieved by reducing the dose at the interface between the target and OAR. In order to maintain the total amount of dose to the target, the optimization then increases the dose in the center, leading to less uniformity overall. Since scenario shifts are portrayed accurately in both methods (see

fig. 18.2), the difference between worst-case optimization with APM and importance reweighting is likely mainly due to the difference between pencil beam and MC dose calculations rather than the scenario computation itself.

18.3.2 Expected value optimization

Second, we consider expected value optimization based on a squared deviation objective as described in section 17.2.2 and chapter 18. Figure 18.5 compares the variance influence matrices computed with APM and Monte Carlo importance reweighting for each volume of interest. Beamlets in the matrix are grouped per ray (all beamlets with the same spatial position) and ordered by their lateral position from low to high and left to right. Thus beamlets in the middle of the matrix are positioned in the center of the domain and beamlets with lower/higher indices are increasingly more left and lower/right and higher than the center (see fig. 18.4).

While for all volumes of interest, block structures representing groups of beamlets within the same ray are visible and beamlets with closer spatial proximity tend to show higher correlations, some differences in structure can be observed especially between the target and organ at risk. Due to the positioning of the organ to the sides of the central tumor, beamlets in the middle of the domain cause higher variance in the dose delivered to the target, while beamlets at the lower and upper end of the index spectrum have a larger influence on the dose variance in the organ at risk.

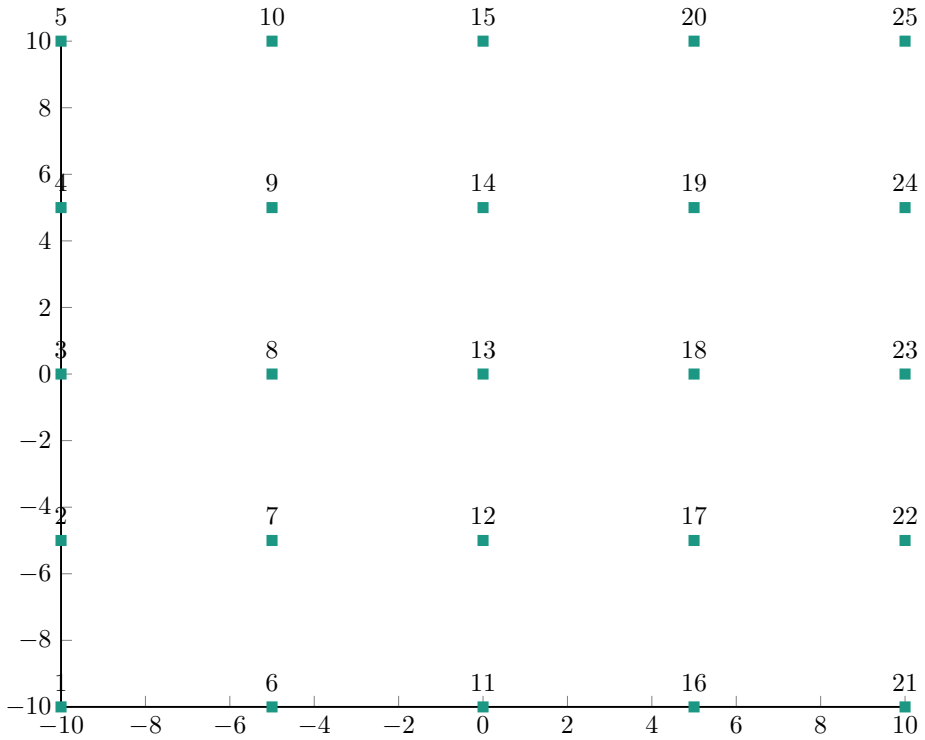
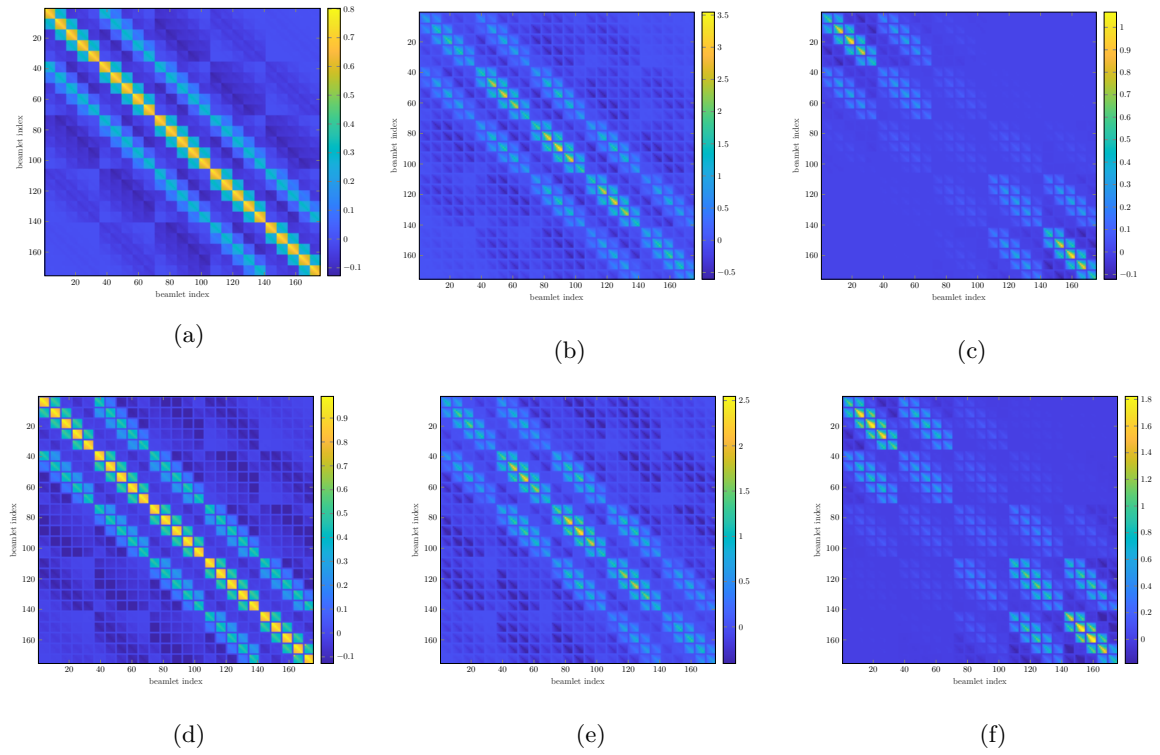


Figure 18.4: Spatial position and index in variance influence matrix of each of the 25 rays. Each ray contains 7 beamlets with the same spatial position, but different energies.



(a) (b) (c)

(d) (e) (f)

Figure 18.5: Variance influence matrix for a plan with 175 beamlets in a water phantom (fig. 18.1) and a global set-up error with 3mm standard deviation. (a) Influence matrix for the complete body computed with APM, (b) influence matrix for the target computed with APM, (c) influence matrix for the organ at risk computed with APM and (c) influence matrix for the complete body computed with MC importance reweighting, (d) influence matrix for the target computed with MC importance reweighting as well as (e) influence matrix for the organ at risk computed with MC importance reweighting.

The described structures are captured similarly in the variance influence matrices computed with both different methods. The total magnitude of variance values is however slightly higher in the target and lower in the organ at risk for the APM method compared to Monte Carlo reweighting. This also affects the results of the robust optimization presented in fig. 18.6.

Here, again both methods can reduce the dose delivered to the organ at risk while increasing the maximum dose in the target. Further, the uniformity of the dose distribution in the target is reduced in both cases. However, APM shows a larger difference to the conventional optimization in the target, while Monte Carlo importance reweighting has a larger effect in the organ at risk. This is likely due to the differences in their variance influence matrices: APM expects and tries to avoid a larger variance in the target, while the organ at risk is not expected to be affected by uncertainties as much and vice versa for MC reweighting. Which is a more desired outcome depends largely on the specific patient case and the marginal benefits of a little less dose in the specific OAR compared to more dose in the target.

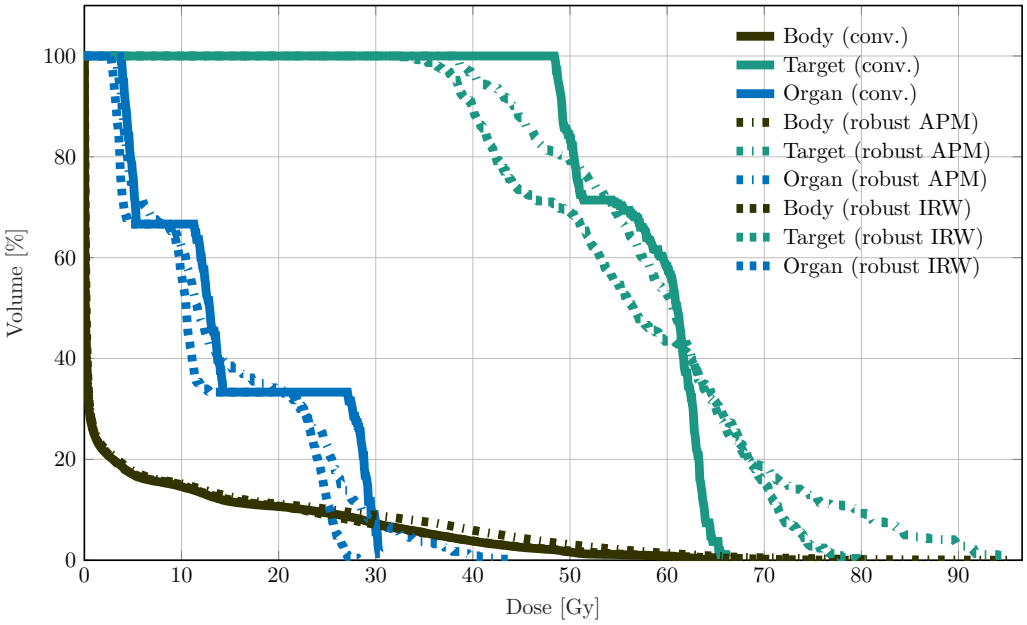


Figure 18.6: Comparison of conventional with robust DVHs computed with expected value optimization based on a pencil beam method and APM vs. Monte Carlo dose calculation with importance reweighting (IRW), respectively.

CHAPTER 19

Discussion

The main use of uncertainty statistics in radiation therapy is for the choice or configuration of more robust treatment plans. In this part, we have discussed two ways of incorporating knowledge about uncertainties into the plan optimization: One based on a discrete set of worst-case scenarios and the other on the optimization of the expected value of the objective function. The main challenge for robust or probabilistic optimization in radiation therapy lies in the high computational costs associated first with computing the dose and further the uncertainty estimates. As a proof-of-concept, we discuss how importance reweighting can be applied to reconcile costly Monte Carlo dose calculations with robust optimization and derive an algorithm to precompute variance influence matrices for expected value optimization.

The optimized plan as well as computed scenarios and variance influences are then compared between the Monte Carlo-based importance reweighting approach and a completely analytical pencil beam algorithm with APM for the uncertainty propagation. The reweighting approach allows a direct construction of scenarios and influence matrices from a regular simulation and significantly decreases the costs of uncertainty quantification, which makes optimization feasible also for Monte Carlo dose engines. We see similar effects of the optimization for both robustness models and both uncertainty propagation methods. Due to the difference already in the dose calculation methods, it is however unclear whether the small deviations that we observe have to be attributed to the dose calculation or uncertainty quantification approach. Further analysis of different test cases would be necessary to investigate this and validate the results. Further, different, time-dependent uncertainty models as discussed in chapter 11 could be incorporated to achieve robustness in a wider, more realistic set of scenarios. The optimization design and solution algorithms presented within this work are merely a meant as a proof-of concept and represent a small excerpt of the possible approaches. Robust or probabilistic optimization is a research field of its own, which could be further explored to find more efficient ways of handling the multiple optimization criteria, uncertainties and constraints in treatment plan optimization. Since Monte Carlo-based optimization,

even with the described adjustments, is costly compared to a simplified analytical approach, the use of dose and uncertainty computations based on the dynamical low-rank approximation is a promising field for future work. Especially in combination with optimization strategies that may exploit the low-rank structure or underlying system of partial differential equations, this approach could be a good alternative to the cheap but very simplified APM pencil beam method and the accurate but costly Monte Carlo-based approach.

Part V

Summary & outlook

Radiation therapy treatment planning especially under consideration of uncertainties offers a variety of mathematical challenges. In this thesis, we have introduced a novel software framework for deterministic dose calculations, tackeled the high-dimensional phase space of the (uncertain) transport equation using the dynamical low-rank approximation and introduced novel efficient methods for uncertainty propagation with MC or determinstic solvers. The following chapter summarizes our findings and discusses limitations as well as potential for future work building on this thesis.

CHAPTER 20

Summary

In this thesis, we have investigated different strategies to reduce the high computational costs associated with treatment planning for charged particle radiation therapy under uncertainties.
First, we considered the core problem of solving the linear Boltzmann transport equation to determine the delivered dose for a given beam set-up. We introduced the transport equation for electrons and protons as well as the continuous slowing down approximation, which allows a use of the energy as a pseudo-time. We then recalled the three most common types of numerical methods used for dose calculations: The Monte Carlo method, which is based on a stochastic interpretation of particle transport as a random walk; the pencil beam algorithm, a simplified probabilistic approach based on Gaussian lateral representations of the beam convolved with measured depth dose data; and lastly deterministic Boltzmann solvers, which take a classic numerical approach of discretizing the six-dimensional phase space of the transport equation.
While the Monte Carlo method is often seen as the gold standard in terms of accuracy, it is also very computationally expensive and converges slowly. Pencil beam models are computationally efficient, however, they can only describe layered heterogeneities in depth (and not lateral direction) accurately. Thus simulation results are inaccurate, especially in cases including air cavities or other inhomogeneities. We therefore focused on the deterministic Boltzmann solvers, which are relatively unknown in the field of medical physics. We introduced a solver framework – KiT-RT – implementing the three most well-known deterministic approaches, P_N, S_N and M_N, for the continuous slowing down approximation. While we could see that the application of deterministic Boltzmann equations for dose calculation can achieve similar accuracy as MC simulations, they also exhibited high computational complexity and memory costs. Therefore, we introduced a novel approach to reduce the computational costs and memory associated with a deterministic solution of the Boltzmann equation: The dynamical low-rank approximation (DLRA) for radiation therapy. Here, the solution is evolved on a low-rank manifold in time. To facilitate the use of boundary conditions and reduce the overall rank, we split the radiation transport equation into col-

lided and uncollided particles through a collision source method. Further, stiff scattering terms were treated through an efficient implicit energy discretization and a rank adaptive integrator was chosen to dynamically adapt the rank in energy. The method can be shown to be L^2-stable under a time step restriction. We then demonstrated that the dynamical low-rank approximation can significantly reduce run-times by more than an order of magnitude for electrons while maintaining high accuracy compared to the full-rank solution as well as a Monte Carlo reference. The method further proved to be well-suited for GPU implementations and especially efficient for proton transport, which requires only a tenth of the rank needed in electron transport.

In part III, we extended the phase space and investigated the transport equation with uncertain parameters. We first discussed relevant uncertainties in radiation therapy and common assumptions or models used in uncertainty propagation. These were found to be extremely simplified and not able to represent more high-dimensional or time-dependent uncertainties. With view to the novel uncertainty propagation approaches discussed later within this part, we introduced efficient ways of modeling uncertainties in tissue density as well as time-dependent spatial uncertainties within the multivariate Gaussian model of the particle phase space.
We then introduced two novel uncertainty propagation methods: The first aimed to reconcile costly Monte Carlo dose calculations with uncertainty quantification. Here, we considered realizations of the uncertain parameter as changes in the probability distribution of the initial phase space parameters and used importance sampling to reconstruct scenario and expected dose estimates from one set of simulated particle trajectories. We further showed that these dose estimates are unbiased and that the overall error when reconstructing several quantities from the same initial sample can be reduced by sampling from a mixture, i.e. weighted sum, of the target probability functions. The method was applied to several patient cases, where run times could be reduced by more than an order of magnitude for all cases.
Since the Monte Carlo-based methods were still relatively costly and additionally required more memory when storing simulated trajectories for later reconstruction, we next considered a use of DLRA also for uncertainty quantification. We demonstrated how a tensor DLRA approach can be applied to the continuous slowing down approximation with uncertainties. This approach allowed a direct propagation of time-dependent uncertainties in the tissue densities. The computed expected value and variance of the dose were validated against a collocation approach. Even in this first proof-of-concept implementation, the results agreed well with the reference and the computational costs could be reduced significantly.

Lastly, in part IV, we joined together the previous methods to determine the optimal treatment set-up while considering uncertainties. Two options for the design of the robust optimization problem were discussed, one based on a discrete set of error scenarios in addition to the nominal dose, and the other based on the expected value of a squared deviation objective function. The latter required the computation of the dose expected value as well as a variance term. We further elaborated on how the optimization itself can be sped up by precomputing influence matrices or tensors for a set of viable beam options and merely optimizing the weights used for their accumulation. We derived an algorithm to efficiently compute these influence terms using importance reweighting of a MC dose simulation and argue that it produces positive semidefinite covariance matrices. We find that this approach makes robust optimization (more) feasible in combination with MC dose calculations and observe similar effects on the plan quality for both objective function designs and in comparison to an analytical pencil beam-based approach.

CHAPTER 21

Outlook

The presented work leaves several open questions and tasks that could be tackled in future research, both with view to a better mathematical understanding and in bringing the methods closer to clinical applicability.

Dynamical low rank approximation

To improve the accuracy of dose calculations for proton transport, more exact physical models are required. We have seen in chapter 7 that incorporating nuclear scattering could close the majority of the gap to the Monte Carlo reference. Unfortunately, cross section data for nuclear scattering is only available for specific materials and medium to high energy ranges. Since available mathematical models are usually obtained by regressing on this data, they inherit their limitations. Future work could aim at finding better mathematical models for nuclear scattering, e.g., by translating the often more exact interaction models used for Monte Carlo simulations to differential scattering cross sections. Similarly, the Rutherford formula can accurately model single Coulomb scattering events. However, the macroscopic effect of multiple Coulomb scattering cannot be accurately modeled with Rutherford scattering, as the accumulation of individual Rutherford events cannot be realistically resolved. Models derived from, for example, Molière's theory would need to be adapted to the formulation. The use of dedicated multiple Coulomb scattering models could also aid in alleviating problems with the singularity of Rutherford cross sections around zero angular deflection. In contrast to Rutherford's formula, multiple scattering typically employs aggregate models for the many small-angle scattering events, resulting in a nearly Gaussian distribution in angle. [215] state that the Boltzmann scattering operator is not well-suited for small-angle scattering and proposes a hybrid Fokker-Planck-Boltzmann method. Future research could investigate such a hybrid

approach using Rutherford scattering in comparison to a Boltzmann method with a multiple Coulomb scattering model.

So far, time evolution equations for the proposed dynamical low-rank approximation have only been derived based on the continuous slowing down approximation to the linear Boltzmann transport equation. This allows a treatment of the energy as a pseudo-time. Certain physical effects relevant to proton transport, such as range/energy straggling however introduce stochasticity to the energy loss which cannot be modeled exactly within a continuous slowing down setting. Further, the most commonly used treatment modality relies on photons for which the continuous slowing down approximation does not hold. Thus, future work could explore the use of the dynamical low-rank approximation for the classical Boltzmann transport equation. Possible approaches could include adapting [115], who propose to use the iteration index in a power iteration scheme as a pseudo-time when solving a neutron transport problem with multi-group discretizations of the energy variable. In photon transport, a hybrid approach might be taken, which uses the dynamical low-rank approximation for the electron transport equation (where continuous slowing down can be assumed) and a coarse deterministic or pencil beam method for the coupled photon equation.

With view to a clinical application, the optimization of the GPU implementation as well as thorough quantitative comparisons to state-of-the-art commercial dose calculation engines would be of interest to realistically judge the speed and memory capabilities on workstations used in treatment planning. Further, the current work is limited to the computation of single (pencil) beams. An extension to an IMRT setting could be achieved using Gaussian mixtures for the boundary conditions modeling the radiation source, where the weights in the mixture model are determined during the enveloping optimization process. Monte Carlo methods typically compute beams individually and merely use variance reduction techniques such as importance sampling to reduce the number of trajectories that need to be simulated. Therefore, it would be especially interesting to explore the effect of more complex boundary conditions on the costs of dose calculation using the dynamical low-rank approximation.

Lastly, the dynamical low-rank approximation is still a relatively young topic and very much developing. In the time during and after this work was conducted a lot of new and interesting research, e.g., on fully parallel integrators has been published, which could be built upon and brought to an application in radiation therapy. Especially the topic of tensor DLRA has merely been touched upon here. While we have seen promising results for an application for uncertainty quantification in radiation therapy, there is a lot of potential for future work. This could be in exploring different tensor formats or a use for updates in (robust) plan optimization as suggested in [139]. Further, uncertainty quantification for dose computations with the dynamical low-rank approximation could be handled non-intrusively. For example, the choice of rank could be utilized as the lever for different levels of accuracy and cost in a multi-level Monte Carlo approach.

Uncertainty Modeling

The uncertainty models introduced in chapter 11 exploit the time-dependency of beam application patterns in intensity-modulated radiation therapy to include spatial shifts during the treatment in the initial phase space distribution. In this work, we have only considered the uses of these models as input for un-

certainty propagation methods. This however limits the format to multivariate (Gaussian) distributions and fixes the temporal resolution to the intervals between beam application times. More complex and higher resolution models could be directly used for robust optimization which assumes spatial invariance [43] or for real-time predictions of tumor or organ positions needed in inline radiation therapy.

Further, recent developments in adaptive (real-time) radiotherapy interact with uncertainty modeling and robustness requirements. Uncertainty mitigation techniques, such as gating (irradiation is stopped if the patient moves out of a predefined window) or tracking (beam follows the movement of the patient) can affect the distribution of the net displacement of the patient in relation to the beam. This needs to be included in the uncertainty models to account for such effects in robust optimization and avoid over-corrections. For example, [92] propose a truncated Gaussian model to account for gating during margin computations. Tracking on the other hand could result in a narrowed distribution within the tolerance window. Such probabilistic models can be directly included in our proposed uncertainty framework. In the case of tracking, however, additional range uncertainties would have to be considered, since both the target and beam can constantly change their lateral position and encounter different tissue compositions.

Importance (re-)weighting for MC dose calculations

Importance (re-)weighting was introduced as a method to improve the efficiency of sampling or collocation-based uncertainty quantification for Monte Carlo simulations. It relies on the derivation of probability distributions representing the realizations of uncertain parameters. In our work so far, we have used the common assumption of Gaussian phase space and uncertainty distributions. This is however not a reasonable assumption for all types of uncertainties or treatment modalities. For example, photons are typically physically forced into a flattened nearly uniform distribution. Future work could consider the use of non-Gaussian probability distributions and investigate to which extent joint and conditional probabilities can be derived in these cases.

One could further lean more into the combination of importance (re-)weighting with non-intrusive spectral projections. While we have seen that this does not improve efficiency for expected value or variance computations, the functional representation of dose in terms of uncertain parameters could be used for sensitivity analyses or in robust optimization.

Robust optimization

As the main focus of this work lies on radiation transport and uncertainty quantification, we have merely scratched the surface of also using this information for robust treatment plan optimization. There are many options for future work, ranging from different optimization designs, e.g., using probabilistic constraints to different optimization algorithms. Within a closer vicinity to the methodology used here, expected value optimization could be extended to a wider range of objective functions. For a squared deviation objective, we saw that the expected value could be simplified and brought into a form where expensive quantities can be precomputed. Future work could explore whether similar results can be obtained for other common objective functions, e.g. based on equivalent uniform dose (EUD) or dose volume histograms.

More broadly, we could study how the uncertainty propagation techniques discussed in this work can most efficiently be combined with different ways of incorporating robustness and balancing the trade-off between different objectives. For example, the variance influence matrices for individual volumes of interest computed using importance reweighting could be used for Pareto optimization. On the other hand, as discussed in previous sections, deterministic dose calculation methods and specifically the dynamical low-rank approximation could lend themselves to very cost-efficient optimization algorithms.

Appendix

A.1 Detailed results of γ-analysis in chapter 14

Table A1 presents more detailed results of the γ-analysis shown in chapter 14, including a breakdown into passrates per volume of interest in each patient and additional data on the liver patient for set-up uncertainties.

Table A1: $\gamma_{2\,\%}^{2\,\mathrm{mm}}$-pass rates in volumes of interest (VOI) of the water phantom, liver and prostrate patient for (a) set-up errors, (b) range errors and (c) set-up and range errors. All estimates were computed from the nominal distribution S_0 and in (b) and (c) also compared against such from the expected distribution $\mathcal{S}$.

(a) Set-up errors

Patient	VOI									
Waterbox	Total	Body	Target							
$\mathbb{E}[\boldsymbol{D}]$	99.95	99.95	99.81							
$\sigma(\boldsymbol{D})$	98.04	98.04	98.44							
Liver	Total	GTV	Liver	Heart	CTV	Contour	PTV			
$\mathbb{E}[\boldsymbol{D}]$	99.01	98.07	97.71	99.83	94.99	97.83	96.59			
$\sigma(\boldsymbol{D})$	99.81	100	99.79	99.67	100	99.72	99.49			
Prostrate	Total	Rectum	Penile bulb	Lymph nodes	Femoral heads	Prostate bed	PTV 68	PTV 56	Bladder	Body
$\mathbb{E}[\boldsymbol{D}]$	99.99	100	100	99.75	100	99.71	99.85	99.88	99.96	
$\sigma(\boldsymbol{D})$	99.82	99.85	95.92	100	100	100	100	99.36	99.94	99.79

(b) Range errors

Distribution	Patient	VOI						
	Waterbox	Total	Body	Target				
S_0	$\mathbb{E}[\boldsymbol{D}]$	99.58	99.58	99.44				
	$\sigma(\boldsymbol{D})$	91.63	91.63	92.58				
$\mathcal{S}$	$\boldsymbol{D}$	99.98	99.98	100				
	$\sigma(\boldsymbol{D})$	93.55	93.55	87.88				
	Liver	Total	GTV	Liver	Heart	CTV	Contour	PTV
S_0	$\mathbb{E}[\boldsymbol{D}]$	99.91	99.71	99.91	100	99.74	99.93	99.81
	$\sigma(\boldsymbol{D})$	73.32	79.47	70.68	83.38	66.49	72.16	57.38
$\mathcal{S}$	$\boldsymbol{D}$	99.93	99.90	99.96	100	99.85	99.85	99.91
	$\sigma(\boldsymbol{D})$	93.70	99.05	92.17	95.51	90.06	93.80	86.22

(c) Set-up and range errors

Distribution	Patient	VOI						
	Waterbox	Total	Body	Target				
S_0	$\mathbb{E}[\boldsymbol{D}]$	99.39	99.39	99.81				
	$\sigma(\boldsymbol{D})$	95.10	95.10	82.76				
$\mathcal{S}$	$\boldsymbol{D}$	99.93	99.93	99.81				
	$\sigma(\boldsymbol{D})$	99.50	99.50	97.04				
	Liver	Total	GTV	Liver	Heart	CTV	Contour	PTV
S_0	$\mathbb{E}[\boldsymbol{D}]$	99.86	99.71	99.86	99.93	99.61	99.91	99.67
	$\sigma(\boldsymbol{D})$	95.56	96.76	89.87	87.77	95.85	91.74	91.89
$\mathcal{S}$	$\boldsymbol{D}$	99.87	100	99.94	100	99.77	99.97	99.86
	$\sigma(\boldsymbol{D})$	96.64	100	93.39	84.96	98.67	92.50	92.58

A.2 Full Gaussian phase space model

The simplified model used in 14 can be extended to include distributions in the initial momentum direction $\boldsymbol{\Omega}_0$. The additional variables are included in the Gaussian mixture model:

$$S_0(\boldsymbol{Z}) = \sum_{b=1}^{B} w_b S_0^b(\boldsymbol{Z}) \, , S_0^b(\boldsymbol{Z}) = S_0^b(\boldsymbol{r}_0, \boldsymbol{\Omega}_0, E_0) = \mathcal{N}(\boldsymbol{\mu}_Z^b, \boldsymbol{\Sigma}_Z^b) \tag{1}$$

where the respective entries in the covariance matrix Σ_Z^b can be chosen $\neq 0$ to model randomness in the momentum direction as well as correlations of the momentum directions with primary particle positions.

Table A4: $\gamma_{2\,\%}^{2\,\mathrm{mm}}$-pass rates in volumes of interest (VOI) of the liver patient computed using the full phase space parameterizations.

	From S_0			From $\mathcal{S}$		
Error type	Set-up					
Liver	$\boldsymbol{D}$	$\mathbb{E}[\boldsymbol{D}]$	$\boldsymbol{\sigma(D)}$	$\boldsymbol{D}$	$\mathbb{E}[\boldsymbol{D}]$	$\boldsymbol{\sigma(D)}$
Overall	100	99.86	95.88	99.85	100	96.32
GTV	100	99.52	95.09	99.23	100	97.21
Liver	100	99.96	94.63	99.93	100	95.32
Heart	100	100	95.10	100	100	97.04
CTV	100	99.79	92.78	99.67	100	94.22
Contour	100	99.99	95.89	99.97	100	96.30
PTV	100	99.89	90.92	99.78	100	89.93

Figure 1 presents results for the nominal dose, expected dose and standard deviation in a liver patient, for set-up uncertainties with 3 mm standard deviation, 0.2 standard deviation in the momentum direction

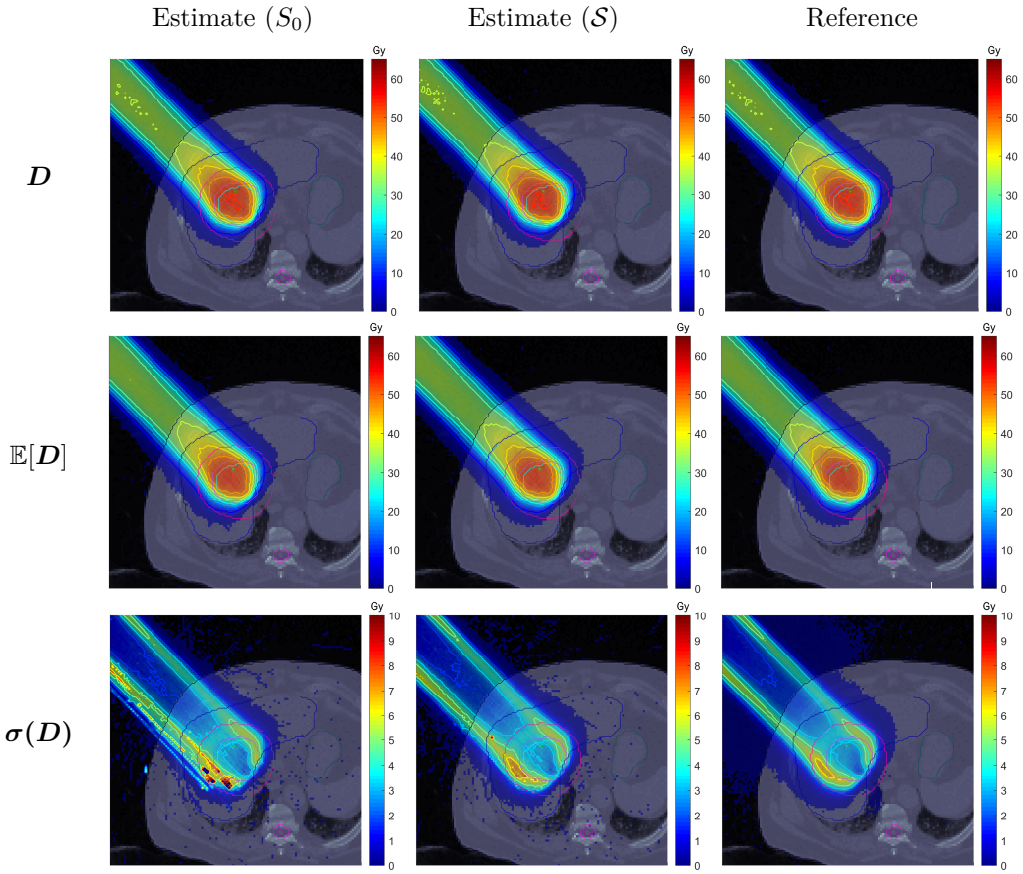


Figure 1: Nominal dose, expected dose and standard deviation w.r.t. set-up uncertainties with 3 mm standard deviation for one beam (couch angle 0°, gantry angle 315°), computed using the full phase space parameterizations.

and 0.3 correlation between Ω_v and r_v, $v \in \{x, y\}$. Estimates were computed based on the convolution function $\mathcal{S}$ of the error and beam parameter densities, as well as the nominal parameter density S_0. The corresponding global γ-analysis pass rates can be found in table A4.

Bibliography

[1] Abdessalem, Anis Ben, Dervilis, Nikolaos, Wagg, David J, and Worden, Keith. "Automatic kernel selection for gaussian processes regression with approximate bayesian computation and sequential monte carlo". In: *Frontiers in Built Environment* 3 (2017), p. 52 (cit. on p. 88).

[2] Adams, Marvin L and Larsen, Edward W. "Fast iterative methods for discrete-ordinates particle transport calculations". In: *Progress in nuclear energy* 40.1 (2002), pp. 3–159 (cit. on p. 50).

[3] Ahnesjö, Anders, Saxner, Mikael, and Trepp, Avo. "A pencil beam model for photon dose calculation". In: *Medical Physics* 19.2 (1992), pp. 263–273. DOI: https://doi.org/10.1118/1.596856 (cit. on p. 35).

[4] Aistleitner, Christoph and Dick, Josef. "Functions of bounded variation, signed measures, and a general Koksma-Hlawka inequality". In: *Acta Arithmetica* 167.2 (2015), pp. 143–171 (cit. on p. 102).

[5] Alcouffe, RE, O'Dell, RD, and Brinkley Jr, FW. "A first-collision source method that satisfies discrete Sn transport balance". In: *Nuclear Science and Engineering* 105.2 (1990), pp. 198–203 (cit. on p. 51).

[6] Alldredge, Graham W, Frank, Martin, and Hauck, Cory D. "A regularized entropy-based moment method for kinetic equations". In: *SIAM Journal on Applied Mathematics* 79.5 (2019), pp. 1627–1653 (cit. on pp. 38, 42).

[7] An, Yu, Liang, Jianming, Schild, Steven E, Bues, Martin, and Liu, Wei. "Robust treatment planning with conditional value at risk chance constraints in intensity-modulated proton therapy". In: *Medical physics* 44.1 (2017), pp. 28–36 (cit. on p. 146).

[8] Arndt, Richard A, Strakovsky, Igor I, and Workman, Ron L. "Nucleon-nucleon elastic scattering to 3 GeV". In: *Physical Review C* 62.3 (2000), p. 034005 (cit. on p. 27).

[9] Attix, Frank Herbert. *Introduction to radiological physics and radiation dosimetry*. John Wiley & Sons, 2008 (cit. on p. 29).

[10] Bangert, Mark, Hennig, Philipp, and Oelfke, Uwe. "Analytical Probabilistic Modeling for Radiation Therapy Treatment Planning". In: *Physics in Medicine and Biology* 58.16 (2013), pp. 5401–5419. DOI: 10.1088/0031-9155/58/16/5401 (cit. on pp. 35, 81, 83, 107, 108, 141, 142).

[11] Bangert, Mark and Ziegenhein, Peter. "Bestrahlungsplanung". In: *Medizinische Physik: Grundlagen – Bildgebung – Therapie – Technik*. Ed. by Wolfgang Schlegel, Christian P. Karger, and Oliver Jäkel. Berlin, Heidelberg: Springer, 2018, pp. 485–514. DOI: 10.1007/978-3-662-54801-1_24 (cit. on pp. v, 13).

[12] Barnard, Richard, Frank, Martin, and Herty, Michael. "Optimal radiotherapy treatment planning using minimum entropy models". In: *Applied Mathematics and Computation* 219.5 (2012), pp. 2668–2679 (cit. on p. 38).

[13] Barschall, H. H., Chadwick, M. B., Jones, D. T. L., Meulders, J. P., Schuhmacher, H., and Young, P. G. "ICRU Report 63, Nuclear Data for Neutron and Proton Radiotherapy and for Radiation Protection". In: *Journal of the ICRU* os-32.2 (2000). URL: https://journals.sagepub.com/toc/crub/os-32/2 (cit. on pp. 27, 69).

[14] Baskar, Rajamanickam, Lee, Kuo Ann, Yeo, Richard, and Yeoh, Kheng-Wei. "Cancer and radiation therapy: current advances and future directions". In: *International journal of medical sciences* 9.3 (2012), p. 193 (cit. on p. iv).

[15] Beckman, Richard J, Mckay, Michael D, Beckman, Richard J, and Mckay, Michael D. "Monte Carlo Estimation Under Different Distributions Using the Same Simulation". In: 29.2 (1987), pp. 153–160 (cit. on p. 110).

[16] Benhamou, Eric. *A Few Properties of Sample Variance*. 2018 (cit. on p. 117).

[17] Berger, MJ, Coursey, JS, Zucker, MA, and Chang, J. "Stopping-Power & Range Tables for Electrons, Protons, and Helium Ions". In: *NIST Standard Reference Database 124, NIST* (2017) (cit. on pp. 24, 25, 28, 69).

[18] Bethe, Hans. "Zur theorie des durchgangs schneller korpuskularstrahlen durch materie". In: *Annalen der Physik* 397.3 (1930), pp. 325–400 (cit. on pp. 21, 28, 82).

[19] Bethe, Hans A. "Passage of radiations through matter". In: *Experimental nuclear physics* (1953) (cit. on pp. 29, 82).

[20] Bichsel, Hans and Hiraoka, T. "Energy spectra and depth-dose curves for 70 MeV protons". In: *International Journal of Quantum Chemistry* 36.S23 (1989), pp. 565–574 (cit. on p. 29).

[21] Billingsley, P. "Probability and Measure. 3rd Wiley". In: *New York* (1995) (cit. on p. 101).

[22] Blatman, Géraud and Sudret, Bruno. "An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis". In: *Probabilistic Engineering Mechanics* 25.2 (2010), pp. 183–197 (cit. on p. 105).

[23] Blatman, Géraud and Sudret, Bruno. "Adaptive sparse polynomial chaos expansion based on least angle regression". In: *Journal of computational Physics* 230.6 (2011), pp. 2345–2367 (cit. on p. 105).

[24] Bloch, Felix. "Zur bremsung rasch bewegter teilchen beim durchgang durch materie". In: *Annalen der Physik* 408.3 (1933), pp. 285–320 (cit. on pp. 21, 28, 82).

[25] Bortfeld, Thomas. "An analytical approximation of the Bragg curve for therapeutic proton beams". In: *Medical Physics* 24.12 (1997), pp. 2024–2033. DOI: https://doi.org/10.1118/1.598116. (Visited on 04/16/2021) (cit. on pp. 29, 36, 70, 82, 108).

[26] Bortfeld, Thomas, Chan, Timothy CY, Trofimov, Alexei, and Tsitsiklis, John N. "Robust management of motion uncertainty in intensity-modulated radiation therapy". In: *Operations Research* 56.6 (2008), pp. 1461–1473 (cit. on p. 145).

[27] Butkevich, AV, Kokoulin, RP, Matushko, GV, and Mikheyev, SP. "Comments on multiple scattering of high-energy muons in thick layers". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 488.1-2 (2002), pp. 282–294 (cit. on p. 27).

[28] Caflisch, Russel E. "Monte Carlo and quasi-Monte Carlo methods". In: *Acta Numerica* 7 (1998), pp. 1–49. DOI: 10.1017/S0962492900002804. (Visited on 12/14/2020) (cit. on p. 101).

[29] Cambou, Mathieu, Hofert, Marius, and Lemieux, Christiane. "Quasi-random numbers for copula models". In: *Statistics and Computing* 27.5 (2017), pp. 1307–1329 (cit. on pp. 101, 102).

[30] Cameron, Robert H and Martin, William T. "The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals". In: *Annals of Mathematics* (1947), pp. 385–392 (cit. on p. 99).

[31] Camminady, Thomas, Frank, Martin, and Kusch, Jonas. "Highly uniform quadrature sets for the discrete ordinates method". In: *Nuclear Science and Engineering* (2021) (cit. on p. 38).

[32] Case, Kenneth M and Zweifel, Paul Frederick. "Linear transport theory". In: (1967) (cit. on p. 37).

[33] Casiraghi, M, Albertini, F, and Lomax, Antony J. "Advantages and limitations of the 'worst case scenario'approach in IMPT treatment planning". In: *Physics in Medicine & Biology* 58.5 (2013), p. 1323 (cit. on p. 147).

[34] Ceruti, Gianluca. "Unconventional contributions to dynamical low-rank approximation of tree tensor networks". PhD thesis. Eberhard Karls Universität Tübingen, 2021 (cit. on p. 137).

[35] Ceruti, Gianluca, Kusch, Jonas, and Lubich, Christian. "A rank-adaptive robust integrator for dynamical low-rank approximation". In: *BIT Numerical Mathematics* (2022), pp. 1–26 (cit. on pp. 50, 54, 61, 63).

[36] Ceruti, Gianluca and Lubich, Christian. "An unconventional robust integrator for dynamical low-rank approximation". In: *BIT Numerical Mathematics* (2021), pp. 1–22 (cit. on pp. 49, 54, 136, 137).

[37] Chu, Millie, Zinchenko, Yuriy, Henderson, Shane G, and Sharpe, Michael B. "Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty". In: *Physics in Medicine & Biology* 50.23 (2005), p. 5463 (cit. on p. 146).

[38] Chu, Millie, Zinchenko, Yuriy, Henderson, Shane G., and Sharpe, Michael B. "Robust Optimization for Intensity Modulated Radiation Therapy Treatment Planning under Uncertainty". In: *Physics in Medicine and Biology* 50.23 (2005), pp. 5463–5477. DOI: 10.1088/0031-9155/50/23/003 (cit. on p. 118).

[39] Clark, K et al. "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository". In: *Journal of Digital Imaging* 26.6 (2013), pp. 1045–1057. DOI: 10.1007/s10278-013-9622-7 (cit. on p. 45).

[40] Clark, Kenneth, Vendt, Bruce, Smith, Kirk, Freymann, John, Kirby, Justin, Koppel, Paul, Moore, Stephen, Phillips, Stanley, Maffitt, David, Pringle, Michael, et al. "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository". In: *Journal of digital imaging* 26.6 (2013), pp. 1045–1057 (cit. on p. 90).

[41] Clenshaw, Charles W and Curtis, Alan R. "A method for numerical integration on an automatic computer". In: *Numerische Mathematik* 2.1 (1960), pp. 197–205 (cit. on p. 103).

[42] Craft, David, Bangert, Mark, Long, Troy, Papp, Dávid, and Unkelbach, Jan. "Shared Data for Intensity Modulated Radiation Therapy (IMRT) Optimization Research: The CORT Dataset". In: *GigaScience* 3.1 (2014). DOI: 10.1186/2047-217X-3-37 (cit. on pp. 95, 120).

[43] Craig, T, Battista, J, and Van Dyk, J. "Incorporating geometric uncertainties into dose calculations with convolution: the effect of spatial invariance". In: *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*. Vol. 4. IEEE. 2000, pp. 3253–3256 (cit. on pp. 146, 163).

[44] Davison, Boris. *Neutron Transport Theory*. Oxford University Press, 1957 (cit. on p. 34).

[45] Davisson, Charlotte Meaker and Evans, Robley D. "Gamma-Ray Absorption Coefficients". In: *Rev. Mod. Phys.* 24 (2 1952), pp. 79–107. DOI: 10.1103/RevModPhys.24.79 (cit. on p. 31).

[46] De Lathauwer, Lieven, De Moor, Bart, and Vandewalle, Joos. "A multilinear singular value decomposition". In: *SIAM journal on Matrix Analysis and Applications* 21.4 (2000), pp. 1253–1278 (cit. on p. 136).

[47] Dektor, Alec, Rodgers, Abram, and Venturi, Daniele. "Rank-adaptive tensor methods for high-dimensional nonlinear PDEs". In: *Journal of Scientific Computing* 88.2 (2021), pp. 1–27 (cit. on p. 50).

[48] Deng, Lih-Yuan and Bowman, Dale. "Developments in pseudo-random number generators". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 9.5 (2017), e1404 (cit. on p. 101).

[49] Drzymala, R.E., Mohan, R., Brewster, L., Chu, J., Goitein, M., Harms, W., and Urie, M. "Dose-volume histograms". In: *International Journal of Radiation Oncology*Biology*Physics* 21.1 (1991). Three-Dimensional Photon Treatment Planning Report of the Collaborative Working Group on the Evaluation of Treatment Planning for External Photon Beam Radiotherapy, pp. 71–78. DOI: https://doi.org/10.1016/0360-3016(91)90168-4 (cit. on p. 16).

[50] Dürichen, Robert, Pimentel, Marco A. F., Clifton, Lei, Schweikard, Achim, and Clifton, David A. "Multitask Gaussian Processes for Multivariate Physiological Time-Series Analysis". In: *IEEE transactions on bio-medical engineering* 62.1 (2015), pp. 314–322. DOI: 10.1109/TBME.2014.2351376 (cit. on pp. 85, 86).

[51] Duvenaud, David. "Automatic Model Construction with Gaussian Processes". PhD thesis. 2014 (cit. on pp. 86, 88).

[52] Ebrahimi, Saba and Lim, Gino J. "A reinforcement learning approach for finding optimal policy of adaptive radiation therapy considering uncertain tumor biological response". In: *Artificial Intelligence in Medicine* 121 (2021), p. 102193 (cit. on p. 79).

[53] Ehrgott, Matthias, Güler, Çiğdem, Hamacher, Horst W, and Shao, Lizhen. "Mathematical optimization in intensity modulated radiation therapy". In: *Annals of Operations Research* 175.1 (2010), pp. 309–365 (cit. on pp. 12–14).

[54] Einkemmer, Lukas. "A low-rank algorithm for weakly compressible flow". In: *SIAM Journal on Scientific Computing* 41.5 (2019), A2795–A2814 (cit. on p. 54).

[55] "Elastic Scattering of Electrons and Positrons". In: *Journal of the ICRU* 7.1 (2007). PMID: 24170956, pp. iii–iii. DOI: 10.1093/jicru_ndm001 (cit. on p. 24).

[56] Eldred, Michael and Burkardt, John. "Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification". In: *47th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition*. 2009, p. 976 (cit. on p. 102).

[57] Fernández-Varea, JM, Mayol, R, and Salvat, F. "Cross sections for elastic scattering of fast electrons and positrons by atoms". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 82.1 (1993), pp. 39–45 (cit. on p. 27).

[58] Fippel, Matthias and Soukup, Martin. "A Monte Carlo dose calculation algorithm for proton therapy". In: *Medical physics* 31.8 (2004), pp. 2263–2273 (cit. on p. 27).

[59] Fredriksson, Albin. "A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization". In: *Medical physics* 39.8 (2012), pp. 5169–5181 (cit. on pp. 146, 147).

[60] Fredriksson, Albin and Bokrantz, Rasmus. "A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning". In: *Medical physics* 41.8Part1 (2014), p. 081701 (cit. on p. 146).

[61] Fredriksson, Albin, Forsgren, Anders, and Hårdemark, Björn. "Minimax optimization for handling range and setup uncertainties in proton therapy". In: *Medical Physics* 38.3 (2011), pp. 1672–1684. DOI: 10.1118/1.3556559 (cit. on pp. 35, 81, 107).

[62] Ganapol, BD. "Homogeneous infinite media time-dependent analytic benchmarks for X-TM transport methods development". In: *Los Alamos National Laboratory* (1999) (cit. on pp. 46, 63).

[63] Ganapol, BD. "Analytical benchmarks for nuclear engineering applications". In: *Case Studies in Neutron Transport Theory* (2008) (cit. on p. 63).

[64] Garrett, C Kristopher and Hauck, Cory D. "A comparison of moment closures for linear kinetic transport equations: The line source benchmark". In: *Transport Theory and Statistical Physics* 42.6-7 (2013), pp. 203–235 (cit. on p. 63).

[65] Ghanem, Roger G. and Spanos, Pol D. *Stochastic Finite Elements: A Spectral Approach*. New York: Springer-Verlag, 1991 (cit. on p. 98).

[66] Giles, Michael B. "Multilevel monte carlo path simulation". In: *Operations research* 56.3 (2008), pp. 607–617 (cit. on p. 135).

[67] Goldstein, Herbert, Poole, Charles, and Safko, John. *Classical mechanics*. 2002 (cit. on p. 27).

[68] Gottschalk, Bernard, Koehler, AM, Schneider, RJ, Sisterson, JM, and Wagner, MS. "Multiple Coulomb scattering of 160 MeV protons". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 74.4 (1993), pp. 467–490 (cit. on pp. 36, 108).

[69] Hachiya, Hirotaka, Akiyama, Takayuki, Sugiayma, Masashi, and Peters, Jan. "Adaptive Importance Sampling for Value Function Approximation in Off-Policy Reinforcement Learning". In: *Neural Networks* 22.10 (2009), pp. 1399–1410. DOI: 10.1016/j.neunet.2009.01.002 (cit. on p. 110).

[70] Halton, J. H. "Radical-Inverse Quasi-Random Point Sequence". In: *Commun. ACM* 7.12 (1964), pp. 701–702. DOI: 10.1145/355588.365104 (cit. on p. 102).

[71] Hastings, W. K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970). Publisher: [Oxford University Press, Biometrika Trust], pp. 97–109. DOI: 10.2307/2334940. (Visited on 01/28/2021) (cit. on pp. 110, 113).

[72] Hauck, Cory D and McClarren, Ryan G. "A collision-based hybrid method for time-dependent, linear, kinetic transport equations". In: *Multiscale Modeling & Simulation* 11.4 (2013), pp. 1197–1227 (cit. on pp. 50, 51).

[73] Heinrich, Stefan. "Monte Carlo complexity of global solution of integral equations". In: *Journal of complexity* 14.2 (1998), pp. 151–175 (cit. on p. 135).

[74] Hensel, Hartmut, Iza-Teran, Rodrigo, and Siedow, Norbert. "Deterministic model for dose calculation in photon radiotherapy". In: *Physics in Medicine & Biology* 51.3 (2006), p. 675 (cit. on pp. 29, 31).

[75] Hesterberg, Tim. "Weighted Average Importance Sampling and Defensive Mixture Distributions". In: *Technometrics* 37.2 (1995), pp. 185–194. DOI: 10.1080/00401706.1995.10484303 (cit. on p. 116).

[76] Hesthaven, Jan S, Pagliantini, Cecilia, and Ripamonti, Nicolò. "Rank-adaptive structure-preserving reduced basis methods for Hamiltonian systems". In: *arXiv preprint arXiv:2007.13153* (2020) (cit. on p. 50).

[77] Highland, Virgil L. "Some practical remarks on multiple scattering". In: *Nuclear Instruments and Methods* 129.2 (1975). Erratum in Nucl. Instr. and Meth. 161 (1979) 171, pp. 497–499 (cit. on p. 27).

[78] Hlawka, Edmund and Mück, R. "Über eine Transformation von gleichverteilten Folgen II". In: *Computing* 9.2 (1972), pp. 127–138 (cit. on p. 102).

[79] Hochbruck, Marlis, Neher, Markus, and Schrammer, Stefan. "Rank-adaptive dynamical low-rank integrators for first-order and second-order matrix differential equations". In: CRC 1173 Preprint 2022.13 (2022). DOI: 10.5445/IR/1000143198 (cit. on p. 50).

[80] Hogstrom, Kenneth R and Almond, Peter R. "Review of electron beam therapy physics". In: *Physics in Medicine & Biology* 51.13 (2006), R455 (cit. on p. 23).

[81] Hogstrom, Kenneth R, Mills, Michael D, and Almond, Peter R. "Electron beam dose calculations". In: *Physics in Medicine & Biology* 26.3 (1981), p. 445 (cit. on p. 35).

[82] Holtz, Sebastian, Rohwedder, Thorsten, and Schneider, Reinhold. "The alternating linear scheme for tensor optimization in the tensor train format". In: *SIAM Journal on Scientific Computing* 34.2 (2012), A683–A713 (cit. on p. 140).

[83] Homma, Noriyasu, Sakai, Masao, Endo, Haruna, Mitsuya, Masatoshi, Takai, Yoshihiro, and Yoshizawa, Makoto. "A New Motion Management Method for Lung Tumor Tracking Radiation Therapy". In: *WSEAS TRANSACTIONS on SYSTEMS* 8 (2009), pp. 471–480 (cit. on pp. 85, 86).

[84] Hong, Linda, Goitein, Michael, Bucciolini, Marta, Comiskey, Robert, Gottschalk, Bernard, Rosenthal, Skip, Serago, Chris, and Urie, Marcia. "A pencil beam algorithm for proton dose calculations". In: *Physics in Medicine & Biology* 41.8 (1996), p. 1305 (cit. on p. 35).

[85] Hosder, Serhat, Walters, Robert, and Balch, Michael. "Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables". In: *48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 2007, p. 1939 (cit. on p. 106).

[86] Hu, A. and Qi, S. "Time Series Analysis of Interfraction Patient Setup in Image Guided Radiation Therapy". In: *International Journal of Radiation Oncology, Biology, Physics* 84.3 (2012), S736–S737. DOI: 10.1016/j.ijrobp.2012.07.1970 (cit. on p. 85).

[87] Hu, Jingwei and Wang, Yubo. "An adaptive dynamical low rank method for the nonlinear Boltzmann equation". In: *arXiv preprint arXiv:2112.02695* (2021) (cit. on pp. 50, 60).

[88] Hugo, Geoffrey D, Weiss, Elisabeth, Sleeman, William C, Balik, Salim, Keall, Paul J, Lu, Jun, and Williamson, Jeffrey F. *Data from 4D lung imaging of NSCLC patients*. 2016. DOI: 10.7937/K9/TCIA.2016.ELN8YGLE (cit. on pp. 90, 92, 139).

[89] Hussein, Mohammad, Clark, Catharine, and Nisbet, Andrew. "Challenges in Calculation of the Gamma Index in Radiotherapy – Towards Good Practice". In: *Physica Medica* 36 (2017), pp. 1–11. DOI: 10.1016/j.ejmp.2017.03.001 (cit. on p. 18).

[90] Ivanchenko, VN, Kadri, O, Maire, M, and Urban, L. "Geant4 models for simulation of multiple scattering". In: *Journal of Physics: Conference Series*. Vol. 219. 3. IOP Publishing. 2010, p. 032045 (cit. on p. 27).

[91] Jablonski, A, Salvat, F, and Powell, CJ. "NIST electron elastic-scattering cross-section database". In: *NIST Standard Reference Database* 64.1 (2010) (cit. on p. 24).

[92] Janssen, Tomas M., van der Heide, Uulke A., Remeijer, Peter, Sonke, Jan-Jakob, and van der Bijl, Erik. "A margin recipe for the management of intra-fraction target motion in radiotherapy". In: *Physics and Imaging in Radiation Oncology* 24 (2022), pp. 159–166. DOI: https://doi.org/10.1016/j.phro.2022.11.008 (cit. on pp. 145, 163).

[93] Jarrel, J.J. and Adams, M.L. "Discrete-ordinates quadrature sets based on linear discontinuous finite elements". In: *Proc. International Conference on Mathematics and Computational Methods applied to Nuclear Science and Engineering* (2011) (cit. on p. 38).

[94] Jörres, Christian. "Numerical methods for Boltzmann transport equations in radiotherapy treatment planning". PhD thesis. Universitätsbibliothek der RWTH Aachen, 2015 (cit. on p. 20).

[95] Kahn, H. "Random Sampling (Monte Carlo) Techniques in Neutron Attenuation Problems–I". In: *Nucleonics* 6.5 (1950), 27, passim (cit. on pp. 110, 113).

[96] Kanai, Tatsuaki, Kawachi, Kiyomitsu, Kumamoto, Yoshikazu, Ogawa, Hirotsugu, Yamada, Takanobu, Matsuzawa, Hideo, and Inada, Tetsuo. "Spot Scanning System for Proton Radiotherapy". In: *Medical Physics* 7.4 (1980), pp. 365–369. DOI: 10.1118/1.594693 (cit. on p. 120).

[97] Karger, Christian P. "Der Strahlentherapie-Prozess". In: *Medizinische Physik: Grundlagen – Bildgebung – Therapie – Technik*. Ed. by Wolfgang Schlegel, Christian P. Karger, and Oliver Jäkel. Berlin, Heidelberg: Springer, 2018, pp. 399–404. DOI: 10.1007/978-3-662-54801-1_24 (cit. on p. 11).

[98] Kawrakow, Iwan and Bielajew, Alex F. "On the condensed history technique for electron transport". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 142.3 (1998), pp. 253–280 (cit. on p. 24).

[99] Kieri, E., Lubich, C., and Walach, H. "Discretized dynamical low-rank approximation in the presence of small singular values". In: *SIAM J. Numer. Anal.* 54.2 (2016), pp. 1020–1038. DOI: 10.1137/15M1026791 (cit. on p. 53).

[100] Kimstrand, Peter, Traneus, Erik, Ahnesjö, Anders, Grusell, Erik, Glimelius, Bengt, and Tilly, Nina. "A beam source model for scanned proton beams". In: *Physics in Medicine and Biology* 52.11 (2007), pp. 3151–3168. DOI: 10.1088/0031-9155/52/11/015. (Visited on 05/31/2021) (cit. on p. 82).

[101] Knopf, Antje-Christin, Boye, Dirk, Lomax, Antony, and Mori, Shininchiro. "Adequate margin definition for scanned particle therapy in the incidence of intrafractional motion". In: *Physics in Medicine & Biology* 58.17 (2013), p. 6079 (cit. on p. 145).

[102] Koch, O. and Lubich, C. "Dynamical low-rank approximation". In: *SIAM J. Matrix Anal. Appl.* 29.2 (2007), pp. 434–454. DOI: 10.1137/050639703 (cit. on pp. 49, 52, 53).

[103] Koch, Othmar and Lubich, Christian. "Dynamical tensor approximation". In: *SIAM Journal on Matrix Analysis and Applications* 31.5 (2010), pp. 2360–2375 (cit. on p. 135).

[104] Kohda, Ehiichi and Shigematsu, Naoyuki. "Measurement of lung density by computed tomography: implication for radiotherapy". In: *The Keio Journal of Medicine* 38.4 (1989), pp. 454–463 (cit. on p. 45).

[105] Kollig, Thomas and Keller, Alexander. "Efficient multidimensional sampling". In: *Computer Graphics Forum*. Vol. 21. 3. Wiley Online Library. 2002, pp. 557–563 (cit. on p. 102).

[106] Kristopher Garrett, C., Hauck, Cory, and Hill, Judith. "Optimization and large scale computation of an entropy-based moment closure". In: *Journal of Computational Physics* 302 (2015), pp. 573–590 (cit. on p. 42).

[107] Kuepper, Kerstin. "Models, numerical methods, and uncertainty quantification for radiation therapy". PhD thesis. Universitätsbibliothek der RWTH Aachen, 2016 (cit. on p. 22).

[108] Kusch, Jonas. "Realizability-preserving discretization strategies for hyperbolic and kinetic equations with uncertainty". PhD thesis. Karlsruhe Institute of Technology, 2020. URL: https://doi.org/10.5445/IR/1000121168 (cit. on p. 104).

[109] Kusch, Jonas, Alldredge, Graham W, and Frank, Martin. "Maximum-principle-satisfying second-order intrusive polynomial moment scheme". In: *The SMAI journal of computational mathematics* 5 (2019), pp. 23–51 (cit. on p. 42).

[110] Kusch, Jonas, Ceruti, Gianluca, Einkemmer, Lukas, and Frank, Martin. "Dynamical low-rank approximation for Burgers' equation with uncertainty". In: *arXiv preprint arXiv:2105.04358* (2021). URL: https://arxiv.org/abs/2105.04358 (cit. on pp. 60, 135).

[111] Kusch, Jonas, Einkemmer, Lukas, and Ceruti, Gianluca. "On the stability of robust dynamical low-rank approximations for hyperbolic problems". In: *arXiv preprint arXiv:2107.07282* (2021) (cit. on pp. 49, 50, 54, 56, 58).

[112] Kusch, Jonas, Schotthöfer, Steffen, Stammer, Pia, Wolters, Jannick, and Xiao, Tianbai. "KiT-RT: An extendable framework for radiative transfer and therapy". In: *arXiv preprint arXiv:2205.08417* (2022) (cit. on pp. vii, 41, 43, 46).

[113] Kusch, Jonas and Stammer, Pia. *Numerical testcases for "A robust collision source method for rank adaptive dynamical low-rank approximation in radiation therapy"*. https://github.com/JonasKu/publication-A-robust-collision-source-method-for-rank-adaptive-dynamical-low-rankapproximation.git. 2021 (cit. on pp. vii, 62).

[114] Kusch, Jonas and Stammer, Pia. "A Robust Collision Source Method for Rank Adaptive Dynamical Low-Rank Approximation in Radiation Therapy". In: *ESAIM: Mathematical Modelling and Numerical Analysis* (2022). DOI: 10.1051/m2an/2022090 (cit. on pp. vii, 61, 62).

[115] Kusch, Jonas, Whewell, Benjamin, McClarren, Ryan, and Frank, Martin. "A low-rank power iteration scheme for neutron transport criticality problems". In: *arXiv preprint arXiv:2201.12340* (2022) (cit. on p. 162).

[116] l'Ecuyer, Pierre. "Randomized quasi-Monte Carlo: An introduction for practitioners". In: *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*. Springer. 2016, pp. 29–52 (cit. on p. 102).

[117] Laine, Samuli and Karras, Tero. "Stratified sampling for stochastic transparency". In: *Computer Graphics Forum*. Vol. 30. 4. Wiley Online Library. 2011, pp. 1197–1204 (cit. on p. 102).

[118] Larsen, E. W. and Tolar, D. R. "A "Transport" Condensed History Method". In: *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications*. Ed. by Andreas Kling, Fernando J. C. Barão, Masayuki Nakagawa, Luis Távora, and Pedro Vaz. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 49–54 (cit. on p. 24).

[119] Larsen, Edward W. "A theoretical derivation of the condensed history algorithm". In: *Annals of Nuclear Energy* 19.10-12 (1992), pp. 701–714 (cit. on p. 24).

[120] Larsen, Edward W, Miften, Moyed M, Fraass, Benedick A, and Bruinvis, Iain AD. "Electron dose calculations using the method of moments". In: *Medical physics* 24.1 (1997), pp. 111–125 (cit. on p. 21).

[121] Lathrop, Kaye D. "Ray effects in discrete ordinates equations". In: *Nuclear Science and Engineering* 32.3 (1968), pp. 357–369 (cit. on p. 38).

[122] Leong, Joseph. "Implementation of random positioning error in computerised radiation treatment planning systems as a result of fractionation". In: *Physics in Medicine & Biology* 32.3 (1987), p. 327 (cit. on p. 79).

[123] LeVeque, Randall J. *Numerical methods for conservation laws (2. ed.)*. Lectures in mathematics. Birkhäuser, 1992, pp. 1–214 (cit. on p. 40).

[124] Levermore, C. "Moment closure hierarchies for kinetic theories". In: *Journal of Statistical Physics* 83 (1996), pp. 1021–1065 (cit. on pp. 37, 38).

[125] Levermore, C. David. "Entropy-based moment closures for kinetic equations". In: *Transport Theory and Statistical Physics* 26.4-5 (1997), pp. 591–606 (cit. on pp. 37, 38).

[126] Lewis, Elmer Eugene and Miller, Warren F. "Computational methods of neutron transport". In: (1984) (cit. on p. 38).

[127] Li, P, Wang, S, Li, T, Lu, J, Huang Fu, Y, and Wang, D. "A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis". In: *The Cancer Imaging Archive* (2020). DOI: https://doi.org/10.7937/TCIA.2020.NNC2-0461 (cit. on p. 45).

[128] Lim, Gino J, Kardar, Laleh, Ebrahimi, Saba, and Cao, Wenhua. "A risk-based modeling approach for radiation therapy treatment planning under tumor shrinkage uncertainty". In: *European Journal of Operational Research* 280.1 (2020), pp. 266–278 (cit. on p. 79).

[129] Lin, Yuting, Liu, Tian, Yang, Wells, Yang, Xiaofeng, and Khan, Mohammad K. "The non-Gaussian nature of prostate motion based on real-time intrafraction tracking". In: *International Journal of Radiation Oncology* Biology* Physics* 87.2 (2013), pp. 363–369 (cit. on p. 79).

[130] Liu, Wei, Zhang, Xiaodong, Li, Yupeng, and Mohan, Radhe. "Robust Optimization of Intensity Modulated Proton Therapy". In: *Medical Physics* 39.2 (2012), pp. 1079–1091. DOI: 10.1118/1.3679340 (cit. on pp. 20, 31, 78, 118, 142).

[131] Lomax, A. "Intensity Modulation Methods for Proton Radiotherapy". In: *Physics in Medicine and Biology* 44.1 (1999), pp. 185–205. DOI: 10.1088/0031-9155/44/1/014 (cit. on p. 120).

[132] Lomax, A. J. "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties." In: *Physics in medicine and biology* 53.4 (2008), pp. 1027–1042. DOI: 10.1088/0031-9155/53/4/014 (cit. on pp. v, 20, 31, 78, 79, 81, 82).

[133] Lomax, A. J. "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions." In: *Physics in medicine and biology* 53.4 (2008), pp. 1043–1056. DOI: 10.1088/0031-9155/53/4/015 (cit. on pp. v, 78).

[134] Longoni, Gianluca. "Advanced quadrature sets and acceleration and preconditioning techniques for the discrete ordinates method in parallel computing environments". PhD thesis. University of Florida, 2004 (cit. on p. 38).

[135] Low, D. A., Harms, W. B., Mutic, S., and Purdy, J. A. "A Technique for the Quantitative Evaluation of Dose Distributions". In: *Medical Physics* 25.5 (1998), pp. 656–661. DOI: 10.1118/1.598248 (cit. on p. 17).

[136] Lowe, Matthew, Albertini, Francesca, Aitkenhead, Adam, Lomax, Antony John, and MacKay, Ranald I. "Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors". In: *Physics in Medicine and Biology* 61.1 (2016). Publisher: IOP Publishing, pp. 413–429. DOI: 10.1088/0031-9155/61/1/413. (Visited on 09/23/2016) (cit. on p. 79).

[137] Lubich, C. and Oseledets, I. V. "A projector-splitting integrator for dynamical low-rank approximation". In: *BIT* 54.1 (2014), pp. 171–188. DOI: 10.1007/s10543-013-0454-0 (cit. on pp. 49, 53).

[138] Lubich, C., Vandereycken, B., and Walach, H. "Time integration of rank-constrained Tucker tensors". In: *SIAM J. Numer. Anal.* 56.3 (2018), pp. 1273–1290. DOI: 10.1137/17M1146889 (cit. on p. 137).

[139] Lubich, Christian, Rohwedder, Thorsten, Schneider, Reinhold, and Vandereycken, Bart. "Dynamical approximation by hierarchical Tucker and tensor-train tensors". In: *SIAM Journal on Matrix Analysis and Applications* 34.2 (2013), pp. 470–494 (cit. on pp. 135, 140, 162).

[140] Lynch, Gerald R and Dahl, Orin I. "Approximations to multiple Coulomb scattering". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 58.1 (1991), pp. 6–10 (cit. on p. 27).

[141] Marchuk, G and Lebedev, V I. "Numerical methods in the theory of neutron transport". In: (1986). URL: https://www.osti.gov/biblio/7057084 (cit. on p. 38).

[142] Mathews, Kirk A. "On the propagation of rays in discrete ordinates". In: *Nuclear science and engineering* 132.2 (1999), pp. 155–180 (cit. on p. 38).

[143] Matoušek, Jiřié. "On the L2-discrepancy for anchored boxes". In: *Journal of Complexity* 14.4 (1998), pp. 527–556 (cit. on p. 102).

[144] McClarren, Ryan G and Hauck, Cory D. "Robust and accurate filtered spherical harmonics expansions for radiative transfer". In: *Journal of Computational Physics* 229.16 (2010), pp. 5597–5614 (cit. on p. 37).

[145] McGowan, S E, Burnet, N G, and Lomax, A J. "Treatment planning optimisation in proton therapy". In: *The British Journal of Radiology* 86.1021 (2013). Publisher: The British Institute of Radiology, pp. 20120288–20120288. DOI: 10.1259/bjr.20120288. (Visited on 05/27/2021) (cit. on p. 79).

[146] Medaglia, Andrea, Colelli, Giulia, Farina, Lisa, Bacila, Ana, Bini, Paola, Marchioni, Enrico, Figini, Silvia, Pichiecchio, Anna, and Zanella, Mattia. "Uncertainty quantification and control of kinetic models of tumour growth under clinical uncertainties". In: *International Journal of Non-Linear Mechanics* 141 (2022), p. 103933 (cit. on p. 79).

[147] Metropolis, Nicholas and Ulam, Stanislaw. "The monte carlo method". In: *Journal of the American statistical association* 44.247 (1949), pp. 335–341 (cit. on p. 32).

[148] Mohan, Radhe and Grosshans, David. "Proton therapy–present and future". In: *Advanced drug delivery reviews* 109 (2017), pp. 26–44 (cit. on p. 26).

[149] Molière, Gert. "Theorie der Streuung schneller geladener Teilchen II Mehrfach-und Vielfachstreuung". In: *Zeitschrift für Naturforschung A* 3.2 (1948), pp. 78–97. DOI: doi:10.1515/zna-1948-0203 (cit. on p. 27).

[150] Morel, JE, Wareing, TA, Lowrie, RB, and Parsons, DK. "Analysis of ray-effect mitigation techniques". In: *Nuclear science and engineering* 144.1 (2003), pp. 1–22 (cit. on p. 38).

[151] Musharbash, Eleonora and Nobile, Fabio. "Dual dynamically orthogonal approximation of incompressible Navier Stokes equations with random boundary conditions". In: *Journal of Computational Physics* 354 (2018), pp. 135–162 (cit. on p. 60).

[152] Newhauser, Wayne D and Zhang, Rui. "The physics of proton therapy". In: *Physics in Medicine & Biology* 60.8 (2015), R155 (cit. on pp. 21, 24, 28, 29, 82).

[153] Nguyen, Thi Phuoc Van. *Normal/high/low breathing rate measured by Continuous Wave Radar sensor*. 2018. DOI: 10.21227/q833-jb44 (cit. on p. 88).

[154] Nguyen-Bui, Thanh-Ha, Larouze, Alexandre, Bourgeois, Rémi, and Dubroca, Bruno. "Multi-Dimensional Numerical Modelling of Heavy Ion Transport in Matter Using a Fast and Accurate Method". 2021 (cit. on pp. 27, 28, 38).

[155] Niederreiter, Harald. "Methods for estimating discrepancy". In: *Applications of number theory to numerical analysis*. Elsevier, 1972, pp. 203–236 (cit. on p. 102).

[156] Njeh, CF et al. "Tumor delineation: The weakest link in the search for accuracy in radiotherapy". In: *Journal of medical physics* 33.4 (2008), p. 136 (cit. on pp. v, 79).

[157] Nocedal, Jorge and Wright, Stephen J. *Numerical optimization*. Springer, 1999 (cit. on p. 14).

[158] Nonnenmacher, Achim and Lubich, Christian. "Dynamical low-rank approximation: applications and numerical experiments". In: *Mathematics and Computers in Simulation* 79.4 (2008), pp. 1346–1357 (cit. on pp. 135, 136).

[159] Oelfke, U and Bortfeld, T. "Intensity modulated radiotherapy with charged particle beams: studies of inverse treatment planning for rotation therapy". In: *Medical Physics* 27.6 (2000), pp. 1246–1257 (cit. on p. 12).

[160] Olbrant, Edgar. "Models and numerical methods for time-and energy-dependent particle transport". 2012 (cit. on pp. 24, 27–29).

[161] Olbrant, Edgar and Frank, Martin. "Generalized Fokker–Planck theory for electron and photon transport in biological tissues: application to radiotherapy". In: *Computational and mathematical methods in medicine* 11.4 (2010), pp. 313–339 (cit. on pp. 22, 27).

[162] Ostermann, Alexander, Piazzola, Chiara, and Walach, Hanna. "Convergence of a low-rank Lie–Trotter splitting for stiff matrix differential equations". In: *SIAM Journal on Numerical Analysis* 57.4 (2019), pp. 1947–1966 (cit. on p. 50).

[163] Owen, Art B. "Randomly permuted (t, m, s)-nets and (t, s)-sequences". In: *Monte Carlo and quasi-Monte Carlo methods in scientific computing*. Springer, 1995, pp. 299–317 (cit. on p. 102).

[164] Owen, Art B. "Monte Carlo variance of scrambled net quadrature". In: *SIAM Journal on Numerical Analysis* 34.5 (1997), pp. 1884–1910 (cit. on p. 102).

[165] Owen, Art B. *Monte Carlo Theory, Methods and Examples*. 2013 (cit. on pp. 115–117).

[166] Paganetti, Harald. "Range uncertainties in proton therapy and the role of Monte Carlo simulations". In: *Physics in Medicine and Biology* 57.11 (2012). Publisher: IOP Publishing ISBN: 0031-9155, R99–R117. DOI: 10.1088/0031-9155/57/11/R99. (Visited on 10/13/2016) (cit. on p. 79).

[167] Papanicolaou, Alex. "Taylor approximation and the delta method". In: 28 (2009) (cit. on p. 82).

[168] Papanikolaou, Nikos, Battista, Jerry J, Boyer, Arthur L, Kappas, Constantin, Klein, Eric, Mackie, T Rock, Sharpe, Michael, and Van Dyk, Jake. "Tissue inhomogeneity corrections for megavoltage photon beams". In: *AAPM Task Group* 65 (2004), pp. 1–142 (cit. on p. iv).

[169] Park, Peter C. et al. "Statistical Assessment of Proton Treatment Plans under Setup and Range Uncertainties." In: *International journal of radiation oncology, biology, physics* 86.5 (2013), pp. 1007–1013. DOI: 10.1016/j.ijrobp.2013.04.009 (cit. on p. 78).

[170] Pedroni, E, Scheib, S, Böhringer, T, Coray, A, Grossmann, M, Lin, S, and Lomax, A. "Experimental characterization and physical modelling of the dose distribution of scanned proton pencil beams". In: *Physics in Medicine and Biology* 50.3 (2005), pp. 541–561. DOI: 10.1088/0031-9155/50/3/011. (Visited on 05/31/2021) (cit. on p. 82).

[171] Peng, Zhuogang and McClarren, Ryan G. "A high-order/low-order (HOLO) algorithm for preserving conservation in time-dependent low-rank transport calculations". In: *Journal of Computational Physics* 447 (2021), p. 110672 (cit. on p. 63).

[172] Peng, Zhuogang, McClarren, Ryan G, and Frank, Martin. "A low-rank method for two-dimensional time-dependent radiation transport calculations". In: *J. Comput. Phys.* 421 (2020), p. 109735 (cit. on p. 63).

[173] Perkó, Zoltán, Gilli, Luca, Lathouwers, Danny, and Kloosterman, Jan Leen. "Grid and basis adaptive polynomial chaos techniques for sensitivity and uncertainty analysis". In: *Journal of Computational Physics* 260 (2014), pp. 54–84 (cit. on p. 105).

[174] Perkó, Zoltán, Voort, Sebastian R. van der, Water, Steven van de, Hartman, Charlotte M. H., Hoogeman, Mischa, and Lathouwers, Danny. "Fast and Accurate Sensitivity Analysis of IMPT Treatment Plans Using Polynomial Chaos Expansion". In: *Physics in Medicine and Biology* 61.12 (2016), pp. 4646–4664. DOI: 10.1088/0031-9155/61/12/4646 (cit. on pp. 35, 78, 81, 107, 120, 132, 147).

[175] Perl, J., Shin, J., Schumann, J., Faddegon, B., and Paganetti, H. "TOPAS: An Innovative Proton Monte Carlo Platform for Research and Clinical Applications". In: *Medical Physics* 39.11 (2012), pp. 6818–6837. DOI: 10.1118/1.4758060 (cit. on pp. 23, 26, 30, 43, 72, 108, 119).

[176] Peshkin, Leonid, Bld, Dworkin, Shelton, Christian R, and Bld, Gates. "Christian R. Shelton: Learning from Scarce Experience". In: (1992) (cit. on p. 110).

[177] Petti, Paula L. "Differential-pencil-beam dose calculations for charged particles". In: *Medical physics* 19.1 (1992), pp. 137–149 (cit. on p. 35).

[178] Pflugfelder, D., Wilkens, J. J., and Oelfke, U. "Worst Case Optimization: A Method to Account for Uncertainties in the Optimization of Intensity Modulated Proton Therapy." In: *Physics in medicine and biology* 53.6 (2008), pp. 1689–1700. DOI: 10.1088/0031-9155/53/6/013 (cit. on pp. 83, 118, 142).

[179] Poëtte, Gaël. "A gPC-intrusive Monte-Carlo Scheme for the Resolution of the Uncertain Linear Boltzmann Equation". In: *Journal of Computational Physics* (2018) (cit. on p. 35).

[180] Quinonero-Candela, Joaquin and Rasmussen, Carl Edward. "A unifying view of sparse approximate Gaussian process regression". In: *The Journal of Machine Learning Research* 6 (2005), pp. 1939–1959 (cit. on p. 88).

[181] Radig, Jean. "Modelisation of Uncertainties for Treatment Planning". MA thesis. ETH Zürich, 2022 (cit. on pp. viii, 88).

[182] Rasmussen, Carl Edward. "Gaussian processes in machine learning". In: *Summer school on machine learning*. Springer. 2003, pp. 63–71 (cit. on p. 88).

[183] Roberts, Stephen, Osborne, Michael, Ebden, Mark, Reece, Steven, Gibson, Neale, and Aigrain, Suzanne. "Gaussian processes for time-series modelling". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1984 (2013), p. 20110550 (cit. on pp. 88, 89).

[184] Rutherford, E. "The scattering of α and β particles by matter and the structure of the atom". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 21.125 (1911), pp. 669–688. DOI: 10.1080/14786440508637080 (cit. on p. 26).

[185] Salvat, Francesc, Fernández-Varea, José M, Sempau, Josep, et al. "PENELOPE-2008: A code system for Monte Carlo simulation of electron and photon transport". In: *Workshop Proceedings, Barcelona, Spain*. Vol. 30. 2008 (cit. on p. 24).

[186] Schaffner, Barbara, Pedroni, Eros, and Lomax, Antony. "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". In: *Physics in Medicine & Biology* 44.1 (1999), p. 27 (cit. on p. 36).

[187] Schattschneider, Peter. *Fundamentals of inelastic electron scattering*. Springer Science & Business Media, 2012 (cit. on p. 24).

[188] Schneider, Uwe, Pedroni, Eros, and Lomax, Antony. "The calibration of CT Hounsfield units for radiotherapy treatment planning". In: *Physics in Medicine & Biology* 41.1 (1996), p. 111. DOI: 10.1088/0031-9155/41/1/009 (cit. on pp. 11, 65, 79).

[189] Schneider, Wilfried, Bortfeld, Thomas, and Schlegel, Wolfgang. "Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions". In: *Physics in Medicine & Biology* 45.2 (2000), p. 459. DOI: 10.1088/0031-9155/45/2/314 (cit. on pp. 11, 65, 79).

[190] Schotthöfer, Steffen, Xiao, Tianbai, Frank, Martin, and Hauck, Cory. "Structure Preserving Neural Networks: A Case Study in the Entropy Closure of the Boltzmann Equation". In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato. Vol. 162. Proceedings of Machine Learning Research. PMLR, 2022, pp. 19406–19433. URL: https://proceedings.mlr.press/v162/schotthofer22a.html (cit. on p. 38).

[191] Scott, David W. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015 (cit. on p. 108).

[192] Scott, William T. "The theory of small-angle multiple scattering of fast charged particles". In: *Reviews of modern physics* 35.2 (1963), p. 231 (cit. on p. 27).

[193] Seco, Joao, Robertson, Daniel, Trofimov, Alexei, and Paganetti, Harald. "Breathing interplay effects during proton beam scanning: simulation and statistical analysis". In: *Physics in Medicine & Biology* 54.14 (2009), N283 (cit. on p. 84).

[194] Seco, Joao and Verhaegen, Frank. *Monte Carlo techniques in radiation therapy*. CRC press Boca Raton, FL: 2013 (cit. on p. v).

[195] Seibold, Benjamin and Frank, Martin. "StaRMAP—A Second Order Staggered Grid Method for Spherical Harmonics Moment Equations of Radiative Transfer". In: *ACM Trans. Math. Softw.* 41.1 (2014). DOI: 10.1145/2590808 (cit. on p. 43).

[196] Shirakawa, Shinichi, Akimoto, Youhei, Ouchi, Kazuki, and Ohara, Kouzou. “Sample Reuse via Importance Sampling in Information Geometric Optimization”. In: (2018), pp. 1–24 (cit. on p. 110).

[197] Siddon, Robert L. “Fast calculation of the exact radiological path for a three-dimensional CT array”. In: *Medical physics* 12.2 (1985), pp. 252–255 (cit. on p. 36).

[198] Sobol’, Il’ya Meerovich. “On the distribution of points in a cube and the approximate evaluation of integrals”. In: *Zhurnal Vychislitel’noi Matematiki i Matematicheskoi Fiziki* 7.4 (1967), pp. 784–802 (cit. on p. 102).

[199] Sobotta, B, Söhn, M, and Alber, M. “Robust optimization based upon statistical theory”. In: *Medical physics* 37.8 (2010), pp. 4019–4028 (cit. on p. 146).

[200] Soukup, Martin, Fippel, Matthias, and Alber, Markus. “A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations”. In: *Physics in Medicine & Biology* 50.21 (2005), p. 5089 (cit. on pp. 36, 108).

[201] Soukup, Martin, Fippel, Matthias, and Alber, Markus. “A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations”. In: *Physics in Medicine and Biology* 50.21 (2005), pp. 5089–5104. DOI: 10.1088/0031-9155/50/21/010. (Visited on 05/31/2021) (cit. on pp. 36, 82).

[202] Souris, Kevin, Montero, Ana Barragan, Janssens, Guillaume, Perri, Dario Di, Sterpin, Edmond, and Lee, John A. “Technical Note: Monte Carlo methods to comprehensively evaluate the robustness of 4D treatments in proton therapy”. In: *Medical Physics* 46.10 (2019), pp. 4676–4684. DOI: https://doi.org/10.1002/mp.13749. (Visited on 05/27/2021) (cit. on p. 84).

[203] Stammer, P., Burigo, L., Jäkel, O., Frank, M., and Wahl, N. “Efficient Uncertainty Quantification for Monte Carlo Dose Calculations Using Importance (Re-)Weighting”. In: 66.20 (2021), p. 205003. DOI: 10.1088/1361-6560/ac287f (cit. on pp. vii, viii, 110).

[204] Stammer, P., Burigo, L., Jäkel, O., Frank, M., and Wahl, N. “PO-1728 Efficient modeling and quantification of time-dependent errors in IMPT”. In: *Radiotherapy and Oncology* 170 (2022), S1529–S1531 (cit. on pp. viii, 85).

[205] Stammer, Pia, Burigo, Lucas, Jäkel, Oliver, Frank, Martin, and Wahl, Niklas. “O044 - Efficient uncertainty estimates in Monte Carlo dose calculation using importance re-weighting”. In: *International Journal of Particle Therapy, Proceedings to the 59th Annual Conference of the Particle Therapy Cooperative Group (PTCOG59 2021 Online)* (2022), p. 24. DOI: 10.14338/IJPT-22-PTCOG59-9.3 (cit. on p. vii).

[206] Stammer, Pia, Burigo, Lucas, Jäkel, Oliver, Frank, Martin, and Wahl, Niklas. “Multivariate error modeling and uncertainty quantification using importance (re-)weighting for Monte Carlo simulations in particle transport”. In: *Journal of Computational Physics* 473 (2023), p. 111725. DOI: https://doi.org/10.1016/j.jcp.2022.111725 (cit. on pp. vii, viii, 33, 110).

[207] Stroud, A.H. and Secrest, D. *Gaussian Quadrature Formulas*. Prentice-Hall series in automatic computation. Prentice-Hall, 1966. URL: https://books.google.de/books?id=0bI-AAAAIAAJ (cit. on p. 103).

[208] Supanitsky, A. D. and Medina-Tanco, G. “Effect of Multiple Reusing of Simulated Air Showers in Detector Simulations”. In: *Astroparticle Physics* 30.5 (2008), pp. 264–269. DOI: 10.1016/j.astropartphys.2008.10.001 (cit. on p. 110).

[209] Szymanowski, Hanitra and Oelfke, Uwe. "Two-dimensional pencil beam scaling: an improved proton dose algorithm for heterogeneous media". In: *Physics in Medicine & Biology* 47.18 (2002), p. 3313 (cit. on p. 36).

[210] Taussky, Olga, Taussky, Olga, and Todd, John. *Generation and testing of pseudo-random numbers*. US Department of Commerce, National Institute of Standards and Technology, 1954 (cit. on p. 101).

[211] Teng, Tong, Chen, Jie, Zhang, Yehong, and Low, Bryan Kian Hsiang. "Scalable variational Bayesian kernel selection for sparse Gaussian process regression". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 04. 2020, pp. 5997–6004 (cit. on p. 88).

[212] Tourovsky, A, Lomax, A J, Schneider, U, and Pedroni, E. "Monte Carlo dose calculations for spot scanned proton therapy". In: *Physics in Medicine and Biology* 50.5 (2005), pp. 971–981. DOI: 10.1088/0031-9155/50/5/019. (Visited on 05/31/2021) (cit. on p. 82).

[213] Trefethen, Lloyd N. "Cubature, Approximation, and Isotropy in the Hypercube". In: *SIAM Rev.* 59.3 (2017), pp. 469–491. DOI: 10.1137/16M1066312 (cit. on p. 104).

[214] Tukey, John W. "Configural Polysampling". In: *Siam Review* 29.1 (1987), pp. 1–20 (cit. on p. 110).

[215] Uilkema, Sander. "Proton therapy planning using the SN method with the Fokker–Planck approximation". In: *Delft University of Technology* (2012) (cit. on pp. 27–29, 161).

[216] Ulmer, W. and Matsinos, Evangelos. "Theoretical methods for the calculation of Bragg curves and 3D distributions of proton beams". In: *The European Physical Journal Special Topics* 190 (2011), pp. 1–81. DOI: 10.1140/epjst/e2010-01335-7 (cit. on p. 82).

[217] Unkelbach, Jan, Chan, Timothy CY, and Bortfeld, Thomas. "Accounting for range uncertainties in the optimization of intensity modulated proton therapy". In: *Physics in Medicine & Biology* 52.10 (2007). Publisher: IOP Publishing, p. 2755 (cit. on pp. v, 35, 79, 81, 107).

[218] Van Herk, Marcel. "Errors and margins in radiotherapy". In: *Seminars in radiation oncology*. Vol. 14. 1. Elsevier. 2004, pp. 52–64 (cit. on pp. 79, 145).

[219] Vassiliev, Oleg N. "Introduction". In: *Monte Carlo Methods for Radiation Transport: Fundamentals and Advanced Topics*. Ed. by Oleg N. Vassiliev. Biological and Medical Physics, Biomedical Engineering. Cham: Springer International Publishing, 2017, pp. 1–14. DOI: 10.1007/978-3-319-44141-2_1 (cit. on pp. iv, v).

[220] Vassiliev, Oleg N. "Particle Trajectories, Tallies, and Variance Reduction". In: *Monte Carlo Methods for Radiation Transport: Fundamentals and Advanced Topics*. Ed. by Oleg N. Vassiliev. Biological and Medical Physics, Biomedical Engineering. Cham: Springer International Publishing, 2017, pp. 105–139. DOI: 10.1007/978-3-319-44141-2_4 (cit. on p. 35).

[221] Vassiliev, Oleg N. "The Boltzmann Equation". In: *Monte Carlo Methods for Radiation Transport: Fundamentals and Advanced Topics*. Ed. by Oleg N. Vassiliev. Biological and Medical Physics, Biomedical Engineering. Cham: Springer International Publishing, 2017, pp. 49–104. DOI: 10.1007/978-3-319-44141-2_3 (cit. on pp. 34, 35).

[222] Vidlicková, Eva. "Dynamical low rank approximation for uncertainty quantification of time-dependent problems". PhD thesis. Lausanne: MATHICSE, 2022, p. 222. DOI: 10.5075/epfl-thesis-9087 (cit. on p. 135).

[223] Von Neumann, John. "13. various techniques used in connection with random digits". In: *Appl. Math Ser* 12.36-38 (1951), p. 3 (cit. on p. 32).

[224] Wächter, Andreas and Biegler, Lorenz T. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". In: *Mathematical programming* 106.1 (2006), pp. 25–57 (cit. on pp. 14, 150).

[225] Wagner, John C, Peplow, Douglas E, Mosher, Scott W, Evans, Thomas M, et al. "Review of hybrid (deterministic/Monte Carlo) radiation transport methods, codes, and applications at Oak Ridge National Laboratory". In: *Progress in nuclear science and technology* 2 (2011), pp. 808–814 (cit. on p. v).

[226] Wahl, Niklas, Hennig, Philipp, Wieser, Hans-Peter, and Bangert, Mark. "Efficiency of Analytical and Sampling-Based Uncertainty Propagation in Intensity-Modulated Proton Therapy". In: *Physics in Medicine and Biology* 62.14 (2017), pp. 5790–5807. DOI: 10.1088/1361-6560/aa6ec5 (cit. on pp. 35, 81, 83, 107).

[227] Wahl, Niklas, Hennig, Philipp, Wieser, Hans-Peter, and Bangert, Mark. "Analytical incorporation of fractionation effects in probabilistic treatment planning for intensity-modulated proton therapy". In: *Medical physics* 45.4 (2018), pp. 1317–1328 (cit. on p. 147).

[228] Walach, Hanna Maria. "Time integration for the dynamical low-rank approximation of matrices and tensors". PhD thesis. Universitätsbibliothek Tübingen, 2019 (cit. on p. 137).

[229] Webb, Steve. *Intensity-modulated radiation therapy*. CRC Press, 2015 (cit. on p. 12).

[230] Whitlock, Paula A and Kalos, MH. *Monte Carlo Methods*. Wiley, 1986 (cit. on pp. 33, 34).

[231] Wiener, Norbert. "The homogeneous chaos". In: *American Journal of Mathematics* 60.4 (1938), pp. 897–936 (cit. on p. 98).

[232] Wieser, Hans-Peter, Hennig, Philipp, Wahl, Niklas, and Bangert, Mark. "Analytical probabilistic modeling of RBE-weighted dose for ion therapy". In: *Physics in Medicine & Biology* 62.23 (2017), p. 8959 (cit. on p. 107).

[233] Wieser, Hans-Peter, Karger, Christian P., Wahl, Niklas, and Bangert, Mark. "Impact of Gaussian Uncertainty Assumptions on Probabilistic Optimization in Particle Therapy". In: *Physics in Medicine & Biology* 65.14 (2020), p. 145007. DOI: 10.1088/1361-6560/ab8d77 (cit. on pp. 35, 81, 107).

[234] Wieser, Hans-Peter et al. "Development of the Open-Source Dose Calculation and Optimization Toolkit matRad." In: *Medical physics* 44.6 (2017), pp. 2556–2568. DOI: 10.1002/mp.12251 (cit. on pp. 14, 18, 119, 146, 150).

[235] Woo, M. K. and Cunningham, J. R. "The validity of the density scaling method in primary electron transport for photon and electron beams". In: *Medical Physics* 17.2 (1990), pp. 187–194. DOI: https://doi.org/10.1118/1.596497 (cit. on p. 22).

[236] Wu, Qiuwen and Mohan, Radhe. "Algorithms and functionality of an intensity modulated radiotherapy optimization system". In: *Medical physics* 27.4 (2000), pp. 701–711 (cit. on pp. 12–14).

[237] Xiu, Dongbin and Karniadakis, George Em. "The Wiener–Askey polynomial chaos for stochastic differential equations". In: *SIAM journal on scientific computing* 24.2 (2002), pp. 619–644 (cit. on pp. 99, 100).

[238] Yang, Ming, Zhu, X. Ronald, Park, Peter C., Titt, Uwe, Mohan, Radhe, Virshup, Gary, Clayton, James E., and Dong, Lei. "Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration". In: *Physics in Medicine and Biology* 57.13 (2012), pp. 4095–4115. DOI: 10.1088/0031-9155/57/13/4095 (cit. on p. 82).

[239] Yang, Mingru and White, Steven R. "Time-dependent variational principle with ancillary Krylov subspace". In: *Phys. Rev. B* 102.9 (2020), p. 094315 (cit. on p. 50).