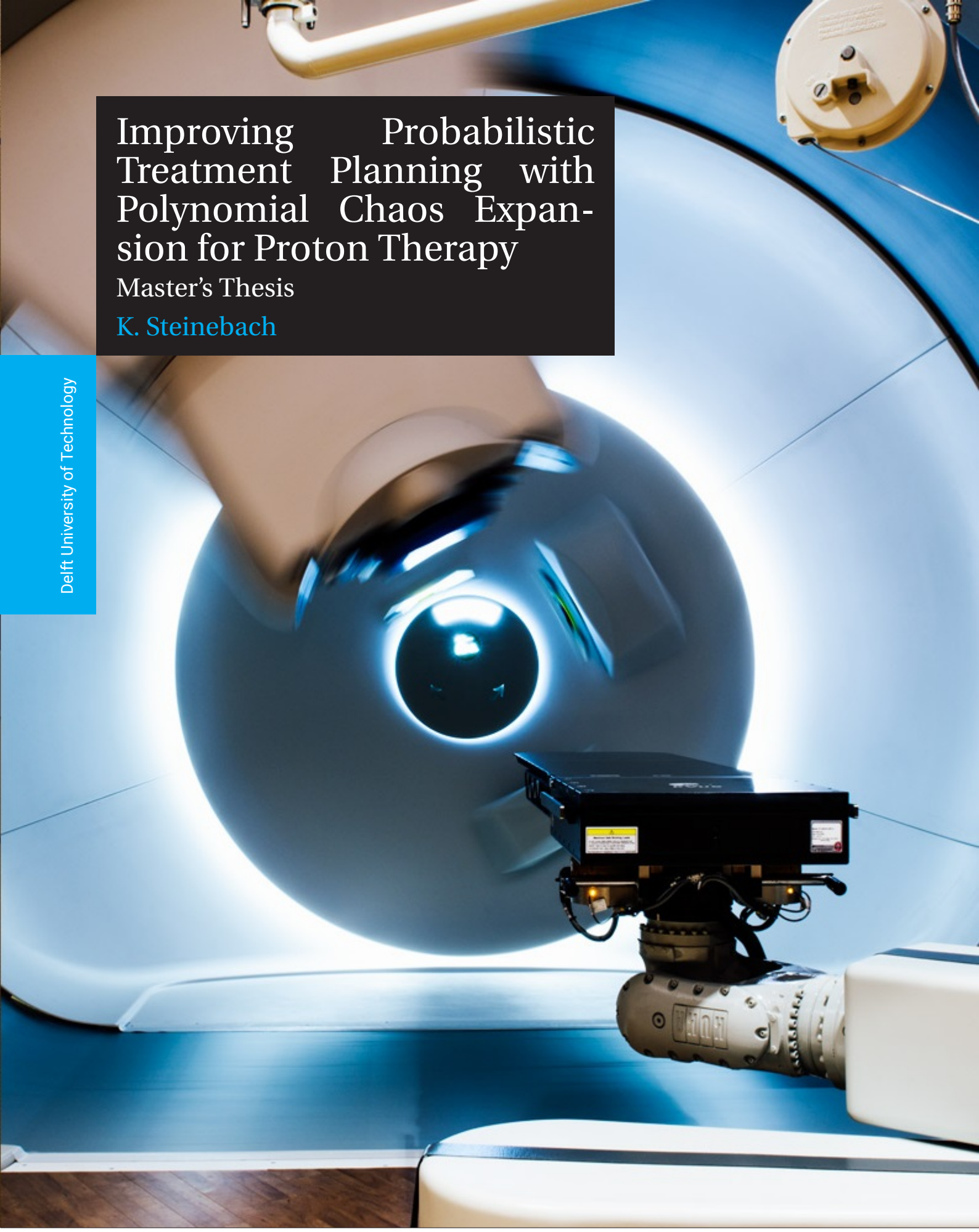


Improving Probabilistic Treatment Planning with Polynomial Chaos Expansion for Proton Therapy

Master's Thesis

K. Steinebach

Delft University of Technology

TU Delft Delft University of Technology

Challenge the future

Improving Probabilistic Treatment Planning with Polynomial Chaos Expansion for Proton Therapy

Master's Thesis

by

K. Steinebach

To obtain the degree of Master of Science in Applied Mathematics at the faculty of Electrical Engineering, Mathematics & Computer Science of Delft University of Technology, to be defended publicly on Friday November 6, 2020 at 9:00 AM CET.

Master of Science
in Applied Mathematics

Under supervision of the Numerical Analysis group
Delft Institute of Applied Mathematics
Faculty of Electrical Engineering, Mathematics & Computer Science
Delft University of Technology

And under supervision of the Medical Physics & Technology group
Department of Radiation Science & Technology
Faculty of Applied Sciences
Delft University of Technology

Student Number:	4372549	
Project Duration:	January 6, 2020 – November 6, 2020	
Thesis Committee:	Dr. Ir. M.B. van Gijzen	TU Delft (EEMCS; Chair & Supervisor)
	Dr. Z. Perkó	TU Delft (AS; Supervisor)
	Dr. Ir. D. Lathouwers	TU Delft (AS; Supervisor)
	Prof. Dr. Ir. A.W. Heemink	TU Delft (EEMCS)

An electronic version of this thesis is available at: `http://repository.tudelft.nl/`.

TU Delft Delft University of Technology

ABSTRACT

In proton therapy, robust treatment planning is currently used to account for uncertainties in patient alignment and proton beam range. A way to overcome the limitations of robust treatment planning is to use probabilistic treatment planning, which can be computationally expensive due to the calculations of statistical measures of stochastic dose parameters. For this purpose, the method of Polynomial Chaos Expansion (PCE) has been introduced to alleviate computational costs. PCEs are helpful to evaluate the statistical measures analytically, or via sampling, and can usually be evaluated faster than through calculations with the dose engine. In this research, the work of Salverda et al. (2019) is extended upon improving probabilistic treatment planning with Polynomial Chaos Expansion for proton therapy. Probabilistic treatment plan optimizations with patient data in ErasmusMC's iCycle were found to be too costly for clinical use with computation times in the order of 1-3 weeks. The goal of this research is to alleviate the computational cost of probabilistic treatment planning, in which percentiles of stochastic dose volume parameters are used, with the aim of a proof of principle for clinical use.

New methodologies on the calculation of the value, gradient and Hessian of percentiles of dose volume parameters, and treatment plan optimizations with these methods, are tested on a simplified 3D geometry with a tumor and an organ in Matlab. The optimizations are performed with the *fmincon* solver, using the default interior-point algorithm. The construction of PCEs is performed with the OpenGPC package, developed by Z. Perkó et al. (2016).

Three exact improvements on the calculation of the value, gradient and Hessian of percentiles of dose volume parameters have been implemented and show an overall reduction in computation time of 30%. These improvements yield a similar speed-up in the total optimization time of a probabilistic treatment plan, depending on the dose parameters in the objectives and constraints, and their computation times. Furthermore, the accuracy of the gradient and Hessian of percentiles of dose volume parameters has been improved through the introduction of a monotonicity constraint on the PCE coefficients. To reduce the computation time even further, an approximation method is devised for the calculation of the gradient and Hessian of percentiles of dose volume parameters. This method shows a decrease in computation time from 5,600 seconds to only 220 seconds per calculation of the value, gradient and Hessian of a percentile of a dose volume parameter for the 3D geometry and a sample size of 500,000. The problem is that the approximation method is quite heuristic and its accuracy cannot be predicted in general. However, relatively accurate percentile approximations are produced with small absolute errors. Optimizations of probabilistic treatment plans are performed with the exact and approximate method, yielding small differences in terms of dose distributions due to differences in sample size. The cases with the approximation method yield the fastest optimization times and are shown to be 7.6 times faster, from 5.3 days to 16.5 hours, than the case with the exact method and a sample size of 500,000. The cases with the exact method show to be 2.5 and 1.2 times faster for a sample size of 50,000 and 250,000 than for a sample size of 500,000 respectively. The higher number of iterations in cases with the approximation method and cases with the exact method for smaller sample sizes outweigh the rapid increase in computation time for higher sample sizes with the exact method.

A decrease in computation time is shown for probabilistic treatment planning, in which percentiles of dose volume parameters are used, in general, and also in implementations in the 3D geometry with and without the approximation method. In future research, the decrease in computation time with the exact improvements should be investigated in optimizations with patient data in iCycle. Also, the approximation method should be tested on different geometries and for multiple uncertain parameters in the geometry to investigate whether this method is robust enough to be tested in optimizations with patient data in iCycle. As a conclusion, the new methodologies in this research do improve probabilistic treatment planning with Polynomial Chaos Expansion for proton therapy by alleviating computational cost, but a proof of principle for clinical use is not yet achieved.

ACKNOWLEDGEMENTS

This thesis has been written for the partial fulfilment of the Master of Science degree in Applied Mathematics at Delft University of Technology. First of all, I would like to thank my supervisors Dr. Zoltán Perkó and Dr. ir. Danny Lathouwers for their help and valuable feedback during the project and in our many progress meetings, and for giving me the possibility to perform my master's thesis research project at the Medical Physics & Technology group in such an interesting research field. I want to speak out just as much gratitude towards Dr. ir. Martin van Gijzen for his guidance and advice as supervisor from the Numerical Analysis group of the Mathematics faculty. Even though Martin is not an expert in proton therapy research, he was always able to help me to put parts of the research into perspective and to keep the overview in such a lengthy project. At the same time, our one-on-one meetings were a great learning experience in explaining the theory to someone who is less familiar with the topics. Furthermore, I would like to thank Prof. dr. ir. Arnold Heemink for taking part in my thesis committee as independent member, besides Zoltán, Danny and Martin.

As we all know, the COVID-19 pandemic has struck the world, and so also the Netherlands, hard this year. In many ways it affected our society, resulting in a great distortion of our everyday work/study/social lifes. Luckily for me, these study changes were relatively good to handle in this thesis. The first three months of the project, from January until March, I was able to work at the Reactor Institute Delft, and shortly at the ErasmusMC. The months April until June everything had to be done from home, but from July onwards there have been study places available at the faculty. All software was available through my own laptop, so I was still able to perform my 'experiments'. The weekly progress meetings were held through Skype, which worked out fine.

During the project, I gained a lot of knowledge about cancer treatment in general, radiotherapy, in particular proton therapy, the possibilities of the method of Polynomial Chaos Expansion and different concepts of treatment planning. Besides the gained knowledge, I experienced quite some personal developments in improving with project planning, in being able to cope with different setbacks, in presenting results, in keeping myself motivated and in managing my expectations during a lengthy project of autonomous research. I am sure that these experiences will help me in my future career.

This master's thesis marks the end of my time as a student. I am grateful for the wonderful moments with all my friends here in Delft. A special thank you to my girlfriend Sunny for motivating me through hard times. Last, but definitely not least, I would like to say a big thank you to my parents Paul & Ellen for the unconditional support they have given me throughout the past six years of studying at Delft University of Technology.

K. Steinebach
Delft, November 2020

CONTENTS

NOMENCLATURE

SYMBOLS

Γ	Set of Polynomial Chaos basis vectors
$\delta_{k,l}$	Kronecker delta
Δ_{lev}	Increased quadrature level for the single dimensional directions of a Smolyak sparse grid
μ_R	Expected value of stochastic response R
$\boldsymbol{\xi}$	Uncertainty vector
$\boldsymbol{\xi}^s$	Realization of uncertainty vector in error scenario s
Σ	Standard deviation of systematic set-up error
σ	Standard deviation of random set-up error
σ_R	Standard deviation of stochastic response R
ϕ	Polynomial Chaos basis function
Ψ	Polynomial Chaos basis vector
$\mathcal{B}$	Index set of beamspots
D	Dose deposition matrix
$\boldsymbol{d}$	Dose vector
$\boldsymbol{d}^P$	Prescribed dose vector
$d_{\alpha\%}$	Dose volume parameter
$d_{\alpha\%}^{\beta\%}$	Probabilistic dose volume parameter
$\hat{\boldsymbol{e}}_n$	Unit vector along the n-th dimension
$f(\cdot)$	Arbitrary objective function
$g(\cdot)$	Arbitrary stochastic variable
$\mathcal{G}$	Arbitrary cubature grid
He_k	k-th order probabilists' Hermite polynomial
h	Step size of finite difference scheme
$h_{m,n}$	Norm of the n-th order PC basis function corresponding to the m-th stochastic parameter
h_k	Norm of the k-th order PC basis vector
L	Grid level
$LxEyOz$	Notation for Polynomial Chaos Expansion with $L = x$, $\Delta_{lev} = y$ and $O = z$
$\mathcal{M}_{\boldsymbol{\xi}}[\cdot]$	Arbitrary statistical measure over uncertainty vector $\boldsymbol{\xi}$
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with expected value μ and variance σ^2
N	Number of components in uncertainty vector
N_b	Number of beamspots
N_s	Number of samples
N_v	Number of voxels
O	Polynomial order
$p_\xi(\xi)$	Probability density function of stochastic parameter ξ
$P_{\alpha\%}^i\{\cdot\}$	α-th percentile of the set of values distinguished by the index i
$Q_{lev}^{(i)}$	i-th dimensional integral approximation with quadrature level lev
$R(\cdot)$	Arbitrary stochastic response
r	Polynomial Chaos Expansion coefficient
$\mathcal{S}$	Index set of samples
$\mathcal{V}$	Index set of voxels
$\boldsymbol{w}$	Weight vector
$\boldsymbol{x}$	Beam intensity vector

ABBREVIATIONS

CDF	Cumulative distribution function
CT	Computed tomography
CTV	Clinical target volume
DPH	Dose population histogram
DVH	Dose volume histogram
GTV	Gross tumor volume
HPC	High performance computing
IMPT	Intensity modulated proton therapy
MU	Monitor unit
OAR	Organ-at-risk
PBS	Pencil beam scanning
PCE	Polynomial Chaos Expansion
PDF	Probability density function
PTV	Planning target volume
RR	Range robustness
SOBP	Spread-out Bragg peak
SR	Set-up robustness

1

INTRODUCTION

1.1. CANCER TREATMENT

According to the World Health Organization, cancer has the second highest mortality rate of diseases worldwide with an estimated 9.6 million deaths in 2018 [1]. When a patient is diagnosed with cancer, there are several treatment methods to consider. The three most common modalities to treat cancer are surgery, chemotherapy and radiotherapy [2–4]. The first mentioned treatment method is removing the tumor through surgery. When the tumor cannot be reached without sparing enough healthy tissue and organs, chemotherapy or radiotherapy is often used. Chemotherapy uses chemical drugs to kill cancer cells. These drugs can be injected directly into the blood stream, or can be consumed by the patient in pill form. Chemotherapy avoids the use of surgery, but can cause serious side effects due to the powerful drugs that are used, which also reach other regions of the body besides the tumor [5]. Radiotherapy is the third method that is frequently used to treat cancer. It is estimated that half of all cancer patients are treated with radiotherapy [4]. Radiotherapy makes use of high doses of ionizing radiation, which induce damage to the DNA of tumor cells, finally resulting in cell death. All three methods can be used separately, or in a combined manner, as palliative treatment, focusing on pain and symptom relief, or as curative treatment to remove the tumor as much as possible and improve the patient's life expectancy.

1.2. RADIOTHERAPY

Radiotherapy itself covers a variety of treatment concepts, such as brachytherapy and external beam radiotherapy. In brachytherapy a sealed radiation source, containing a radioactive isotope, is placed inside or near the region of the body that requires treatment. The isotope irradiates the patient with ionizing particles. In external beam radiotherapy the patient will, most frequently, be irradiated with a concentrated beam of photons, electrons or protons. Nowadays, photon therapy is the conventional way of radiotherapy. However, the use of proton therapy is on the rise in the field of radiotherapy, with three proton therapy centers having already been opened in the Netherlands between 2018 and 2019. These centers are located in Delft, Groningen and Maastricht.

1.2.1. PROTON THERAPY

In proton therapy patients are irradiated with proton beams via a gantry, a medical machine that can rotate around the patient, while the patient is placed on the gantry table. Figure 1.1 shows a schematic illustration of a gantry and the gantry table. The protons in the beams deposit their energy along their path through the patient's body. Due to the differences in density between types of tissue, organs and the tumor, the amount of deposited energy differs per location. Therefore, the resulting energy deposition profile depends on the trajectory of the beam. However, it also depends on the type and energy of the particles in the beam and the beam's intensity, which is correlated to the number of particles in the beam.

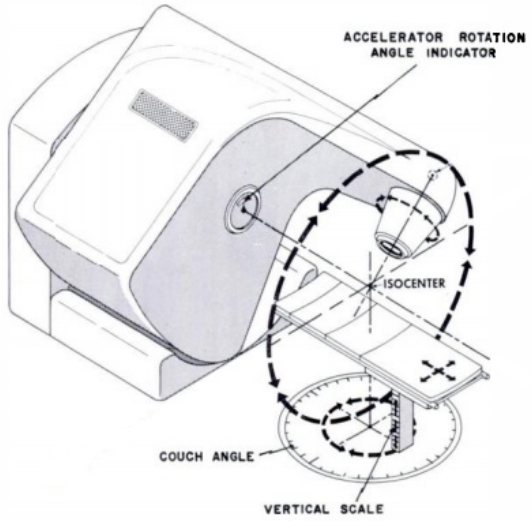


Figure 1.1: Schematic illustration of a medical proton therapy gantry, which can rotate around the patient. The patient lays on the gantry table, which can also be rotated and/or displaced. All these movements change the position of the isocenter, which is the center of irradiation, in the patient's body. Figure from [6].

Protons have a different energy deposition profile along the trajectory in the patient's body than photons. Photons deposit their energy starting with a peak at a small depth and a gradual decrease afterwards along their trajectory, whereas protons deposit very little energy, except for a peak at a specific position in depth. This peak of energy deposition for protons is called the Bragg peak. Figure 1.2 shows the difference in energy deposition profiles between a photon and a proton beam.

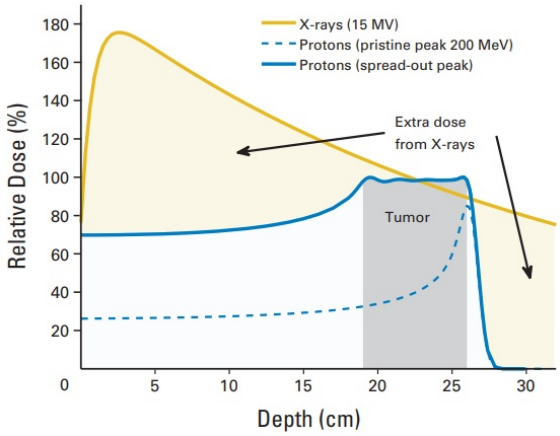


Figure 1.2: Comparison of the energy deposition profiles of a photon beam (X-ray) and a proton beam. On the horizontal and vertical axis, the depth along the trajectory in the patient's body and the delivered dose as a fraction of the prescribed dose in the tumor are shown respectively. The yellow line corresponds to a photon beam, and the blue dashed line corresponds to a single proton beam with its characteristic Bragg peak. When combining multiple proton beams with different energies together, a spread-out Bragg peak (SOBP) is formed, which corresponds to the solid blue line. Figure from [7].

Even though the use of proton therapy is on the rise, it had already been proposed by American physicist R.R. Wilson in 1946 [8]. The narrow Bragg peak and the sudden drop afterwards is a big advantage of proton therapy in comparison with conventional photon therapy, as more healthy tissue and organs can be spared. Conversely, the Bragg peak causes the dose distribution in proton therapy to be more sensitive to uncertainties. Translational and rotational errors in patient alignment, called set-up uncertainty, and errors in the energies of the proton beams that cause shifts in the Bragg peaks, called range uncertainty, are two important examples of such uncertainties.

1.2.2. HANDLING UNCERTAINTIES IN RADIOTHERAPY

Radiotherapy treatment planning starts with a CT scan of the patient. The tumor that is visible on the CT images is called the gross tumor volume (GTV). Due to the possible existence of microscopic extensions of the tumor, which are not visible on the CT images, a small margin around the GTV is considered. The GTV and this margin together form the clinical target volume (CTV). In photon therapy treatment planning, uncertainties are accounted for by irradiating a larger volume than the CTV. The CTV together with this extra margin is called the planning target volume (PTV). Figure 1.3 shows a schematic illustration of these margin definitions and three example scenarios of CTV motion in which the CTV receives the desired target dose in the case that the full PTV is irradiated with the target dose.

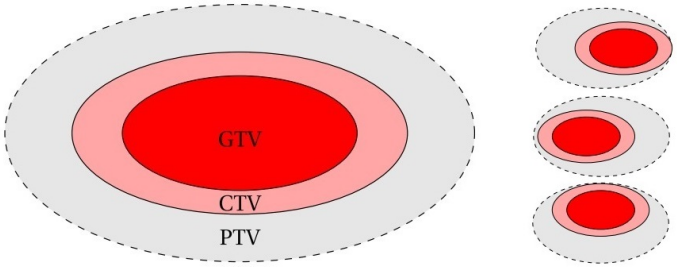


Figure 1.3: In the left image, a schematic illustration of the definitions of the gross tumor volume (GTV: tumor that is visible on the CT images), the clinical target volume (CTV: margin around GTV to account for microscopic tumor extensions) and the planning target volume (PTV: margin around CTV to account for set-up and range uncertainties) is visible. On the right, three scenarios of CTV motion are shown in which the CTV receives the desired target dose in the case that the full PTV is irradiated with the target dose. Figure from [9].

Unfortunately, this methodology is not suitable for proton therapy, because the energy deposition of protons is more sensitive to density variations in the body, as compared to photons. In proton therapy, robust treatment planning is currently used to account for set-up and range type of uncertainties [10]. In robust treatment planning, a discrete set of error scenarios is defined, with an error scenario being a specific realization of the considered uncertain parameters. With this discrete set of error scenarios a worst-case optimization is performed over the different scenarios. This means that the resulting treatment plan is robust against all these predefined discrete scenarios. However, each of these scenarios has statistically a zero probability of occurrence during treatment.

A way to overcome the limitations of robust treatment planning is to use probabilistic treatment planning. In probabilistic treatment planning one assumes the uncertain parameters to follow a certain probability density function (PDF) and an optimization is performed on statistical measures of stochastic dose parameters. A downside is that optimization for probabilistic treatment planning can become computationally expensive when these statistical measures are not analytically available, but dose distributions need to be sampled for many error scenarios with the dose algorithm to approximate these metrics.

The method of Polynomial Chaos Expansion (PCE) can be used to alleviate the computational cost in both robust and probabilistic treatment planning. PCE is a mathematical tool that can be used to approximate the functional dependence of the desired stochastic dose parameters on the considered uncertain parameters. The advantage of this method is that the PCEs can be evaluated fast for many error scenarios, alleviating computational cost in the evaluation of the objective functions and constraints, and their gradients and Hessians, in either robust or probabilistic treatment plan optimization.

1.3. PREVIOUS RESEARCH

In the past five years S.R. van der Voort (2015) and C.E. ter Haar (2018) have investigated the use of PCE to describe stochastic dose parameters and to construct robustness recipes for robust treatment planning within proton therapy [9, 11].

Furthermore, J.H. Salverda (2019) has performed research on the implementation of PCE in probabilistic treatment planning for proton therapy [12]. This research involved describing several statistical measures of stochastic dose parameters with PCE and, for some of those metrics, investigating their use in treatment planning. For the probabilistic function the expected value of the sum of the quadratic differences between the delivered dose and the prescribed dose explicit expressions were found in terms of characteristics of PCEs.

This probabilistic function was used in probabilistic planning on a simplified one- and three-dimensional geometry and good results were found with relatively low computational cost. For the 1D geometry also the effect of fractionation, dividing the treatment in separate parts, on the dose distribution has been investigated, however, this will not be covered in this research.

From a clinical perspective statistical measures like the expected value are less valuable than, for example, the percentile of a stochastic dose parameter, from which one can make statements about the percentage of error scenarios in which a treatment plan will be satisfied. Unfortunately, the percentile of a dose parameter cannot be expressed explicitly in terms of characteristics of PCEs, but sampling is needed to produce approximations. Implementation of the percentile as statistical measure in probabilistic planning on data of a patient with skull base meningioma, a type of brain tumor, with ErasmusMC's treatment planning software environment iCycle was found too costly for clinical use with computation times in the order of 1-3 weeks and has, in some cases, lead to problems with convergence [12].

1.4. RESEARCH GOAL

This research will contribute to the work of J.H. Salverda by improving probabilistic treatment planning with PCE for proton therapy. This will consist of developing new methods to speed up the calculation of the value, gradient and Hessian of percentiles of stochastic dose volume parameters and to test the new methods in treatment plan optimization. The goal of this research is to alleviate the computational cost of probabilistic treatment planning, in which percentiles of stochastic dose volume parameters are used, with the aim of a proof of principle for clinical use.

1.5. THESIS OUTLINE

This thesis is structured as follows. In Chapter 1, an introduction about different types of cancer treatment, and in particular external beam radiotherapy, is given with a short explanation of handling uncertainties in photon and proton therapy. Also, recent proton therapy research with the use of PCE within both robust and probabilistic treatment planning is discussed. In Chapter 2, the relevant theory of proton therapy physics and the concept of the pencil beam scanning technique in intensity modulated proton therapy are discussed. Also, different uncertainties in proton therapy are explained in more detail. In Chapter 3, several clinical treatment parameters, different concepts of treatment planning and treatment planning with ErasmusMC's software environment iCycle are explained. Chapter 4 contains the relevant mathematical background of PCE and the theory of numerical integration that is needed for the construction of PCEs. Also, the construction of PCEs with the OpenGPC package is explained. Chapter 5 shows how three probabilistic functions for this research can be described with PCEs. In Chapter 6, the different implementations of this research are explained. In Chapter 7, the results of this research are presented. Chapter 8 contains the important discussion points and conclusions of this research. Also, recommendations for future research are mentioned.

2

PROTON THERAPY

As discussed in Chapter 1.1, one of the methods to treat a patient with cancer is through the use of proton therapy. The radiation dose of the proton beams induces damage to the DNA of tumor cells, finally resulting in cell death. A patient is only treated completely when all cancerous cells are killed and do not return [13]. However, even after being cured, cancer patients can still encounter side effects of the radiotherapy treatment, due to damage to healthy tissue and organs near the tumor or along the trajectory of the proton beams.

It is important to understand how interactions between protons and matter occur, and what the effects are. These interactions affect the energy deposition of the protons along the trajectory in the patient's body. Chapter 2.1 covers the physics behind proton therapy, Chapter 2.2 introduces the application of intensity modulated proton therapy and Chapter 2.3 addresses three different uncertainties in proton therapy.

2.1. PHYSICS OF PROTON THERAPY

This section covers three important interactions of protons with matter that affect the energy deposition of the protons in a beam along the trajectory in the patient's body, and thus also the location of the Bragg peak. For this purpose, the definition of the range of a proton beam is introduced.

2.1.1. PROTON BEAM INTERACTIONS

Unlike the photons used in conventional radiotherapy, protons have a mass and are positively charged. Therefore, the interactions of protons with matter are different from that of photons. Most interactions of protons are with electrons and nuclei, possibly resulting in a change of kinetic energy and trajectory. Figure 2.1 shows the three most common types of interactions, in which p and p' denote a proton and e, n and γ denote an electron, a neutron and a photon respectively.

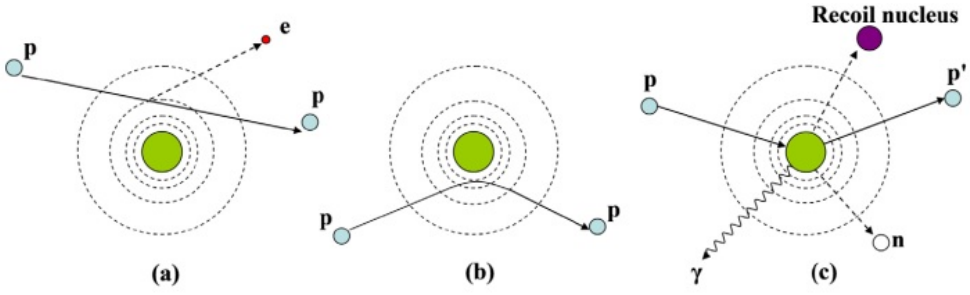


Figure 2.1: The three most common types of interactions between protons and other particles. **(a)** Inelastic Coulombic scattering. **(b)** Elastic Coulombic scattering. **(c)** Non-elastic nuclear interaction. p and p' denote a proton and e, n and γ denote an electron, a neutron and a photon respectively. Figures from [14].

The first type of interaction is inelastic Coulombic scattering of a proton with an electron. This causes a decrease in kinetic energy of the proton, but the trajectory of most protons is not changed due to the fact

that the rest mass of a proton is 1832 times the rest mass of an electron [14]. Figure 2.1a shows a schematic illustration of this type of interaction.

The second type of interaction is elastic Coulombic scattering of a proton with a nucleus, which is caused by the repulsive force between the positively charged proton and the positively charged nucleus. This repulsive force changes the kinetic energy and trajectory of the proton, as Figure 2.1b shows in a schematic illustration.

The third type of interaction is the non-elastic nuclear interaction of a proton with a nucleus, which is less common than the first two types of interactions [14]. This type of interaction occurs when there is direct impact of a proton with a nucleus. Figure 2.1c shows a schematic illustration of an example of this type of interaction.

All these interactions are not deterministic, but stochastic processes, and affect the kinetic energy and trajectory of the protons in a proton beam. In external beam radiotherapy, the choice can be made by considering the overall energy and trajectory of the proton beam, or instead looking at single protons. The latter is done by Monte Carlo dose calculations, however, in this research the former is used.

2.1.2. RANGE OF A PROTON BEAM

One important characteristic of the energy deposition profile of a proton beam is its range. The range of a proton beam is defined as the position at which the dose of the beam has decreased to 80% of the maximum dose, i.e., the 80% distal fall-off point [15]. Figure 2.2 illustrates the definition of the range of a single proton beam. The energy loss per unit length of a proton beam depends on the different stopping powers of the absorbing materials along the trajectory of the beam [16]. The range and the energy deposition profile of a proton beam can be determined with the use of these stopping powers. To calculate the energy loss per unit length of a proton beam, a conversion has to be made from the Hounsfield units of a CT scan to units of stopping power. In this conversion, however, uncertainties arise as the stopping power of a material is related to its density and the conversion of Hounsfield units to material density is not exact, but an approximation [17].

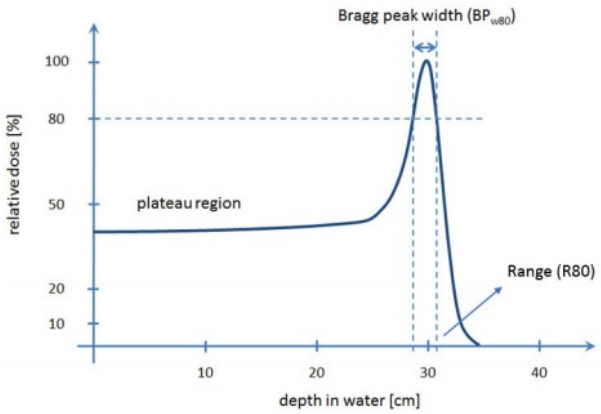


Figure 2.2: Example of an energy deposition profile of a proton beam with the delivered dose as a fraction of the prescribed dose in the tumor drawn as a function of the trajectory of the beam in water. As important dosimetry characteristics the plateau region, the width of the Bragg peak and the range of the proton beam are shown. The range of a proton beam is defined as the position at which the dose of the beam has decreased to 80% of the maximum dose. Figure from [18].

2.2. INTENSITY MODULATED PROTON THERAPY

In this research, the pencil beam scanning (PBS) technique is used. A pencil beam is a narrowly shaped beam of, in this case, protons. Each pencil beam can be characterized by its angle of irradiation, lateral position and energy. The combination of these three characteristics is called a beamspot. This research focuses on intensity modulated proton therapy (IMPT), in which a weight is assigned to each beamspot. This weight is proportional to the number of protons in the beam, defining its intensity. The goal in IMPT is to find the optimal weights for the beamspots, such that the resulting dose distribution has the best possible tumor coverage, whilst sparing the surrounding tissue and organs [19]. Figure 2.3 visualizes how PBS in IMPT works for the different energy layers of the pencil beams.

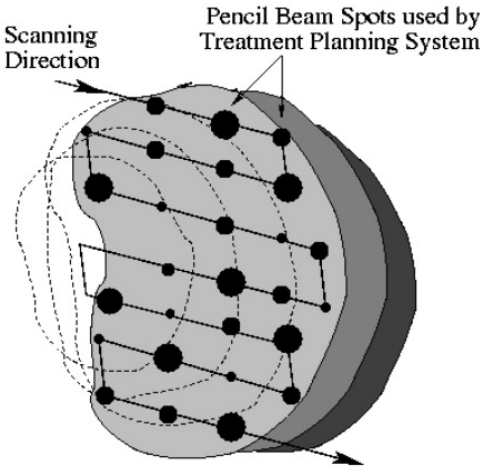


Figure 2.3: Schematic illustration of a cross-section of a tumor. The methodology of the pencil beam scanning technique in intensity modulated proton therapy for the different energy layers of the pencil beams is displayed, resulting in a beamspot grid in the tumor. The beamspots have different weights/intensities, which are visible through the size of the black dots. Figure from [20].

2.3. UNCERTAINTIES IN PROTON THERAPY

Handling uncertainties in proton therapy is of high importance to ensure adequate dose coverage in the tumor and sparing of surrounding tissue and organs. There are four main types of uncertainties in proton therapy, namely set-up and range uncertainties, the uncertainty in the delineation of the tumor and organs on the CT images of a patient and the uncertainty of internal motion of organs and the tumor. In this section, the first three types of uncertainties are discussed, however, in this research only set-up uncertainties are considered.

2.3.1. SET-UP UNCERTAINTY

A treatment plan is made according to a CT scan, and if any errors occur in the positional information from the CT scan and/or in the alignment of the patient during treatment, this causes a shift in the dose distribution, which is not desired. These errors are called set-up errors and can be divided into a systematic and a random component. A systematic set-up error is caused by errors in positional information from the CT scan, and is therefore constant for each treatment fraction. The assumption here is that only one CT scan is regarded for the whole treatment plan. A random set-up error originates from misalignment of the patient during treatment compared to its position during the CT scan, and is, therefore, different for each treatment fraction. Both the positioning of the patient and the gantry table itself contribute to the misalignment of the patient. The set-up error could be a translational or rotational error, or a combination of the two. A rotational set-up error could arise from a tilted gantry table.

Any uncertainty vector will be denoted by $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)^T$, having N components. Usually, both a systematic and a random set-up error have three components, corresponding to the three dimensions in Euclidean space. Furthermore, these components are assumed to be Gaussian distributed with zero mean and a positive variance, and are assumed to be mutually independent. The assumption of this type of PDF is in accordance with results of clinical research [21].

For a component of the systematic set-up error, ξ_i, the variance is denoted by Σ_i^2. This means that $\xi_i \sim \mathcal{N}\left(0, \Sigma_i^2\right)$, where $\mathcal{N}(\cdot,\cdot)$ denotes a Gaussian distribution. The PDF $p_{\xi_i}(\xi_i)$ of ξ_i can be formulated as in Equation 2.1.

$$p_{\xi_i}(\xi_i) = \frac{1}{\Sigma_i \sqrt{2\pi}} e^{-\frac{\xi_i^2}{2\Sigma_i^2}} \tag{2.1}$$

Similarly, for a component of the random set-up error, ξ_i, the variance is denoted by σ_i^2. This means that $\xi_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$ and the PDF $p_{\xi_i}(\xi_i)$ of ξ_i can be formulated similarly as Equation 2.1 with the substitution of σ_i^2 at the position of Σ_i^2. In Figure 2.4, a schematic illustration of two different realizations of a systematic set-up error, Σ_1 and Σ_2, is shown with three different realizations of a random set-up error, σ_1-σ_3 and σ_4-σ_6, drawn for both.

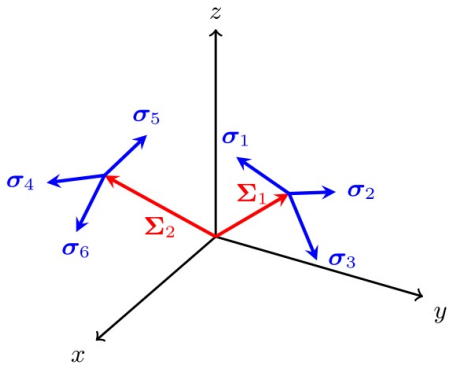


Figure 2.4: Schematic illustration of two different realizations of a systematic set-up error, Σ_1 and Σ_2. For each systematic set-up error three different realizations of a random set-up error, σ_1-σ_3 and σ_4-σ_6, are drawn. Figure from [12].

2.3.2. RANGE UNCERTAINTY

Range errors are present when the range of a proton beam is over- or underestimated. Range errors can be expressed in an absolute or a relative sense. There are two common causes of range errors. The first cause originates from the conversion of Hounsfield units of a CT scan to units of stopping power [22]. This conversion is used to calculate the range of a proton beam and in this calculation certain errors arise, as stated in Chapter 2.1.2. The second cause is a wrong calibration of the gantry. Figure 2.5 shows the differences in changes in energy deposition for a photon beam, a proton beam and a proton SOBP when range errors are present due to a change in the densities of the structures along the beam's trajectory. For the photon beam little changes, however, for the proton beams the Bragg peaks are shifted, resulting in a shifted SOBP, which is not desired when an organ-at-risk (OAR) is close-by. This is also the reason why the PTV margin methodology does not work for proton therapy, since the full PTV would be irradiated as if it is all tumor, but the resulting dose distribution will be different due to a wrong assumption of the densities in the PTV.

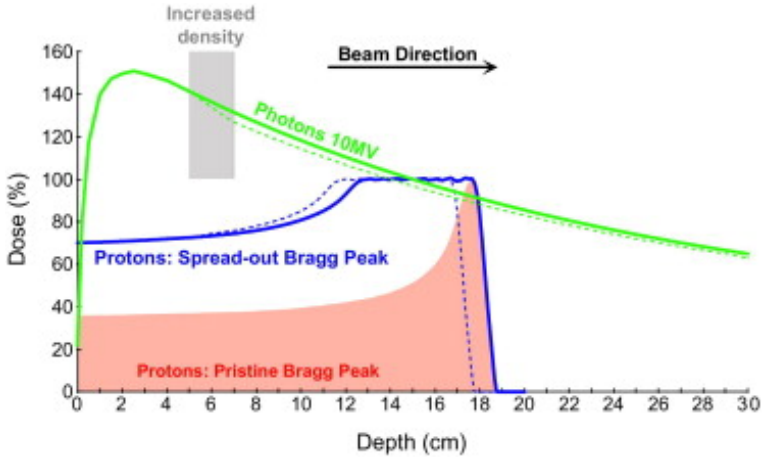


Figure 2.5: Illustration of the differences in changes in energy deposition for a photon beam, a proton beam and a proton SOBP when range errors are present due to a change in the densities of the structures along the beam's trajectory. Figure from [23].

2.3.3. DELINEATION UNCERTAINTY

Radiotherapy treatment planning starts with a CT scan of the patient. The CT scan is made up of images of adjacent cross-sectional slices of the body. Next, a physician will have to delineate the tumor and all organs of interest. The delineation will be done on all 2D images of the cross-sectional slices. A requirement is that at least all structures that will receive a dose are delineated. Besides the tumor, also the OARs get delineated by the physician. Of course, in the end one would like the tumor to receive a relatively high dose, while the OARs receive a relatively low dose. Due to the finite resolution of the CT images and possible issues with contrast, delineation errors arise. Furthermore, this type of error is observer-specific, due to a different interpretation

of the images by different physicians [24]. Figure 2.6 gives an example of the differences in delineation of the GTV for a brain tumor by different physicians.

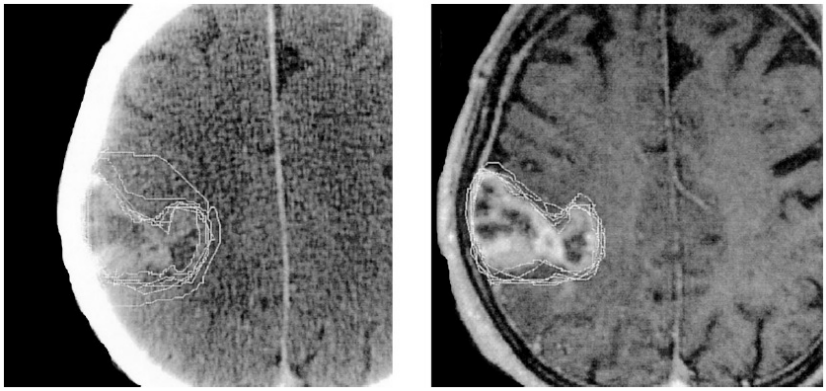


Figure 2.6: GTV delineation by nine different observers on an axial CT image (left) and an axial MRI image (right). Figure from [24].

3

TREATMENT PLANNING

In this chapter, different aspects of treatment planning are discussed. Chapters 3.1 and 3.2 address the definitions of relevant clinical treatment parameters and two ways to examine the quality of a treatment plan respectively. In Chapters 3.3 and 3.4, the methodology of automated treatment planning at the ErasmusMC and three important concepts of treatment planning are explained.

3.1. CLINICAL TREATMENT PARAMETERS

This section covers relevant clinical parameters that are involved in treatment planning. For convenience, the same notations as in [12] are used.

3.1.1. UNCERTAINTY VECTOR

The uncertainty vector is used to describe set-up and range type of errors. The uncertainty vector $\boldsymbol{\xi}$ is formulated according to Equation 3.1.

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)^T \tag{3.1}$$

The number of variables N in the uncertainty vector depends on the considered uncertain parameters. Usually, these variables are the three components of the random and systematic set-up error and the relative or absolute range error. The nominal scenario corresponds to $\boldsymbol{\xi} = \mathbf{0}$, which is when no errors are present.

3.1.2. BEAM INTENSITY VECTOR

In this research, the PBS technique is used in combination with IMPT. Via the gantry, a patient is irradiated with pencil beams of different energies from different positions and angles. As discussed in Chapter 2.2, the combination of the energy, lateral position and angle of irradiation of a pencil beam is called a beamspot. The set of beamspot indices in a treatment plan is denoted by $\mathcal{B} = \{1, \ldots, N_b\}$, with N_b the number of beamspots. In IMPT, a weight is assigned to each beamspot, defining its intensity. The beam intensity vector $\boldsymbol{x}$ contains the intensities of all beamspots, measured in Monitor Unit [MU]. The definition of the beam intensity vector is as follows:

$$\boldsymbol{x} = \left(x_1, \ldots, x_{N_b}\right)^T. \tag{3.2}$$

The elements of the beam intensity vector are physically constrained to be non-negative [19].

3.1.3. DOSE DEPOSITION MATRIX

The dose deposition matrix $D(\boldsymbol{\xi})$ describes the contribution of each beamspot to the deposited dose in each voxel. The set of beamspot indices is again denoted by $\mathcal{B} = \{1, \ldots, N_b\}$ and the set of voxel indices is denoted by $\mathcal{V} = \{1, \ldots, N_v\}$, with N_b and N_v the number of beamspots and voxels respectively. With the element D_{ij} of matrix D, the contribution of beamspot j to the dose received by voxel i, measured in Gray per Monitor Unit [$\mathrm{Gr} \cdot \mathrm{MU}^{-1}$], is described. The definition of the dose deposition matrix is:

$$D(\boldsymbol{\xi}) = \left[D_{ij}(\boldsymbol{\xi}) : i \in \mathcal{V}, j \in \mathcal{B}\right]. \tag{3.3}$$

The elements of the dose deposition matrix are also constrained to be non-negative. The dose deposition matrix depends on the uncertainty vector $\boldsymbol{\xi}$, because set-up and range type of errors affect the contribution of a certain beamspot to the dose received by a certain voxel.

3.1.4. DOSE VECTOR

The dose vector $\boldsymbol{d}(\boldsymbol{x},\boldsymbol{\xi})$ describes the dose received by each voxel, measured in Gray [Gy], and can be determined by the matrix-vector product of the dose deposition matrix $D(\boldsymbol{\xi})$ and the beam intensity vector $\boldsymbol{x}$, as described in Equation 3.4.

$$\boldsymbol{d}(\boldsymbol{x},\boldsymbol{\xi}) = D(\boldsymbol{\xi})\boldsymbol{x} \tag{3.4}$$

3.1.5. PRESCRIBED DOSE VECTOR

The prescribed dose vector $\boldsymbol{d}^P$ describes the clinically prescribed dose at each voxel, and is defined as in Equation 3.5.

$$\boldsymbol{d}^P = \left(d_1^P, \ldots, d_{N_v}^P\right)^T \tag{3.5}$$

The prescribed dose is relatively high at the voxels that correspond to the GTV, and the prescribed dose is relatively low at the voxels that correspond to healthy tissue and OARs.

3.1.6. DOSE VOLUME PARAMETERS

Until now, only dose parameters are considered that give information about a single voxel or beamspot. However, it might also be desirable to have information about how much dose is received by a certain volume fraction, which is useful to examine the quality of a treatment plan. A dose volume parameter can be used to describe such information for a discretized structure with $\mathcal{V}_{str} \subseteq \mathcal{V}$ the set of voxel indices and $N_{v,str} \leq N_v$. The term dose volume parameter is a general name for a family of dose volume-related quantities. An important dose volume parameter is the maximum dose that is received by at least α% of the volume and is denoted by $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ with $\alpha \in [0,100]$. This is equivalent to the maximum lower bound of the dose that is received by α% of the volume. Equation 3.6 formulates this parameter.

$$d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}) = P^i_{(100-\alpha)\%}\left\{d_i(\boldsymbol{x},\boldsymbol{\xi}) : i \in \mathcal{V}_{str}\right\} \tag{3.6}$$

The operator $P^i_{(100-\alpha)\%}\{\cdot\}$ denotes the $(100-\alpha)$-th percentile of the set of values distinguished by the index i. Other forms of dose volume parameters are also possible to define, for example, the minimum dose that is received by at least α% of the volume. For the definition of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ as described in Equation 3.6, there are two volume fractions for which the parameter has a specific name, namely the near-maximum dose $d_{2\%}(\boldsymbol{x},\boldsymbol{\xi})$ and the near-minimum dose $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$.

3.2. QUALITY EXAMINATION OF A TREATMENT PLAN

This section covers two approaches in examining the outcome of a treatment plan.

3.2.1. DOSE VOLUME HISTOGRAM

Dose volume histograms (DVHs) are used to visualize dose parameters as a function of the considered volume fraction. A DVH is an inverse cumulative histogram with on the horizontal axis the dose parameter and on the vertical axis the volume fraction, and shows for each dose value which volume fraction receives at least this dose. Figure 3.1 shows an example of a DVH for an optimized probabilistic treatment plan.

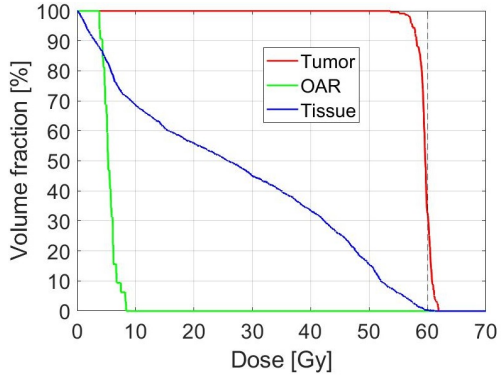


Figure 3.1: Dose volume histogram of the expected dose (weighted average of the dose distributions over the joint PDF of the uncertain parameters) with the red, green and blue line corresponding to the tumor, an OAR and healthy tissue respectively.

3.2.2. DOSE POPULATION HISTOGRAM

A dose population histogram (DPH) is used to visualize the probability distribution of a dose parameter. A DPH is similar to a DVH, but on the vertical axis the probability or population is shown, instead of the volume fraction. Considering a population, a set of patients, or for a single patient a set of error scenarios, could be meant. Figure 3.2 shows an example of a DPH.

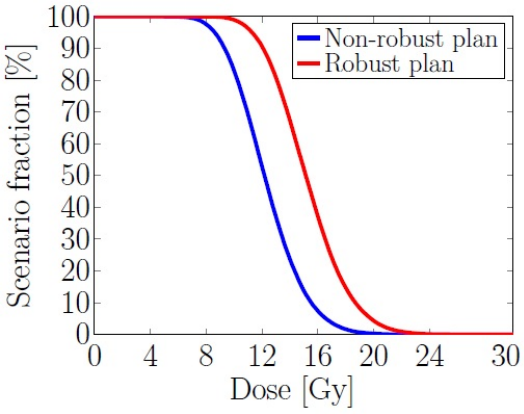


Figure 3.2: Dose population histogram of the mean dose in an OAR as a function of the fraction of error scenarios. The red and blue line correspond to a non-robust and a robust plan respectively. Figure from [25].

3.3. AUTOMATED TREATMENT PLANNING WITH ERASMUSMC'S ICYCLE

The ErasmusMC uses the in-house developed software environment Erasmus-iCycle, iCycle in short, to perform automated treatment planning [26]. The majority of functionalities in iCycle are coded in Matlab. Automated treatment planning in iCycle is performed through a multi-criteria optimization according to a specified wish-list. In this research, treatment plan optimizations are not performed on patient data in iCycle, but are performed on a simplified 3D geometry in Matlab. However, as this research is closely related to recent implementations in iCycle in [12], the information about wish-lists and multi-criteria optimizations in iCycle are shortly discussed to place this research into perspective.

3.3.1. WISH-LIST

The wish-list contains all dose requirements of a treatment plan. Here, two types of requirements are distinguished, namely constraints and objectives. The first type of requirement is a constraint, which should always be met in the treatment plan. The second type of requirement is an objective, which is tried to be reached as close as possible. The wish-list consists of a set of constraints and objective goals, with the objectives having a specified priority of optimization. These criteria are for example the minimum, maximum or mean dose in the GTV or a certain OAR. Also, one needs to specify whether the criterion should be minimized or maximized. A wish-list is specific for the tumor site, but is usually not patient specific [27]. In Table 3.1, an example of a simplified wish-list for a skull base meningioma patient is shown, in which $\bar{d}$ denotes the mean dose.

Table 3.1: An example of a simplified wish-list for a skull base meningioma patient, in which $\bar{d}$ denotes the mean dose. Note that in this wish-list only the nominal scenario is considered. Source: [12].

Constraints:		Volume:	Type:	Limit [Gy]:
		GTV	Maximize $\min_i d_i(\boldsymbol{x},\boldsymbol{0})$	$0.95 \cdot 50.4$
		Optic Nerve (L)	Minimize $\max_i d_i(\boldsymbol{x},\boldsymbol{0})$	50
		Optic Nerve (R)	Minimize $\max_i d_i(\boldsymbol{x},\boldsymbol{0})$	50
Objectives:	Priority:	Volume:	Type:	Goal [Gy]:
	1	Brainstem	Minimize $\bar{d}(\boldsymbol{x},\boldsymbol{0})$	1
	2	Cerebellum	Minimize $\bar{d}(\boldsymbol{x},\boldsymbol{0})$	1

3.3.2. MULTI-CRITERIA OPTIMIZATION

The goal of the treatment plan optimization is to find the configuration of pencil beam directions, beamspots and the beamspot weights that produce the optimal dose distribution with respect to the considered wish-list. The beam directions are optimized through an iterative process of adding beam directions, calculating their contributions to the dose in the GTV and the surrounding OARs and omitting beam directions that contribute the least [28]. In iCycle, the beam directions can also be chosen manually to save computation time. This is usually based on clinical experience. With the beam directions chosen, a similar iterative process is performed for the configuration of beamspots [29]. This process is called beamspot resampling. To alleviate computational cost the iCycle feature voxel reduction can be enabled. This feature uses a representative subset of voxels in a structure to perform the dose calculations.

The multi-criteria optimization in iCycle works according to the two-phase ϵ-constraint method to calculate the corresponding beamspot weights [26, 30]. In the first phase, the objectives are optimized in the order of priority, within the prescribed constraints. The gradient and Hessian of the objective functions and constraints are needed to guide the optimization in a proper direction. The dose distribution is calculated with the dose engine, which is a general term for the dose calculation algorithm. Once the assigned goal of an objective has been reached, this objective is set as a constraint for the next objective. To improve the feasibility of convergence of the next objectives, the limit of this new constraint is slightly relaxed with respect to the goal of the initial objective. This relaxation factor is set to a margin of around 3%. Subsequently, the next objective gets optimized, and this goes on until all objectives are optimized. An objective with a lower priority is less likely to reach its goal, because more constraints are to be met. In the second phase, the objectives that have been optimized in the first phase are optimized again, but this time further than their assigned goal value as far as possible. After the two phases, a Pareto optimum is found, which means that no parameters can be improved without giving up on some other parameter.

A physician or dosimetrist should always review the outcome of the treatment plan. If this person believes that the current treatment plan does not spare certain OARs enough or the resulting dose distribution is not optimal to their knowledge, the directions and weights of the proton pencil beams could be changed and a new treatment plan is constructed.

3.4. CONCEPTS OF TREATMENT PLANNING

With the methodology in Chapter 3.3, a proton therapy treatment plan can be obtained. However, this plan does not necessarily account for uncertainties that are present. In practice, it is a certainty that set-up and range type of errors, as mentioned in Chapter 2.3, will occur. All these errors together result in a different dose distribution than the optimized dose distribution, which is not desired. Therefore, there are different concepts of treatment planning within proton therapy to consider. In Chapters 3.4.1, 3.4.2 and 3.4.3, the concepts of conventional, robust and probabilistic treatment planning are explained respectively.

3.4.1. CONVENTIONAL TREATMENT PLANNING

In conventional treatment planning, no uncertain parameters are considered. Only the nominal scenario, written as $\boldsymbol{\xi} = \mathbf{0}$, is taken into account. If we consider an objective function $f(\boldsymbol{x}, \boldsymbol{\xi})$, with $\boldsymbol{x}$ the beam intensity vector and $\boldsymbol{\xi}$ the uncertainty vector, the optimization as in Equation 3.7 is performed. The resulting treatment plan is guaranteed to satisfy the objective goals and constraints in the nominal scenario, which has statistically a zero probability of occurrence. This coincides with the construction of a treatment plan similarly to the type of objective goals and constraints as in the wish-list in Table 3.1. Conventional treatment planning has been implemented in iCycle.

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}, \mathbf{0}) \tag{3.7}$$

3.4.2. ROBUST TREATMENT PLANNING

In robust treatment planning, a discrete set of error scenarios is taken into account in the optimization of the treatment plan. The resulting treatment plan is robust against these predefined scenarios, however, each of these scenarios has statistically a zero probability of occurrence. Nevertheless, this would be more accurate than conventional treatment planning. Robust treatment planning is, at the moment, a frequently used concept of handling uncertainties in proton therapy [9, 11, 31]. A multi-criteria minimax optimization approach is used in which objective-wise the worst-case values over all predefined scenarios are optimized, within the worst-case values of the constraints, until each objective is optimized. Considering the set of predefined scenarios with indices $s \in \mathcal{S} = \{1, \ldots, N_s\}$ and N_s the number of scenarios, and the objective function $f(\boldsymbol{x}, \boldsymbol{\xi}^s)$ with

$\boldsymbol{\xi}^s$ the realization of the uncertainty vector in scenario s, the optimization can be written as in Equation 3.8. The treatment plan will satisfy the wish-list in all defined scenarios.

$$\min_{\boldsymbol{x}} \max_{s \in \mathcal{S}} f\left(\boldsymbol{x}, \boldsymbol{\xi}^s\right) \tag{3.8}$$

Robust treatment planning has also been implemented in iCycle. In Table 3.2, a possible set of error scenarios is shown, on which a robust treatment plan could be based. In this treatment plan, set-up and range type of errors are considered by defining the so-called set-up robustness (SR) and range robustness (RR) settings.

Table 3.2: A possible set of error scenarios to use in robust treatment planning, based on a defined set-up and range robustness. Source: [9].

Scenario:	x [mm]	y [mm]	z [mm]	Range (%)
Nominal scenario	0	0	0	0
Positive shift in x-direction	SR	0	0	0
Negative shift in x-direction	-SR	0	0	0
Positive shift in y-direction	0	SR	0	0
Negative shift in y-direction	0	-SR	0	0
Positive shift in z-direction	0	0	SR	0
Negative shift in z-direction	0	0	-SR	0
Range overestimation	0	0	0	RR
Range underestimation	0	0	0	-RR

3.4.3. PROBABILISTIC TREATMENT PLANNING

In probabilistic treatment planning, one assumes the uncertain parameters to follow a certain PDF. As mentioned in Chapter 2.3, each uncertain parameter is assumed to be mutually independent and Gaussian distributed with mean zero and a certain positive variance. This defines the continuous sample space of the uncertainty vector $\boldsymbol{\xi}$. In contrast to robust treatment planning, in probabilistic treatment planning also objective goals and constraints containing statistical measures of stochastic dose parameters are used in the wish-list, such that the full sample space or a subset, both containing an infinite amount of error scenarios, can be taken into account when optimizing the treatment plan. This results in the possibility to make probabilistic statements about the optimized treatment plan.

Besides the assumed PDF for the uncertain parameters, also a choice has to be made of which statistical measure to optimize for. As an example, one could guarantee α% of the GTV to receive at least γ Gy in β% of the error scenarios. Requirements are that such a solution exists, while conforming to the complete wish-list, and that the optimization converges. In this example, the statistical measure is the percentile, but of course other statistical measures like the (conditional) expected value, mode or median are also possible to optimize for. However, from a clinical perspective the percentile is a more relevant type of statistical measure, giving information about a percentage of error scenarios in which the wish-list is satisfied.

Considering a statistical measure $\mathcal{M}_{\boldsymbol{\xi}}[\cdot]$ over the uncertainty vector $\boldsymbol{\xi}$, an arbitrary objective function $f(\cdot)$ and an arbitrary stochastic variable $g(\boldsymbol{x}, \boldsymbol{\xi})$, which could represent any dose parameter, the optimization in probabilistic treatment planning is described as in Equation 3.9. Recently, probabilistic treatment planning has also been implemented in iCycle [12].

$$\min_{\boldsymbol{x}} f\left(\mathcal{M}_{\boldsymbol{\xi}}\left[g(\boldsymbol{x}, \boldsymbol{\xi})\right]\right) \tag{3.9}$$

4

POLYNOMIAL CHAOS EXPANSION

This chapter covers the relevant theory on the method of Polynomial Chaos Expansion. In Chapters 4.1 and 4.2, related parts of probability theory and statistics are discussed respectively, based on literature [32, 33]. Chapters 4.3, 4.4 and 4.5 treat the mathematics of PCE, numerical integration and the concept of hyperbolic trimming respectively, based on literature [32–41]. Chapter 4.6 describes how the construction of PCEs is performed and which notation is used for the characteristics of constructed PCEs.

4.1. PROBABILITY THEORY

The concept of probabilistic treatment planning is heavily related to the uncertain parameters that are considered. These uncertain parameters define the underlying sample space Ω that needs to be taken into account when making statements about statistical measures of dose parameters. A random outcome of the sample space is denoted by $\theta \in \Omega$ and is described by a realization of the uncertainty vector $\boldsymbol{\xi}(\theta)$ as:

$$\boldsymbol{\xi}(\theta) = (\xi_1(\theta), \ldots, \xi_N(\theta))^T, \tag{4.1}$$

similarly to Equation 3.1 with N the number of uncertain parameters. The set of possible events is denoted by $\mathcal{F}$, which is a σ-algebra on Ω. The possible events are mapped to their probability of occurrence through the probability measure $P : \mathcal{F} \rightarrow [0, 1]$. Altogether, this constructs the probability space $(\Omega, \mathcal{F}, P)$.

In this research, the dose parameters can be described as real-valued stochastic responses $R(\theta) : \Omega \rightarrow \mathbb{R}$. These responses are assumed to be square-integrable functions, so belonging to the L^2-space, which is defined as:

$$L^2(\Omega, P) = \{R(\theta) : [R(\theta) : \Omega \rightarrow \mathbb{R}] \wedge [\langle R, R \rangle < \infty]\}. \tag{4.2}$$

In the L^2-space the inner product is defined as:

$$\langle Q, R \rangle = \int_\Omega Q(\theta) R(\theta) dP(\theta) = \int_{\mathcal{D}} Q(\boldsymbol{\xi}(\theta)) R(\boldsymbol{\xi}(\theta)) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{4.3}$$

for arbitrary stochastic responses Q and R, and $\mathcal{D}$ being the domain of the uncertainty vector. In the rest of this research the dependence of $\boldsymbol{\xi}$ on θ is omitted, because this relation is not needed throughout the rest of the theory.

4.2. STATISTICS

The different uncertain parameters are assumed to be Gaussian distributed with zero mean ($\mu_{\xi_i} = 0$) and a certain positive variance $\sigma^2_{\xi_i}$, as mentioned in Chapter 2.3. Therefore, the PDF of a single uncertain parameter $\xi_i \sim \mathcal{N}\left(0, \sigma^2_{\xi_i}\right)$ is known as:

$$p_{\xi_i}(\xi_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{\xi_i^2}{2\sigma_i^2}}. \tag{4.4}$$

Furthermore, we assume the uncertain parameters to be mutually independent, which means that the joint PDF of the uncertainty vector $\boldsymbol{\xi}$ can be written as:

$$p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) = \prod_{i=1}^{N} p_{\xi_i}(\xi_i), \tag{4.5}$$

with N the number of uncertain parameters. Another important feature of these mutually independent uncertain parameters is the additivity of their Gaussian distributions. If we assume $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ to be two independent Gaussian random variables, and we define $Z = X + Y$ to be a new random variable, then for the random variable Z we find that $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

For the stochastic response R two important characteristics are its expected value and variance. The expected value, or mean, of the stochastic response R is defined as:

$$\mu_R = \mathrm{E}[R] = \int_{\mathbb{R}^N} R(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{4.6}$$

and the variance is defined as:

$$\sigma_R^2 = \mathrm{Var}(R) = \mathrm{E}\left[(R - \mathrm{E}(R))^2\right] = \int_{\mathbb{R}^N} (R(\boldsymbol{\xi}) - \mu_R)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{4.7}$$

The standard deviation of the stochastic response, σ_R, is defined as the square root of its variance. Here, the shorthand notation of a single multidimensional integral over $\mathbb{R}^N$ is used instead of N nested one-dimensional integrals over $\mathbb{R}$.

In practice, not all error scenarios in the full sample space can occur. The assumption of a continuous non-truncated Gaussian PDF is still accurate, because with the right choice of the variance, these extreme scenarios have almost no effect on the accuracy of statistical measures of stochastic dose parameters that are calculated with PCEs based on the full sample space.

4.3. MATHEMATICS OF PCE

Polynomial Chaos Expansion is a method which is used to describe the functional dependence of an arbitrary stochastic response $R(\boldsymbol{\xi})$ on the uncertainty vector $\boldsymbol{\xi}$ when an explicit analytical expression is unknown. In radiotherapy, the stochastic response can represent any stochastic dose parameter. The advantage is that typically the PCE can be evaluated faster than the stochastic response itself, calculated from the dose engine. The PCE is a polynomial expansion of the stochastic response and consists of the sum of basis vectors $\Psi_k(\boldsymbol{\xi})$ with each basis vector having an expansion coefficient r_k, as defined in Equation 4.8. The PC basis vectors are not actual vectors, but multivariate functions.

$$R(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} r_k \Psi_k(\boldsymbol{\xi}). \tag{4.8}$$

For the PCE to be useful in practice, the expansion is truncated to contain only a finite number of basis vectors, namely $P+1$, yielding the following approximation of the stochastic response:

$$R(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} r_k \Psi_k(\boldsymbol{\xi}). \tag{4.9}$$

The basis vector $\Psi_k(\boldsymbol{\xi})$ is the product of basis functions $\phi_{j,\gamma_{kj}}(\xi_j)$, one for each uncertain parameter, with each basis function being a univariate polynomial. This can be represented as:

$$\Psi_k(\boldsymbol{\xi}) = \prod_{j=1}^{N} \phi_{j,\gamma_{kj}}(\xi_j), \tag{4.10}$$

with the multi-index γ_{kj} being the order of the j-th polynomial basis function for the k-th basis vector. One needs to define which PC basis set and type of polynomial basis functions are used to construct the PCE. This changes the accuracy of the constructed PCE for a given basis set and its rate of convergence to the stochastic response.

4.3.1. ORTHOGONALITY OF BASIS FUNCTIONS AND BASIS VECTORS

For convergence of our polynomial expansion to the stochastic response $R(\boldsymbol{\xi})$, orthogonality of the univariate polynomials $\phi_{j,\gamma_{kj}}(\xi_j)$ is recommended [32]. A property of orthogonal polynomials is that the inner product of two polynomials is only non-zero if their order is the same. With $h_{j,n}$ representing the norm of the n-th order polynomial basis function of the j-th uncertain parameter, $\delta_{k,l}$ the Kronecker delta, and $p_{\xi_j}(\xi_j)$ the PDF weight function in the inner product, this can be formulated as:

$$\left\langle \phi_{j,m}(\xi_j), \phi_{j,n}(\xi_j) \right\rangle = \int_{\infty}^{\infty} \phi_{j,m}(\xi_j)\phi_{j,n}(\xi_j)p_{\xi_j}(\xi_j)d\xi_j = h_{j,n}^2\delta_{m,n}, \quad \forall\, j \in \{1,\ldots,N\}. \tag{4.11}$$

Equation 4.12 shows that the orthogonality of the basis vectors $\Psi_k(\boldsymbol{\xi})$ can be derived from the orthogonality of the basis functions $\phi_{j,\gamma_{kj}}(\xi_j)$ with h_k the norm of the k-th basis vector.

$$\begin{aligned}\langle \Psi_k(\boldsymbol{\xi}), \Psi_l(\boldsymbol{\xi}) \rangle = \int_{\mathbb{R}^N} \Psi_k(\boldsymbol{\xi})\Psi_l(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} &= \int_{\mathbb{R}^N} \prod_{j=1}^{N} \phi_{j,\gamma_{kj}}(\xi_j) \prod_{j'=1}^{N} \phi_{j',\gamma_{lj'}}(\xi_{j'})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= \prod_{j=1}^{N} \int_{\infty}^{\infty} \phi_{j,\gamma_{kj}}(\xi_j)\phi_{j,\gamma_{lj}}(\xi_j)p_{\xi_j}(\xi_j)d\xi_j \\ &= \prod_{j=1}^{N} h_{j,\gamma_{kj}}^2 \delta_{\gamma_{kj},\gamma_{lj}} = \delta_{k,l} \prod_{j=1}^{N} h_{j,\gamma_{kj}}^2 = \delta_{k,l}h_k^2\end{aligned} \tag{4.12}$$

In the derivation of Equation 4.12, Equations 4.10, 4.5 and 4.11 are substituted at the second, third and fourth equal-sign respectively. Next, we need to define the basis functions such that Equations 4.11 and 4.12 hold. Because of the assumption of Gaussian distributed uncertain parameters, the probabilists' Hermite polynomials will be chosen as basis functions, due to the fact that they are orthogonal with respect to the Gaussian distribution on $\mathbb{R}$ [39]. The relations between the type of distribution and its corresponding type of orthogonal polynomials are in literature provided by the Wiener-Askey scheme. The part of the Wiener-Askey scheme for continuous distributions is shown in Table 4.1.

Table 4.1: Part of the Wiener-Askey scheme for continuous distributions and corresponding type of polynomials with their domain of support for $a, b \in \mathbb{R}$. This scheme provides the best choice of the PC basis functions, when considering a certain continuous distribution for the stochastic input of the PCE, for convergence to the response. Source: [39].

Distribution:	Polynomial:	Domain of support:
Beta	Jacobi	$[a,b]$
Gamma	Laguerre	$[0,\infty)$
Gaussian	Probabilists' Hermite	$(-\infty,\infty)$
Uniform	Legendre	$[a,b]$

4.3.2. PROBABILISTS' HERMITE POLYNOMIALS

The probabilists' Hermite polynomials are defined as:

$$He_k(\xi) = (-1)^k e^{\frac{\xi^2}{2}} \frac{d^k}{d\xi^k}\left(e^{-\frac{\xi^2}{2}}\right), \quad \forall\, k \in \mathbb{N} \cup \{0\}, \tag{4.13}$$

with k the polynomial order [40]. According to Equation 4.13, the first two probabilists' Hermite polynomials are given by $He_0(\xi) = 1$, $He_1(\xi) = \xi$. The second and higher order probabilists' Hermite polynomials can, besides Equation 4.13, also be determined via the following recurrence relation:

$$He_{k+1}(\xi) = \xi He_k(\xi) - kHe_{k-1}(\xi), \quad \forall\, k \in \mathbb{N}, \tag{4.14}$$

with $He_0(\xi) = 1$, $He_1(\xi) = \xi$ given. The first six probabilists' Hermite polynomials are shown in Figure 4.1.

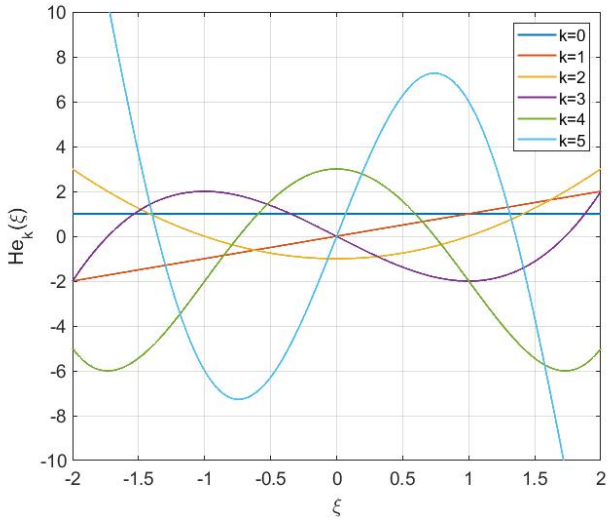


Figure 4.1: The first six probabilists' Hermite polynomials.

The inner product of the k-th order probabilists' Hermite polynomial with itself is related to the polynomial order k when assuming a Gaussian PDF for the variable ξ. This relation is shown in Equation 4.15.

$$\langle He_k(\xi), He_k(\xi)\rangle = k! \tag{4.15}$$

4.3.3. PCE COEFFICIENTS

The derivation of the formula for the PCE coefficients r_k can be performed by multiplying both sides of Equation 4.8 by $\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ and integrating both sides of the expression over $\mathbb{R}^N$, yielding the following result:

$$\begin{aligned}
\langle R(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle = \int_{\mathbb{R}^N} R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} &= \int_{\mathbb{R}^N}\left(\sum_{l=0}^{\infty} r_l\Psi_l(\boldsymbol{\xi})\right)\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\
&= \int_{\mathbb{R}^N}\left(\sum_{l=0}^{P} r_l\Psi_l(\boldsymbol{\xi})\right)\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\
&= \sum_{l=0}^{P} r_l \int_{\mathbb{R}^N} \Psi_l(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\
&= \sum_{l=0}^{P} r_l \langle \Psi_l(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle \\
&= r_k \langle \Psi_k(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle .
\end{aligned} \tag{4.16}$$

It is allowed to interchange summation and integration, since it concerns a finite summation. Orthogonality of the basis vectors is also used. Equation 4.17 shows the resulting expression for the PCE coefficients r_k.

$$r_k = \frac{\langle R(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle}{\langle \Psi_k(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle} = \frac{1}{h_k^2}\langle R(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle = \frac{1}{h_k^2}\int_{\mathbb{R}^N} R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi} \tag{4.17}$$

Notice that in Equation 4.17 the expression for the stochastic response R is still needed to calculate the expansion coefficients, which are contained in the response. Therefore, numerical integration methods are needed, which will be introduced in Chapter 4.4.

4.3.4. EXPECTED VALUE AND VARIANCE OF PCE

As stated in Chapter 4.2, two important characteristics of the stochastic response are its expected value and variance. It would be convenient to have an expression for the expected value and variance of the PCE in terms of characteristics of the PCE. For the expected value of the stochastic response, μ_R, the following can

be derived:

$$\begin{aligned}\mu_R = \int_{\mathbb{R}^N} R(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_{\mathbb{R}^N} \left(\sum_{k=0}^{\infty} r_k \Psi_k(\boldsymbol{\xi}) \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_{\mathbb{R}^N} \left(\sum_{k=0}^{P} r_k \Psi_k(\boldsymbol{\xi}) \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ = \sum_{k=0}^{P} r_k \int_{\mathbb{R}^N} \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ = \sum_{k=0}^{P} r_k \int_{\mathbb{R}^N} \Psi_0(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ = \sum_{k=0}^{P} r_k \langle \Psi_0(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi}) \rangle = \sum_{k=0}^{P} r_k \delta_{0,k} h_k^2 = r_0 h_0^2 = r_0.\end{aligned} \tag{4.18}$$

In Equation 4.18, Equation 4.8 is substituted at the second equal-sign, summation and integration can be interchanged since it concerns a finite summation, the fact that $\Psi_0 = 1$ and thus $h_0^2 = 1$ is used after the fourth and eighth equal-sign respectively, and Equation 4.12 is substituted at the sixth equal-sign. This derivation shows that the expected value of $R(\boldsymbol{\xi})$ is represented by the zeroth order PCE coefficient. For the variance of the stochastic response, σ_R^2, the following can be derived:

$$\begin{aligned}\sigma_R^2 &= \int_{\mathbb{R}^N} \left(R(\boldsymbol{\xi}) - \mu_R \right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^N} R^2(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} - 2 \int_{\mathbb{R}^N} R(\boldsymbol{\xi}) \mu_R p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \int_{\mathbb{R}^N} \mu_R^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &\approx \int_{\mathbb{R}^N} \left(\sum_{k=0}^{P} r_k \Psi_k(\boldsymbol{\xi}) \right) \left(\sum_{k'=0}^{P} r_{k'} \Psi_{k'}(\boldsymbol{\xi}) \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} - 2\mu_R \int_{\mathbb{R}^N} \left(\sum_{k=0}^{P} r_k \Psi_k(\boldsymbol{\xi}) \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \mu_R^2 \int_{\mathbb{R}^N} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \sum_{k=0}^{P} \sum_{k'=0}^{P} r_k r_{k'} \int_{\mathbb{R}^N} \Psi_k(\boldsymbol{\xi}) \Psi_{k'}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} - 2\mu_R \sum_{k=0}^{P} r_k \int_{\mathbb{R}^N} \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \mu_R^2 \\ &= \sum_{k=0}^{P} \sum_{k'=0}^{P} r_k r_{k'} \int_{\mathbb{R}^N} \Psi_k(\boldsymbol{\xi}) \Psi_{k'}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} - 2\mu_R \sum_{k=0}^{P} r_k \int_{\mathbb{R}^N} \Psi_0(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \mu_R^2 \\ &= \sum_{k=0}^{P} \sum_{k'=0}^{P} r_k r_{k'} \langle \Psi_k(\boldsymbol{\xi}), \Psi_{k'}(\boldsymbol{\xi}) \rangle - 2\mu_R \sum_{k=0}^{P} r_k \langle \Psi_0(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi}) \rangle + \mu_R^2 \\ &= \sum_{k=0}^{P} \sum_{k'=0}^{P} r_k r_{k'} \delta_{k,k'} h_k^2 - 2\mu_R \sum_{k=0}^{P} r_k \delta_{0,k} h_k^2 + \mu_R^2 \\ &= \sum_{k=0}^{P} r_k^2 h_k^2 - 2\mu_R r_0 + \mu_R^2 \\ &= \sum_{k=0}^{P} r_k^2 h_k^2 - 2\mu_R^2 + \mu_R^2 = \sum_{k=0}^{P} r_k^2 h_k^2 - \mu_R^2 = \sum_{k=0}^{P} r_k^2 h_k^2 - r_0^2 = \sum_{k=0}^{P} r_k^2 h_k^2 - r_0^2 h_0^2 = \sum_{k=1}^{P} r_k^2 h_k^2.\end{aligned} \tag{4.19}$$

In Equation 4.19, Equation 4.9 is substituted at the third equal-sign, summation and integration can be interchanged since it concerns a finite summation, the fact that $\int_{\mathbb{R}^N} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = 1$ is used after the third equal-sign, the fact that $\Psi_0 = 1$ and thus $h_0^2 = 1$ is used after the fifth and twelfth equal-sign respectively, and Equations 4.12 and 4.18 are substituted at the seventh and eighth equal-sign respectively. This derivation shows that the approximation of the variance of the stochastic response can be expressed as a function of the expansion coefficients and the norms of the PC basis vectors.

4.3.5. POLYNOMIAL CHAOS BASIS SET

Besides the choice for the type of polynomial basis functions to use in the construction of a PCE, a choice has to be made about which orders of polynomials to use for the different uncertain parameters. In Equation 4.9 the full PCE is truncated to contain only a finite number of basis vectors and Equation 4.10 shows that the different basis vectors are built from basis functions. However, the different orders of the basis functions are still to be defined, resulting in the orders of the basis vectors. The following two definitions are introduced to construct the set of PC basis vectors [32]. The first definition is the set of basis vectors Γ_o for a combined

polynomial order o.

$$\Gamma_o = \left\{ \prod_{j=1}^{N} \phi_{j,\gamma_{kj}} : \sum_{j=1}^{N} \gamma_{kj} = o \right\} \tag{4.20}$$

If O is the maximum allowed combined polynomial order, then the second definition, the full set of O-th order PC basis vectors, is defined as:

$$\Gamma(O) = \bigcup_{o \in \{0,1,\ldots,O\}} \Gamma_o = \left\{ \prod_{j=1}^{N} \phi_{j,\gamma_{kj}} : \sum_{j=1}^{N} \gamma_{kj} \leq O \right\}. \tag{4.21}$$

By setting a value to the maximum allowed polynomial order, restrictions are made on the multi-indices γ_{kj} of the basis functions. Altogether, the expression in Equation 4.22 can be retrieved to relate the number of basis vectors, $P+1$, to the number of uncertain parameters, N, and the maximum polynomial order, O [32].

$$P + 1 = \frac{(N+O)!}{N!O!} \tag{4.22}$$

Figure 4.2 illustrates an example of a constructed PC basis set when considering two uncertain parameters, so $N = 2$, and the maximum combined polynomial order defined as $O = 6$. Each blue dot represents a basis vector $\Psi(\boldsymbol{\xi}) = He_i(\xi_1) \cdot He_j(\xi_2)$ with i and j the polynomial order for ξ_1 and ξ_2 respectively. In total 28 basis vectors are shown, which is in correspondence with Equation 4.22.

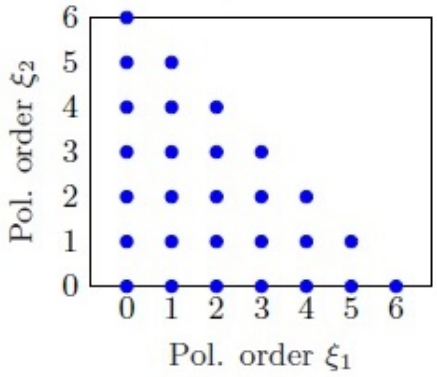


Figure 4.2: Example of a constructed PC basis set when considering two uncertain parameters, so $N = 2$, and the maximum combined polynomial order defined as $O = 6$. Each blue dot represents a basis vector $\Psi(\boldsymbol{\xi}) = He_i(\xi_1) \cdot He_j(\xi_2)$ with i and j the polynomial order for ξ_1 and ξ_2 respectively. In total 28 basis vectors are shown, which is in correspondence with Equation 4.22. Figure from [12].

4.4. NUMERICAL INTEGRATION

Equation 4.17 shows that in the formula for the PCE coefficients an expression of the stochastic response $R(\boldsymbol{\xi})$ is required, which, however, is still unknown. Therefore, to be able to calculate the inner product $\langle R(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi}) \rangle$ a numerical integration method is required. This section covers the relevant aspects of numerical integration methods for this research.

4.4.1. GAUSSIAN QUADRATURE RULE

For a one-dimensional problem, in which the uncertainty vector has only $N = 1$ component, a numerical integration method is called a quadrature rule. Examples are the trapezoidal rule, the Clenshaw-Curtis rule and the Gaussian quadrature rule. There are different versions of Gaussian quadrature rules, for example, the Gauss-Hermite and Gauss-Legendre quadrature rule. A beneficial characteristic of Gaussian quadrature rules is the following: if for an n-point quadrature rule the nodes $\{x_1, \ldots, x_n\}$ are chosen as the n roots of polynomial $p_n(x)$, with $p_n(x)$ being orthogonal to the corresponding weight function $\omega(x)$, then there exist weights $\{w_1, \ldots, w_n\}$ that yield an exact result for polynomials of degree $2n$–1, or less [37]. Equation 4.23 shows this result with $\omega(x)$ the PDF of a single uncertain parameter on the interval $[a, b]$.

$$\int_a^b p_n(x)\omega(x)dx = \sum_{i=1}^{n} w_i p_n(x_i) \tag{4.23}$$

Due to the assumption of Gaussian distributed uncertain parameters and the results in Table 4.1 and Equation 4.23, the Gauss-Hermite quadrature rule is chosen to be used in the PCE construction. However, the calculation will not be exact as the stochastic response is a general function, and not necessarily a polynomial. Equation 4.24 shows the numerical approximation of the function $f(\xi) = R(\xi) \cdot \Psi_k(\xi)$ in the case of a Gauss-Hermite quadrature rule.

$$I^{(1)} f(\xi) = \int_{-\infty}^{\infty} f(\xi) p_\xi(\xi) d\xi \approx Q_{lev}^{(1)} = \sum_{i=1}^{n_{lev}} w_{lev}^{(i)} f\left(\xi_{lev}^{(i)}\right) \tag{4.24}$$

The single dimensional integral approximation $Q_{lev}^{(1)}$ is called a quadrature. In Equation 4.24, the superscript (1) denotes the dimensionality of the integral, the superscripts (i) denote the different quadrature points, lev indicates the quadrature level and n_{lev} denotes the number of quadrature points in the integration rule according to quadrature level lev. A higher quadrature level results in a higher accuracy of the approximation, however, the function has to be evaluated on more quadrature points which makes the approximation computationally more expensive. For Gaussian quadrature rules in one dimension, the number of function evaluations is related to the quadrature level by the expression $n_{lev} = 2 \cdot lev - 1$. For higher dimensions, the number of function evaluations will grow exponentially. Due to the characteristic of Gaussian quadrature rules exact results are found for polynomial orders upto $2 \cdot n_{lev} - 1 = 4 \cdot lev - 3$.

A disadvantage of the Gauss-Hermite quadrature rule is that it has low nestedness, meaning that not all quadrature points of lower quadrature levels are also used in higher quadrature levels. In fact, only the origin is a repeating quadrature point for higher quadrature levels. As a counterexample, the Clenshaw-Curtis rule has full nestedness, meaning that each quadrature point in a certain quadrature level is part of the set of quadrature points of a higher quadrature level. Figure 4.3 shows the difference in quadrature grids for the first four quadrature levels of the Gauss-Hermite and Clenshaw-Curtis quadrature rules in one dimension.

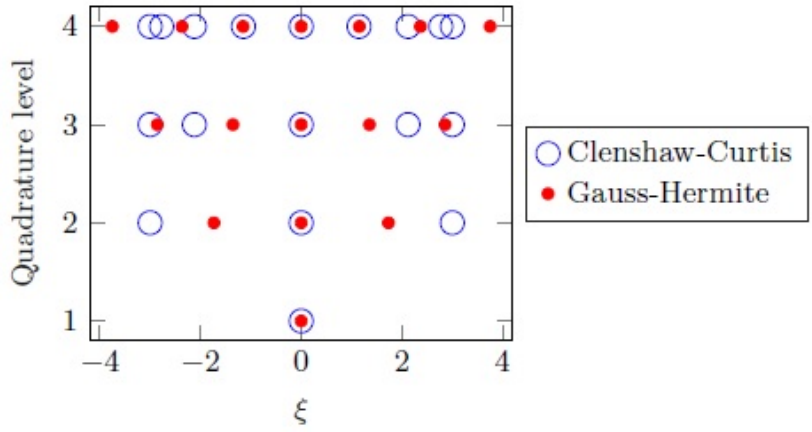


Figure 4.3: Quadrature grids for the first four quadrature levels of the Gauss-Hermite and Clenshaw-Curtis quadrature rules in one dimension. The quadrature points of the Clenshaw-Curtis rule are constructed on the interval [-3,3] and the quadrature points of the Gauss-Hermite rule are based on a standard Gaussian distribution as weight function. Figure from [12].

The full nestedness characteristic of the Clenshaw-Curtis rule could save computation time due to the possibility of re-using function evaluations for different quadrature levels, however, this rule is less accurate than the Gauss-Hermite rule [33]. Therefore, the Gauss-Hermite rule is chosen to be used in this research.

4.4.2. CUBATURE

For higher-dimensional problems, when an uncertainty vector with $N > 1$ parameters is considered, the integral that needs to be approximated is N-dimensional. Cubatures are the extension of quadratures for multidimensional integrals. The multidimensional integral can be approximated by tensorization of the corre-

sponding single dimensional quadratures, as Equation 4.25 shows the extension of Equation 4.24.

$$\begin{aligned} I^{(N)} f = \int_{\mathbb{R}^N} f(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \approx Q^{(N)}_{\boldsymbol{lev}} &= Q^{(1)}_{lev_1} \otimes \ldots \otimes Q^{(1)}_{lev_N} \\ &= \sum_{i_1=1}^{n_{lev_1}} \cdots \sum_{i_N=1}^{n_{lev_N}} w^{(i_1)}_{lev_1} \cdot \ldots \cdot w^{(i_N)}_{lev_N} \cdot f\left(\xi^{(i_1)}_{1,lev_1}, \ldots, \xi^{(i_N)}_{N,lev_N}\right) \\ &= \sum_{i=1}^{n_{total}} w^{(i)} f\left(\boldsymbol{\xi}^{(i)}\right) \end{aligned} \tag{4.25}$$

By extending to a multidimensional problem, a maximum quadrature level can be specified for each dimension. This is denoted by $\boldsymbol{lev} = (lev_1, \ldots, lev_N)^T$ and results in a different number of function evaluations per dimension. In Equation 4.25, the total number of cubature points is defined as $n_{total} = \prod_{j=1}^{N} n_{lev_j}$, the cubature points in the grid are denoted by $\boldsymbol{\xi}^{(i)}$ with index i and the weight corresponding to the cubature point $\boldsymbol{\xi}^{(i)}$ is defined as $w^{(i)} = \prod_{j=1}^{N} w^{(i)}_j$.

4.4.3. SMOLYAK SPARSE GRID

The cubature grid is defined by the chosen quadrature rule and the maximum quadrature level per dimension. Not all cubature points are equally important, and therefore, the number of cubature points can be decreased by omitting higher order quadrature levels in multiple dimensions. In this way, a sparse grid is created. This idea is based on the sparsity-of-effects principle, which states that for the responses in most models lower order terms are more important than higher order terms [36]. Such sparse grids can be used to alleviate computational cost, while keeping the numerical integration method accurate. Multiple types of sparse grids exist, for example, the (extended) Smolyak sparse grid [35].

Grids are created with the use of difference formulas, which are based on the difference between two quadratures. The difference formula is defined as:

$$\Delta^{(1)}_{lev} = Q^{(1)}_{lev} - Q^{(1)}_{lev-1}, \tag{4.26}$$

with $Q^{(1)}_0 = 0$. With the recurrence relation in Equation 4.26, the following expression for $Q^{(1)}_{lev}$ can be derived:

$$Q^{(1)}_{lev} = \sum_{l=1}^{lev} \Delta^{(1)}_l. \tag{4.27}$$

Equation 4.28 yields the result of substituting Equation 4.27 into Equation 4.25.

$$\begin{aligned} Q^{(N)}_{\boldsymbol{lev}} &= Q^{(1)}_{lev_1} \otimes \ldots \otimes Q^{(1)}_{lev_N} \\ &= \left(\sum_{l_1=1}^{lev_1} \Delta^{(1)}_{l_1}\right) \otimes \ldots \otimes \left(\sum_{l_N=1}^{lev_N} \Delta^{(1)}_{l_N}\right) \\ &= \sum_{\boldsymbol{l} \in \mathcal{G}} \Delta^{(N)}_{\boldsymbol{l}} \end{aligned} \tag{4.28}$$

In Equation 4.28 $\mathcal{G}$ denotes the cubature grid and $\boldsymbol{l} = (l_1, \ldots, l_N)^T$ is a multi-index with elements l_j denoting the quadrature level of the difference term in the j-th dimension. Equation 4.29 is used to define a full cubature grid, depending on the maximum quadrature level per dimension defined in $\boldsymbol{lev}$.

$$\mathcal{G}_{\text{Full}}(\boldsymbol{lev}) = \left\{\boldsymbol{l} : l_j \leq lev_j, \quad \forall\, j \in \{1, \ldots, N\}\right\} \tag{4.29}$$

For a Smolyak sparse grid, a maximum overall quadrature level is defined by the grid level L and a rule is applied which defines the permutations of quadrature levels to use. The Smolyak sparse grid is defined as:

$$\mathcal{G}_{\text{Sm}}(L) = \left\{\boldsymbol{l} : \sum_{j=1}^{N} l_j \leq L + N - 1\right\}. \tag{4.30}$$

Figure 4.4 shows an example of a full cubature grid and a Smolyak sparse grid in three dimensions.

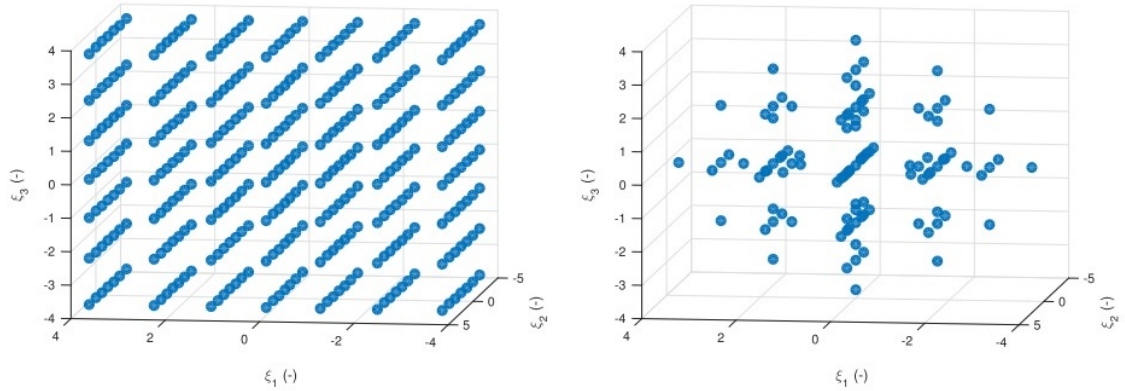


Figure 4.4: Two types of cubature grids in three dimensions. On the left a full grid with $\boldsymbol{lev} = (4,4,4)^T$ and on the right a Smolyak sparse grid with $L = 4$ is shown. Figure from [11].

4.4.4. EXTENDED SMOLYAK SPARSE GRID

Extended Smolyak sparse grids can be used to achieve a more accurate approximation of the single dimensional integrals. As an increase in the grid level is computationally expensive, with extended Smolyak sparse grids it is possible to increase the quadrature level in the single dimensional directions, while keeping the number of extra cubature points limited. The single dimensional grids $\text{Perm}(L,1,\ldots,1)$ for grid level L are replaced by grids $\text{Perm}(L+\Delta_{lev},1,\ldots,1)$, yielding only $2\cdot N\cdot\Delta_{lev}$ extra cubature points, with Δ_{lev} the increase in quadrature level of the single dimensional grids. Equation 4.31 shows the definition of an extended Smolyak sparse grid.

$$\mathcal{G}_{\text{ExtSm}}(L,\Delta_{lev}) = (\mathcal{G}_{\text{Sm}}(L) \cup \text{Perm}(L+\Delta_{lev},1,\ldots,1)) \setminus \text{Perm}(L,1,\ldots,1) \tag{4.31}$$

4.5. HYPERBOLIC TRIMMING

In Chapters 4.4.3 and 4.4.4, the creation of sparse grids is shown by omitting a part of the full cubature grid for high order quadrature levels. Similarly, the sparsity-of-effects principle can be used in the permutations of the polynomial orders of the PC basis functions, which are used to build the PC basis vectors. This concept is called hyperbolic trimming [38]. With hyperbolic trimming of the basis vectors only certain polynomial orders of the basis functions are used in the construction of the basis vectors that meet the following requirement:

$$||\gamma_k||_q \leq O, \tag{4.32}$$

in which O is the maximum combined polynomial order and $||\gamma_k||_q$ is the q-quasi-norm of the multi-indices of the basis functions, γ_{kj}, defined as:

$$||\gamma_k||_q = \left(\sum_{j=1}^{N} \gamma_{kj}^q\right)^{\frac{1}{q}}, \tag{4.33}$$

with $q \in (0,1]$. In the case of $q = 1$, the full O-th order basis set is retrieved, as is shown in Figure 4.2. Figure 4.5 shows an example of hyperbolic trimming on the polynomial orders for a two-dimensional case.

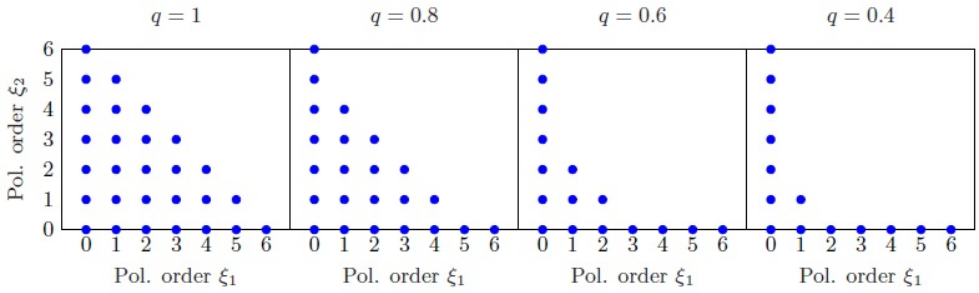


Figure 4.5: Illustration of hyperbolic trimming of a PC basis set with maximum combined polynomial order $O = 6$ in two dimensions for different values of the trim factor q. The blue dots are used to illustrate which polynomial orders are in the PC basis set for the different basis vectors. Figure from [12].

4.6. CONSTRUCTION OF PCE

The construction and evaluation of PCEs will be performed with the OpenGPC package in Matlab, coded by Z. Perkó et al. [25]. This package makes use of all the theory described in Chapter 4.

4.6.1. OPENGPC

The OpenGPC package consists of several scripts that together perform the PCE construction. Figure 4.6 illustrates the different steps in this process. First, input settings have to be supplied. These settings are the characteristics of the uncertain parameters, the type and order of the polynomial basis functions, the grid level and possible extra quadrature levels in the single dimensional directions, the trim factor and the plan file, containing information on the object of which a PCE is constructed and how the response values of this object are calculated on the cubature points. A cut-off value can be introduced to save memory and computation time in the PCE construction by omitting response values that are below the cut-off value.

Then, a check is performed on the PCE settings and the cubature grid is generated. Next, the PCE object is initialized by construction of the PC basis set and the dose mask is calculated, which is a logical vector that shows which response values are greater than or equal to the cut-off value. For these calculations the dose engine is needed. From the dose mask the response values on the cubature grid are retrieved and the PCE coefficients are approximated with a Gaussian quadrature rule. Altogether, this concludes the PCE construction.

In the end, the constructed polynomial expansion can be evaluated on or sampled over the sample space of the uncertain parameters, the mean and variance of the stochastic response can be retrieved and sensitivity analysis can be performed. Also, the derivative of the response with respect to the uncertain parameters is available.

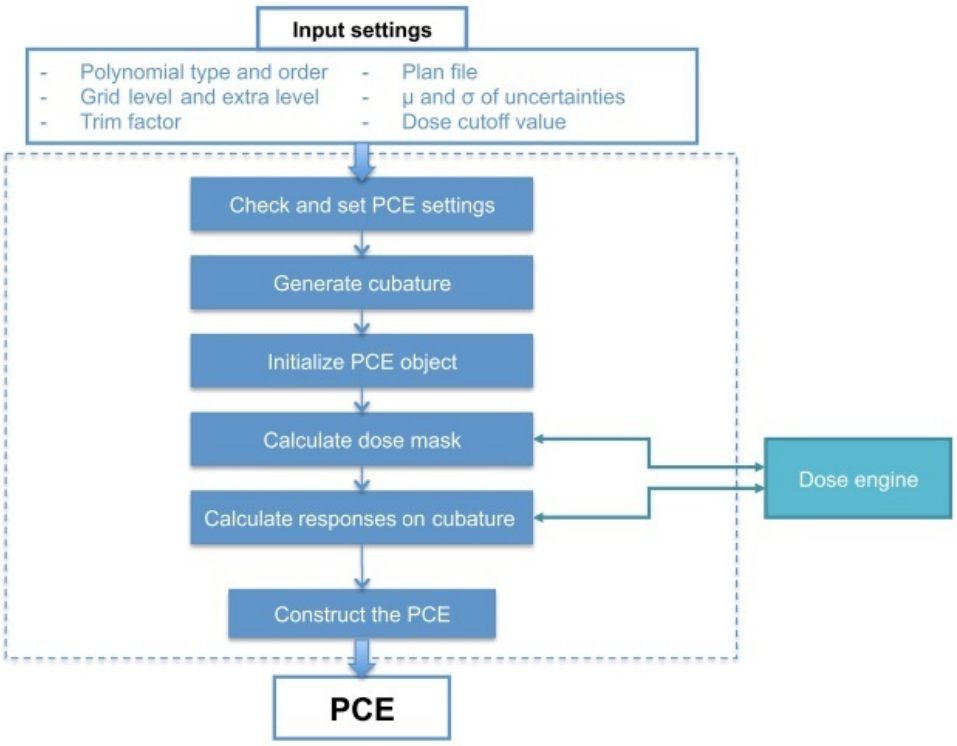


Figure 4.6: Flowchart illustration of the construction of a PCE with the OpenGPC package. Modified figure from [9].

4.6.2. NOTATION

The notation $LxEyOz$, introduced in [12], is used to summarize the main characteristics of a constructed PCE. This notation describes a Smolyak sparse grid with grid level $L = x$, increased single dimensional quadrature level $\Delta_{lev} = y$ and full $O = z$-th order PC basis set. Note that with a full PC basis set no hyperbolic trimming is introduced.

5

PROBABILISTIC FUNCTIONS WITH POLYNOMIAL CHAOS EXPANSION

In probabilistic treatment planning, statistical measures of stochastic dose parameters are used in objective functions and constraints. Any function of a dose parameter, which in turn is dependent on the uncertain parameters, can be used for probabilistic planning. As discussed in Chapter 3.4.3, such a function can be denoted by $g(\boldsymbol{x},\boldsymbol{\xi})$ and choices have to be made about which statistical measure to optimize for and whether the function has to be minimized or maximized. In the end, this boils down to optimizing the function $f\left(\mathcal{M}_{\boldsymbol{\xi}}[g(\boldsymbol{x},\boldsymbol{\xi})]\right)$, as shown in Equation 3.9.

Polynomial Chaos Expansion is used to describe the functional dependence of such probabilistic functions on the uncertain parameters. Obtaining many samples of the dose deposition matrix in different error scenarios from the dose engine and calculating statistical measures with these samples is computationally very expensive. PCE is used to alleviate the computational cost in order to facilitate a faster evaluation of the statistical measures. However, a choice has to be made whether a PCE is constructed for the function $g(\boldsymbol{x},\boldsymbol{\xi})$ as a whole or for a part of this function. Constructing a PCE for the function $g(\boldsymbol{x},\boldsymbol{\xi})$ means that the beam intensity vector is contained in the PCE and thus a PCE has to be constructed in every iteration of the treatment plan optimization, which is not desired. On the other hand, evaluation of statistical measures could be performed relatively fast. Thus, it depends on the probabilistic function which object is most convenient to build a PCE for.

This chapter describes how PCE can be used to approximate several probabilistic functions, and its gradient and Hessian, which are used in this research. In [12], more examples of probabilistic functions described with PCE can be found, as well as an explanation of the use of PCE within probabilistic functions for multiple treatment fractions. However, in the rest of this research the concept of fractionation is not covered.

5.1. EXPECTED QUADRATIC DOSE DIFFERENCE

The expected quadratic dose difference, which is used as an abbreviation for the expected value of the sum of the quadratic differences between the delivered dose and the prescribed dose in a domain, is found to be a suitable objective function for probabilistic treatment planning with PCE for proton therapy [12]. Optimization is performed by minimizing this probabilistic function to find the suitable beam intensity vector $\boldsymbol{x}$ for the treatment plan. The objective function can be denoted as $f(\boldsymbol{x})$. The beamspot indices are $j \in \mathcal{B} = \{1,\ldots,N_b\}$ with N_b the number of beamspots. The domain is discretized to contain N_v voxels with voxel indices $i \in \mathcal{V} = \{1,\ldots,N_v\}$. Each voxel has a corresponding prescribed dose, contained in the vector $\boldsymbol{d}^P$, and also each voxel is assigned a weight, contained in the vector $\boldsymbol{w}$, such that certain voxels are considered more important in the optimization to satisfy the prescribed dose. Therefore, the voxels in the GTV and OARs are given a relatively high weight, whereas the tissue voxels are given a relatively low weight. The expected quadratic dose difference can, with substitution of Equation 3.4 at the second equal-sign, be described as follows:

$$
\begin{aligned}
f(\boldsymbol{x}) &= \int_{\mathbb{R}^N} \sum_{i\in\mathcal{V}} w_i \left(d_i(\boldsymbol{x},\boldsymbol{\xi}) - d_i^P\right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&= \int_{\mathbb{R}^N} \sum_{i\in\mathcal{V}} w_i \left(\sum_{j\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j - d_i^P\right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&= \int_{\mathbb{R}^N} \sum_{i\in\mathcal{V}} w_i \left(\sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}) D_{ij'}(\boldsymbol{\xi}) x_j x_{j'} - 2 d_i^P \sum_{j\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j + \left(d_i^P\right)^2\right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&= \int_{\mathbb{R}^N} \left(\sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}} \left[\sum_{i\in\mathcal{V}} w_i D_{ij}(\boldsymbol{\xi}) D_{ij'}(\boldsymbol{\xi})\right] x_j x_{j'} - 2 \sum_{j\in\mathcal{B}} \left[\sum_{i\in\mathcal{V}} w_i d_i^P D_{ij}(\boldsymbol{\xi})\right] x_j + \sum_{i\in\mathcal{V}} w_i \left(d_i^P\right)^2\right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}.
\end{aligned}
\tag{5.1}
$$

Next, PCEs can be constructed of both expressions in between the square brackets in Equation 5.1. These are formulated as in Equation 5.2 and 5.3 with $P+1$ the number of PC basis vectors.

$$
\sum_{i\in\mathcal{V}} w_i D_{ij}(\boldsymbol{\xi}) D_{ij'}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} R_{jj'}^{(k)} \Psi_k(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} R_{jj'}^{(k)} \Psi_k(\boldsymbol{\xi}) \tag{5.2}
$$

$$
\sum_{i\in\mathcal{V}} w_i d_i^P D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \tag{5.3}
$$

The first PCE consists of the matrix $R^{(k)} \in \mathbb{R}^{N_b \times N_b}$, which contains the k-th order PCE coefficients corresponding to the expression at the left-hand side of Equation 5.2. The indices j and j' are interchangeable, from which we can conclude that the matrix $R^{(k)}$ is symmetric. Therefore, we only have to compute the upper, or lower, triangular part of the matrix, containing $\frac{1}{2}\left(N_b + N_b^2\right)$ elements. The second PCE consists of the vector $\boldsymbol{r}^{(k)} \in \mathbb{R}^{N_b}$, containing the k-th order PCE coefficients corresponding to the expression at the left-hand side of Equation 5.3. Substituting these two PCEs into Equation 5.1, yields:

$$
\begin{aligned}
f(\boldsymbol{x}) &= \int_{\mathbb{R}^N} \left(\sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}} \left[\sum_{k=0}^{\infty} R_{jj'}^{(k)} \Psi_k(\boldsymbol{\xi})\right] x_j x_{j'} - 2 \sum_{j\in\mathcal{B}} \left[\sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi})\right] x_j + \sum_{i\in\mathcal{V}} w_i \left(d_i^P\right)^2\right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&= \sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}} \left(\sum_{k=0}^{\infty} \int_{\mathbb{R}^N} R_{jj'}^{(k)} \Psi_0(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}\right) x_j x_{j'} \\
&\quad - 2 \sum_{j\in\mathcal{B}} \left(\sum_{k=0}^{\infty} \int_{\mathbb{R}^N} r_j^{(k)} \Psi_0(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}\right) x_j + \sum_{i\in\mathcal{V}} w_i \left(d_i^P\right)^2 \int_{\mathbb{R}^N} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&= h_0^2 \sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}} R_{jj'}^{(0)} x_j x_{j'} - 2 h_0^2 \sum_{j\in\mathcal{B}} r_j^{(0)} x_j + \sum_{i\in\mathcal{V}} w_i \left(d_i^P\right)^2 \\
&= \sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}} R_{jj'}^{(0)} x_j x_{j'} - 2 \sum_{j\in\mathcal{B}} r_j^{(0)} x_j + \sum_{i\in\mathcal{V}} w_i \left(d_i^P\right)^2.
\end{aligned}
\tag{5.4}
$$

In the derivation of Equation 5.4 similar steps as in Equation 4.19 are used with $\Psi_0(\boldsymbol{\xi}) = 1$ the zeroth order PC basis vector with norm $h_0^2 = 1$. The derivative of this probabilistic function $f(\boldsymbol{x})$ with respect to an element x_n of the beam intensity vector $\boldsymbol{x}$ results in:

$$
\frac{\partial f(\boldsymbol{x})}{\partial x_n} = \sum_{j\in\mathcal{B}} R_{jn}^{(0)} x_j + \sum_{j'\in\mathcal{B}} R_{nj'}^{(0)} x_{j'} - 2 r_n^{(0)} = 2\left(\sum_{j\in\mathcal{B}} R_{jn}^{(0)} x_j - r_n^{(0)}\right), \quad \forall n \in \{1,\ldots,N_b\}, \tag{5.5}
$$

using the symmetry of the matrix $R^{(0)}$ at the second equal-sign. The elements of the Hessian matrix can be derived as:

$$
\frac{\partial^2 f(\boldsymbol{x})}{\partial x_m \partial x_n} = 2 R_{mn}^{(0)}, \quad \forall m,n \in \{1,\ldots,N_b\}. \tag{5.6}
$$

Altogether, Equations 5.4-5.6 show that in the expression of this probabilistic function, and its gradient and Hessian with respect to the beam intensity vector, only the zeroth order PCE coefficients are required.

5.2. EXPECTED MEAN DOSE

The expected mean dose, equivalent to the mean expected dose through linearity, in a discretized structure with $\mathcal{V}_{str} \subseteq \mathcal{V}$ the set of voxel indices and $N_{v,str} \leq N_v$ the number of voxels is another example of a probabilistic function to be used in treatment planning. In this case, the function $f(\boldsymbol{x})$ is written as:

$$\begin{aligned} f(\boldsymbol{x}) &= \int_{\mathbb{R}^N} \left(\frac{1}{N_{v,str}} \sum_{i \in \mathcal{V}_{str}} d_i(\boldsymbol{x}, \boldsymbol{\xi}) \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^N} \frac{1}{N_{v,str}} \sum_{i \in \mathcal{V}_{str}} \left(\sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \frac{1}{N_{v,str}} \sum_{j \in \mathcal{B}} x_j \int_{\mathbb{R}^N} \sum_{i \in \mathcal{V}_{str}} D_{ij}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \end{aligned} \tag{5.7}$$

with substitution of Equation 3.4 at the second equal-sign. A PCE can be constructed of the row sums of the dose deposition matrix as follows:

$$\sum_{i \in \mathcal{V}_{str}} D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} r_j^{(k)} \Psi_k(\boldsymbol{\xi}). \tag{5.8}$$

The vector $\boldsymbol{r}^{(k)} \in \mathbb{R}^{N_b}$ contains the k-th order PCE coefficients. Note that these are not the same PCE coefficients as in Equation 5.3. Now, Equation 5.8 can be substituted into Equation 5.7, yielding:

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{N_{v,str}} \sum_{j \in \mathcal{B}} x_j \int_{\mathbb{R}^N} \left(\sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \frac{1}{N_{v,str}} \sum_{j \in \mathcal{B}} x_j \sum_{k=0}^{\infty} r_j^{(k)} \int_{\mathbb{R}^N} \Psi_0(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \frac{1}{N_{v,str}} h_0^2 \sum_{j \in \mathcal{B}} r_j^{(0)} x_j \\ &= \frac{1}{N_{v,str}} \sum_{j \in \mathcal{B}} r_j^{(0)} x_j, \end{aligned} \tag{5.9}$$

in which similar derivation steps as in Equation 4.18 are used with $\Psi_0(\boldsymbol{\xi}) = 1$ the zeroth order PC basis vector with norm $h_0^2 = 1$. The derivative of this probabilistic function with respect to an element x_n of the beam intensity vector can easily be derived as:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_n} = \frac{1}{N_{v,str}} r_n^{(0)}, \quad \forall n \in \{1, \ldots, N_b\}. \tag{5.10}$$

This directly shows that the Hessian matrix for this probabilistic function is equal to the zero matrix.

5.3. PROBABILISTIC DOSE VOLUME PARAMETER

In radiotherapy it is desired to have information about how much dose is received by a certain volume fraction of a structure. A dose volume parameter can be used to describe such information. The dose volume parameter $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ is defined as the maximum dose that is received by at least $\alpha\%$ of the considered structure, and is denoted as shown in Equation 3.6. In the optimization of a probabilistic treatment plan objective goals and constraints could be set on percentiles of dose volume parameters for the tumor and organs-at-risk. This is desirable in order to obtain knowledge about the behavior of the dose volume parameter in different error scenarios. Such information is captured in probabilistic dose volume parameters. An example of this parameter is the $\beta\%$-probabilistic $d_{\alpha\%}(\boldsymbol{x})$, being defined as the $(100-\beta)$-th percentile of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$, and is in equations in short denoted as $d_{\alpha\%}^{\beta\%}(\boldsymbol{x})$ for a certain structure. Mathematically, this is written as:

$$d_{\alpha\%}^{\beta\%}(\boldsymbol{x}) = P^s_{(100-\beta)\%} \left\{ P^i_{(100-\alpha)\%} \left\{ d_i\left(\boldsymbol{\xi}^s\right) : i \in \mathcal{V}_{str} \right\} : s \in \mathcal{S} \right\}. \tag{5.11}$$

Two important probabilistic dose volume parameters are the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$, which is called the probabilistic near-minimum dose, and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$, also known as the probabilistic near-maximum dose.

The uncertain parameters are assumed to be Gaussian distributed, which means that the number of possible error scenarios is infinite as the Gaussian distribution is a continuous distribution. The β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ cannot be expressed analytically, and therefore, many error scenarios need to be sampled to accurately calculate this parameter. This means that the gradient and Hessian of such a probabilistic function cannot be derived analytically and the use of a finite difference scheme is needed. In [12] it was concluded that the use of a forward difference scheme for these derivatives is preferred. The approximation of an element of the gradient with a forward finite difference scheme with step size h and $\hat{\boldsymbol{e}}_n$ the unit vector along the n-th dimension of the beam intensity vector is performed as:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_n} \approx \frac{f(\boldsymbol{x}+h\hat{\boldsymbol{e}}_n)-f(\boldsymbol{x})}{h}, \quad \forall n \in \{1,\dots,N_b\}. \tag{5.12}$$

The elements of the Hessian matrix are approximated similarly, using the definition of the gradient in Equation 5.12.

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_m \partial x_n} \approx \frac{f(\boldsymbol{x}+h\hat{\boldsymbol{e}}_m+h\hat{\boldsymbol{e}}_n)-f(\boldsymbol{x}+h\hat{\boldsymbol{e}}_m)-f(\boldsymbol{x}+h\hat{\boldsymbol{e}}_n)+f(\boldsymbol{x})}{h^2}, \quad \forall m,n \in \{1,\dots,N_b\}. \tag{5.13}$$

A choice has to be made on how the probabilistic function $f(\boldsymbol{x}) = d^{\beta\%}_{\alpha\%}(\boldsymbol{x})$ is computed with the use of PCE. In Chapters 5.3.1 and 5.3.2, two different approaches are discussed, namely through a PCE of the dose deposition matrix $D(\boldsymbol{\xi})$ and through a PCE of the dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ respectively. Chapter 5.3.3 explains which approach is used in the rest of this research.

5.3.1. PCE OF DOSE DEPOSITION MATRIX

The first approach is to construct PCEs of the elements of the dose deposition matrix as follows:

$$D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} R^{(k)}_{ij}\Psi_k(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} R^{(k)}_{ij}\Psi_k(\boldsymbol{\xi}), \tag{5.14}$$

with the matrix $R^{(k)} \in \mathbb{R}^{N_b \times N_b}$ containing the k-th order PCE coefficients of the elements of the dose deposition matrix. Next, these PCEs need to be sampled N_s times following the underlying Gaussian sample space. From this the set $\{D(\boldsymbol{\xi}^1),\dots,D(\boldsymbol{\xi}^{N_s})\}$ of sample dose deposition matrices is obtained. Subsequently, the set of corresponding $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}^s)$ values, $\{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}^1),\dots,d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}^{N_s})\}$, is computed as:

$$d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}^s) = P^i_{(100-\alpha)\%}\left\{\sum_{j\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}^s)\,x_j : i \in \mathcal{V}_{str}\right\}. \tag{5.15}$$

The probabilistic function $f(\boldsymbol{x}) = d^{\beta\%}_{\alpha\%}(\boldsymbol{x})$ is then computed as:

$$f(\boldsymbol{x}) = P^s_{(100-\beta)\%}\left\{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}^s) : s \in \mathcal{S}\right\}. \tag{5.16}$$

The forward finite difference scheme can be evaluated with Equation 5.16, in which the beam intensity vector could straightforwardly be increased in the single-dimensional beam intensity directions as these intensities are not contained in the PCEs. Therefore, for the whole optimization all $N_v \times N_b$ PCEs only need to be constructed once. The disadvantage is that sampling is less practical as it requires sampling the PCEs N_s times and calculating the percentile of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ from all these samples. This requires storage of N_s matrices of size $N_v \times N_b$ for all samples of the dose deposition matrix. This quickly becomes too much to handle computationally.

5.3.2. PCE OF DOSE VOLUME PARAMETER

The second approach is to construct a PCE of the dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ itself. As the beam intensities are needed to calculate the dose, a PCE has to be constructed for a certain configuration of beam intensities. This means that the following process has to be repeated for each iteration of the treatment plan optimization. A PCE is constructed of the dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$, yielding:

$$d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{k=0}^{\infty} r^{(k)}(\boldsymbol{x})\Psi_k(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} r^{(k)}(\boldsymbol{x})\Psi_k(\boldsymbol{\xi}), \tag{5.17}$$

with the scalar $r^{(k)}(\boldsymbol{x}) \in \mathbb{R}$ the k-th order PCE coefficient. This PCE can be used to sample the dose volume parameter over many error scenarios, following the underlying Gaussian sample space, obtaining the set $\left\{d_{\alpha\%}\left(\boldsymbol{x},\boldsymbol{\xi}^1\right),\ldots,d_{\alpha\%}\left(\boldsymbol{x},\boldsymbol{\xi}^{N_s}\right)\right\}$. The probabilistic function $f(\boldsymbol{x}) = d^{\beta\%}_{\alpha\%}(\boldsymbol{x})$ is then calculated as:

$$f(\boldsymbol{x}) = P^s_{(100-\beta)\%}\left\{d_{\alpha\%}\left(\boldsymbol{x},\boldsymbol{\xi}^s\right) : s \in \mathcal{S}\right\}. \tag{5.18}$$

Next, Equations 5.12 and 5.13 are used to compute the gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ with respect to the beam intensities. However, as the dependence of this probabilistic function on the beam intensities is unknown, PCEs also need to be constructed for the beam intensity vector plus all the different single and double step size increases of the elements of the beam intensity vector in the finite difference scheme.

The preference of a forward finite difference scheme is due to a minimal number of extra PCEs that is needed, and also because previous research showed a close to linear behavior of the probabilistic near-minimum dose on the beam intensities when a relatively large step size is used [12]. Equation 5.12 shows that for the gradient N_b extra PCEs need to be constructed. For the Hessian matrix this would amount to N_b^2 extra PCEs, but due to the symmetry of the Hessian matrix this can be reduced to $\frac{1}{2}\left(N_b + N_b^2\right)$ extra PCEs. In total one would need $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ PCEs per iteration in the treatment plan optimization. An advantage of this approach is that the PCEs can be evaluated relatively fast for many samples of error scenarios without significant problems with computer memory.

5.3.3. CHOICE OF APPROACH

In theory, both approaches can be used to calculate the value, gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ as probabilistic function if there is enough storage memory to save and process all the N_s samples of either the $N_v \cdot N_b$ or the $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$, which is approximately equal to $\frac{1}{2}N_b^2$ for large enough N_b, PCEs for the first and second approach respectively. The number of voxels is in treatment plan optimization usually larger than the number of beamspots ($N_v > N_b$), so the approach in Chapter 5.3.1 with PCEs of the elements of the dose deposition matrix $D_{ij}(\boldsymbol{\xi})$ does become problematic faster. Therefore, the approach with PCEs of the dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ is used in this research, as has also been used in [12].

6

IMPLEMENTATIONS

In the preceding chapters, relevant theory and probabilistic functions with PCE related to this research have been discussed. This chapter focuses on three implementations that are specific to this research. In Chapter 6.1, the methodology of treatment plan optimization on a simplified three-dimensional geometry with a tumor and an OAR is explained. Chapter 6.2 treats the implementation of the algorithm that is used to calculate the value, gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$. Chapter 6.3 elaborates on a constraint on the PCE coefficients that is required for monotonicity of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ with respect to the beam intensity vector.

6.1. SIMPLIFIED 3D GEOMETRY

In order to test new ideas for proton therapy and to avoid the complications of treatment planning for a real patient in iCycle, a set of matlab scripts has been created by Z. Perkó. These scripts construct a simplified three-dimensional geometry of a patient with a tumor and an OAR. The dose calculations for this geometry are performed according to a pencil beam algorithm, based on [42]. This section describes how this geometry is used in this research and is divided in five parts. In Chapters 6.1.1-6.1.5, the geometry itself, the dose calculations, the considered uncertain parameters, treatment planning in Matlab and the purpose of the different optimizations are discussed respectively.

6.1.1. GEOMETRY

The geometry is a representation of a patient with a tumor at the abdominal side of the spine. The spine is considered an OAR as too much dose to it could result in serious health problems for the patient. The domain is defined in the three dimensions of Euclidean space as $x \in [-40, 40]$, $y \in [0, 10]$ and $z \in [85, 125]$, all units being in millimeters. The spine is a circularly shaped cylinder and the tumor is shaped as a half ring around the spine. A cross-section of this geometry with the location of the spine (OAR) and the tumor is shown in Figure 6.1.

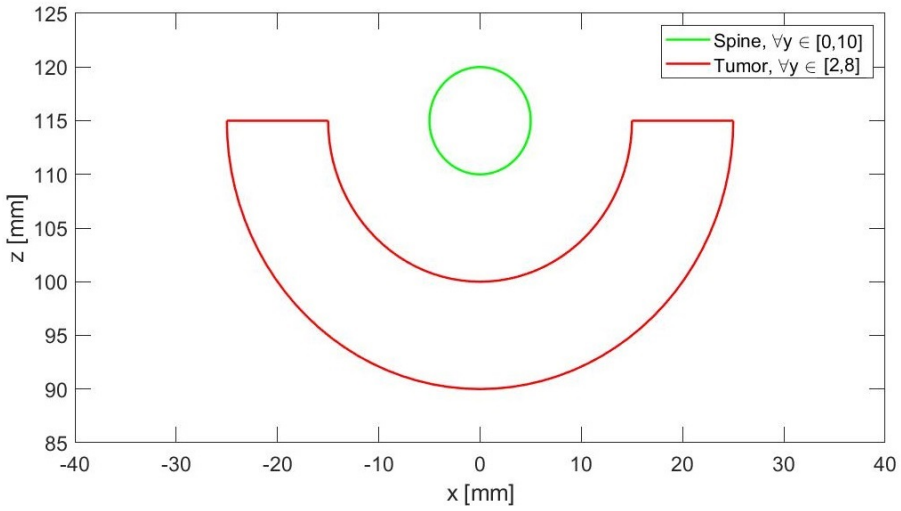


Figure 6.1: Axial cross-section of the simplified three-dimensional geometry, in which the tumor (edge shown in red) is located at the abdominal side of the spine (edge shown in green). The rest of the geometry is healthy tissue. This cross-section holds for $y \in [2,8]$ in millimeters. For values of $y \in [0,2)$ and $y \in (8,10]$ only the spine is present, and not the tumor.

The domain is discretized to create a voxel grid of cubes with equal length sides of 2 mm. This results in $N_{v,x} = 40$, $N_{v,y} = 5$ and $N_{v,z} = 20$ voxels along the x-, y- and z-direction respectively, resulting in a total of $N_v = N_{v,x} \cdot N_{v,y} \cdot N_{v,z} = 4000$ voxels in the domain.

6.1.2. DOSE CALCULATIONS

Proton beams are set up such that they enter the body at $z = 0$ and move in the positive z-direction. The configuration of proton beams is used to form an equidistant beamspot grid, consisting of multiple different energy levels within each proton beam. For a certain energy level, all beams have their Bragg peak at the same depth in the z-direction and the beams are placed in an equidistant grid in the (x, y)-plane with lateral spacings of 3 mm. This same grid is repeated for the different energy levels, the energy levels being chosen such that the Bragg peak spacings in the z-direction are also 3 mm. The center of the beamspot grid is called the isocenter. For this geometry the isocenter is located at $(x, y, z) = (0, 5, 105)$ in millimeters. The number of beamspots along each direction is $N_{b,x} = 26$, $N_{b,y} = 3$ and $N_{b,z} = 12$, yielding a total of $N_b = N_{b,x} \cdot N_{b,y} \cdot N_{b,z} = 936$ beamspots in the domain. Concluding, the beamspot grid with beamspot coordinates (x_i, y_j, z_k) is defined as:

$$\begin{cases} x_i = -40.5 + 3 \cdot i, & \forall i \in \{1, \ldots, 26\}, \\ y_j = -1 + 3 \cdot j, & \forall j \in \{1, \ldots, 3\}, \\ z_k = 85.5 + 3 \cdot k, & \forall k \in \{1, \ldots, 12\}. \end{cases} \tag{6.1}$$

The dose delivered by the protons of a single energy level in a certain proton beam is, as stated in Chapter 2.1.1, not a deterministic, but a stochastic process. The individual protons undergo different interactions with matter, resulting in different energy deposition profiles along the trajectory through the body. Therefore, the dose of a certain beamspot is not a single dose peak at location (x_i, y_j, z_k), but the dose spread in the (x, y)-plane is assumed to be Gaussian distributed with mean (x_i, y_j, z_k) and the variance still to be defined. The spatial energy deposition profile in the z-direction is following the Bragg peak characteristic, as described in Chapter 1.2.1. In the pencil beam algorithm the range straggling, the standard deviation of this Gaussian lateral dose spread at entrance of the body, is set to 3 mm, as used in previous research [12]. The variance increases over depth as there are more possible interactions between protons and matter when considering a longer path length. The pencil beam algorithm takes this into account in the dose calculations.

In contrast to a real patient, in this geometry the spine, the tumor and the surrounding tissue are assumed to have the same homogeneous density. Furthermore, the use of a discretized voxel grid requires conservation of dose to be imposed in each voxel. This is achieved by calculating the mean dose in a voxel with a three-dimensional Gauss-Legendre quadrature rule, based on 8 points, and assigning this mean dose to the center of the voxel.

The optimized beam intensity vectors are used to evaluate both the nominal and the expected dose distribution in the geometry. The nominal dose distribution is obtained by the matrix-vector multiplication of the dose deposition matrix in the nominal scenario and the beam intensity vector. The expected dose distribution, on the other hand, is obtained by the matrix-vector multiplication of the expected value of the dose deposition matrix and the beam intensity vector. This requires the construction of PCEs of the elements of the dose deposition matrix, of which only the zeroth order PCE coefficients are needed as we are interested in the expected value of the dose deposition matrix.

6.1.3. UNCERTAIN PARAMETERS

Uncertainties in the setup of the geometry or the range of the proton beams are yet to be introduced. Set-up errors can be realized by shifting the location of the isocenter. Range errors are usually implemented as a scaling of the values of the CT images. However, as the densities in this simplified 3D geometry are considered to be homogeneous and constant, this is not possible. Range errors can be realized in the pencil beam algorithm by introducing a relative or absolute error on the range of the pencil beams, the realizations being the same for each pencil beam.

In the implementation of this geometry only a systematic set-up uncertainty in the x-direction is considered. This set-up uncertainty is assumed to be Gaussian distributed with a zero mean and a standard deviation of 3 mm.

6.1.4. TREATMENT PLAN OPTIMIZATION

The optimizations on the 3D spine geometry are divided into two parts. The first part of the optimizations is based on the probabilistic functions in Chapters 5.1 and 5.2. Here, two optimizations are considered. In both optimizations the expected value of the sum of the quadratic differences between the prescribed dose and the delivered dose in the full discretized geometry is minimized, and in the one of the two optimizations also a constraint is set on the expected mean dose in the tumor to be greater than or equal to the prescribed dose in the tumor. These two optimizations coincide with cases 1 and 2 in Table 6.1.

The second part of the optimizations is based on the probabilistic functions in Chapters 5.1 and 5.3. Several different optimizations are performed, containing the same objective function as the two optimizations in the first part, but this time a constraint is set on the probabilistic near-minimum dose, the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$, in the tumor to be greater than or equal to a still to be defined value. This optimization problem coincides with case 3 in Table 6.1.

Table 6.1: The three different cases of optimization problems that are considered in this research. For all cases also a constraint is set on the elements of the beam intensity vector to be non-negative, $x_i \geq 0$ for $i \in \{1,\dots,N_b\}$.

Case:	Objective function:	Constraint:
1	Min. $\int_{\mathbb{R}^N} \sum_{i\in\mathcal{V}} w_i \left(d_i(\boldsymbol{x},\boldsymbol{\xi}) - d_i^P\right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$	-
2	Min. $\int_{\mathbb{R}^N} \sum_{i\in\mathcal{V}} w_i \left(d_i(\boldsymbol{x},\boldsymbol{\xi}) - d_i^P\right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$	$\int_{\mathbb{R}^N} \left(\frac{1}{N_{v,tumor}} \sum_{i\in\mathcal{V}_{tumor}} d_i(\boldsymbol{x},\boldsymbol{\xi})\right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \geq d_{tumor}^P$
3	Min. $\int_{\mathbb{R}^N} \sum_{i\in\mathcal{V}} w_i \left(d_i(\boldsymbol{x},\boldsymbol{\xi}) - d_i^P\right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$	$d_{98\%}^{95\%}(\boldsymbol{x}) \geq \dots$

The difference between the different optimizations of case 3 is that for the calculation of the value, gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ different methodologies are used. These methodologies are divided into an exact calculation and a proposed approximation method for the gradient and Hessian, to alleviate the computation time of the optimization. The exact calculation is explained in detail in Chapter 6.2, and several improvements on this algorithm and the proposed approximation method are discussed in Chapter 7.

The optimizations of the beam intensity vector for these treatment plans are performed with the *fmincon* solver in Matlab. This solver is able to find the minimum of a constrained nonlinear multivariate function. Besides the above mentioned objective functions and constraints, a constraint is set on the elements of the beam intensity vector as they are physically constrained to be non-negative. The *interior-point* algorithm is chosen within the *fmincon* solver, as it is the default optimization algorithm, and is treated as a black box. The working of this optimization algorithm will not be discussed here, but detailed information can be found in [43, 44].

The solver is able to approximate the gradient and Hessian matrix of objective functions and constraints through a finite difference scheme, however, as for cases 1 and 2 the gradient and Hessian matrix are analyti-

cally available through the use of PCEs, both will be supplied to the solver. This will improve the convergence speed and accuracy of the optimization, and it will decrease computational cost. For case 3 the accuracy and computation time of the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ is investigated, so these will also be computed and supplied to the solver.

6.1.5. PURPOSE OF DIFFERENT OPTIMIZATIONS

The two optimization problems in cases 1 and 2, without the use of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$, are used for two purposes. The first purpose is to visualize the outcome of probabilistic treatment plans with the use of PCE, of which case 1 is also treated in [12]. The second purpose is to use the optimized beam intensity vectors to test the approximation method that is proposed for the alleviation of computational cost in the calculation of the gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$. The different optimizations for case 3 are intended to compare the performance and computation time of the exact method and the approximation method.

6.2. ALGORITHM FOR VALUE, GRADIENT AND HESSIAN OF PROBABILISTIC DOSE VOLUME PARAMETER

The algorithm for the calculation of the value, gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$, as used in optimizations with patient data in iCycle in [12], works as follows. For the optimization a certain starting vector $\boldsymbol{x}_0 \in \mathbb{R}^{N_b \times 1}$ must be chosen for the beam intensity vector. This beam intensity vector is then expanded by a vertical concatenation with perturbed versions of the beam intensity vector. The concatenated columns are the beam intensity vector plus all the single, for the gradient, and the single and double, for the Hessian, step size increases of the individual beam intensities as are present in the forward finite difference schemes in Equations 5.12 and 5.13. When the gradient is calculated, this expanded matrix X_{grad} looks like:

$$X_{grad} = \begin{bmatrix} \boldsymbol{x}_0 & \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_1 & \dots & \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_{N_b} \end{bmatrix} \in \mathbb{R}^{(N_b)\times(1+N_b)}, \tag{6.2}$$

and when both the gradient and Hessian are calculated, the expanded matrix X_{Hess} looks like:

$$X_{Hess} = \begin{bmatrix} \boldsymbol{x}_0 & \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_1 & \dots & \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_{N_b} & \boldsymbol{x}_0 + 2h\hat{\boldsymbol{e}}_1 & \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_1 + h\hat{\boldsymbol{e}}_2 & \dots & \boldsymbol{x}_0 + 2h\hat{\boldsymbol{e}}_{N_b} \end{bmatrix}, \tag{6.3}$$

with $X_{Hess} \in \mathbb{R}^{(N_b)\times\left(1+N_b+\frac{1}{2}\left(N_b+N_b^2\right)\right)}$. This shows that the submatrix containing the first $1+N_b$ columns of X_{Hess} is equal to the matrix X_{grad}. As an example, for the case $N_b = 3$ the matrix X_{Hess} is defined as:

$$X_{Hess} = \begin{bmatrix} x_1 & x_1+h & x_1 & x_1 & x_1+2h & x_1+h & x_1+h & x_1 & x_1 & x_1 \\ x_2 & x_2 & x_2+h & x_2 & x_2 & x_2+h & x_2 & x_2+2h & x_2+h & x_2 \\ x_3 & x_3 & x_3 & x_3+h & x_3 & x_3 & x_3+h & x_3 & x_3+h & x_3+2h \end{bmatrix}, \tag{6.4}$$

with the $N_b = 3$ single and $\frac{1}{2}\left(N_b + N_b^2\right) = 6$ double perturbations, and $\boldsymbol{x}_0 = [x_1, x_2, x_3]^T$. Next, for the 'base' beam intensity vector (the first column) and all the perturbed versions (all other columns) separate PCEs of the dose volume parameter $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ are constructed, following the methodology as in Chapter 4.6.1. Next, all PCEs are sampled N_s times from the underlying Gaussian sample space, and from these samples the $(100-\beta)$-th percentile values are calculated for each PCE. The value of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is determined via the percentile outcome of the PCE for the base beam intensity vector. The elements of the gradient and Hessian matrix are calculated according to Equations 5.12 and 5.13, using the determined percentile values of all $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ PCEs. This process is repeated at each iteration step of the optimization. Algorithm 1 shows the pseudo-code of this algorithm. The notation $A[i, j]$ is used to denote the (i, j)-th element of a matrix A and $\mathrm{Dim}(A, n)$ denotes the size of matrix A in the n-th dimension.

Algorithm 1 Calculate value, gradient and Hessian matrix of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$

$\boldsymbol{x} = \boldsymbol{x}_0$
if Gradient **then**
  Expand vector $\boldsymbol{x}$ to obtain matrix X_{grad}
else if Gradient & Hessian **then**
  Expand vector $\boldsymbol{x}$ to obtain matrix X_{Hess}
end if
function openGPC:
  Generate cubature grid with set of cubature points $\{\boldsymbol{\xi}_k\}$
  function computeResponses:
    for all $\boldsymbol{\xi} \in \{\boldsymbol{\xi}_k\}$, $i \in \mathcal{V}_{str}$ and $l \in \{1,\dots,\mathrm{Dim}(X_{grad/Hess},2)\}$ **do**
      Compute $(100-\alpha)$-th percentile of $\sum_{j=1}^{N_b} D[i,j](\boldsymbol{\xi}) \cdot X_{grad/Hess}[j,l]$ over index i
  Calculate coefficients of each PCE
Sample all PCEs N_s times with sample set $\{\boldsymbol{\xi}^s\}$
function prctile (Matlab):
  Compute $(100-\beta)$-th percentile for each set of PCE samples
if Gradient **then**
  Calculate gradient elements with Equation 5.12
else if Gradient & Hessian **then**
  Calculate gradient elements with Equation 5.12
  Calculate Hessian matrix elements with Equation 5.13
end if

6.3. MONOTONICITY CONSTRAINT ON PCE COEFFICIENTS

In the ideal case, when calculating the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$, a PCE is constructed of either the dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ itself or the elements of the dose deposition matrix $D_{ij}(\boldsymbol{\xi})$ with an infinite amount of basis functions. The functions of interest in treatment plan optimization are real-valued and continuous stochastic responses in the L^2-space, meaning that the second moment of the response function is finite. Then, from the generalized Polynomial Chaos (gPC) framework it follows that the PCE converges to the function $R(\boldsymbol{\xi})$ for $P \to \infty$ in Equation 4.9 [39]. Approximating $f(\boldsymbol{\xi})$ with a finite number of basis functions gives, therefore, rise to errors.

In treatment plan optimization the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is calculated for the beam intensity vector $\boldsymbol{x}$ and its perturbed versions $\boldsymbol{x} + h\hat{\boldsymbol{e}}_n$ and $\boldsymbol{x} + h\hat{\boldsymbol{e}}_m + h\hat{\boldsymbol{e}}_n$ with an arbitrary step size $h > 0$ and $n, m \in \{1,\dots,N_b\}$. Equation 3.4 shows that the dose in each voxel is a linear combination of the contributions of the different beamspots to the dose in that voxel and the intensities of those beamspots. Physically, the elements of the dose deposition matrix and the beam intensity vector are constrained to be non-negative, and therefore, the dose in an arbitrary voxel in a perturbed case $\boldsymbol{x} + h\hat{\boldsymbol{e}}_n$ or $\boldsymbol{x} + h\hat{\boldsymbol{e}}_m + h\hat{\boldsymbol{e}}_n$ is always greater than or equal to the dose in the base case $\boldsymbol{x}$, yielding:

$$d_i(\boldsymbol{x} + h\hat{\boldsymbol{e}}_n, \boldsymbol{\xi}) \geq d_i(\boldsymbol{x}, \boldsymbol{\xi}), \quad \forall i \in \{1,\dots,N_v\}, \quad \forall h \geq 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^N, \quad \forall n \in \{1,\dots,N_b\}. \tag{6.5}$$

Equation 6.5 holds for every realization of the uncertainty vector $\boldsymbol{\xi}$. This, in turn, means that for a function $f(\boldsymbol{x},\boldsymbol{\xi})$, representing any dose parameter, the function value in a perturbed case is always greater than or equal to the function value in the base case:

$$f(\boldsymbol{x} + h\hat{\boldsymbol{e}}_m + h\hat{\boldsymbol{e}}_n, \boldsymbol{\xi}) \geq f(\boldsymbol{x} + h\hat{\boldsymbol{e}}_n, \boldsymbol{\xi}) \geq f(\boldsymbol{x}, \boldsymbol{\xi}), \quad \forall h \geq 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^N, \quad \forall m, n \in \{1,\dots,N_b\}. \tag{6.6}$$

Because a PCE with an infinite amount of basis functions represents the function $f(\boldsymbol{x},\boldsymbol{\xi})$ exactly, this monotonicity in Equation 6.6 should therefore also hold for the PCE. If this function is approximated with a finite Polynomial Chaos Expansion, this expansion is written as:

$$R(\boldsymbol{x},\boldsymbol{\xi}) \approx \sum_{k=0}^{P} r_k(\boldsymbol{x}) \Psi_k(\boldsymbol{\xi}). \tag{6.7}$$

Monotonicity with respect to the beam intensity vector $\boldsymbol{x}$ then yields

$$\sum_{k=0}^{P} r_k(\boldsymbol{x} + h\hat{\boldsymbol{e}}_n) \Psi_k(\boldsymbol{\xi}) \geq \sum_{k=0}^{P} r_k(\boldsymbol{x}) \Psi_k(\boldsymbol{\xi}), \quad \forall h \geq 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^N, \tag{6.8}$$

where the Polynomial Chaos basis vectors Ψ_k are the same on the left- and right-hand side of Equation 6.8. By spectral projection the following inequality for the PCE coefficients is obtained:

$$r_k(\boldsymbol{x} + h\hat{\boldsymbol{e}}_n) \geq r_k(\boldsymbol{x}), \quad \forall h \geq 0, \quad \forall k \in \{0, \ldots, P\}. \tag{6.9}$$

A similar derivation can be performed for any configuration of positive perturbations of the elements of the beam intensity vector. In the independent construction of the separate PCEs for the base case and all perturbed cases this constraint has not been taken into account before, but is used in this research.

7

RESULTS

In this chapter, the results of research on improving probabilistic treatment planning with PCE for proton therapy are displayed. Chapter 7.1 focuses on the optimizations of cases 1 and 2 for the 3D spine geometry without the use of probabilistic dose volume parameters. In Chapter 7.2, three different improvements on Algorithm 1 are introduced that speed up the algorithm. Chapter 7.3 treats the results of the monotonicity constraint on the PCE coefficients. In chapter 7.4, an approximation method for the gradient and Hessian of probabilistic dose volume parameters is proposed to alleviate the computational cost of Algorithm 1 even further. In Chapter 7.5, the performance and computation times of the improved exact method and the approximation method are compared through optimizations of case 3 with the use of probabilistic dose volume parameters.

7.1. PROBABILISTIC PLANNING WITHOUT PROBABILISTIC DOSE VOLUME PARAMETERS

As discussed in Chapter 6.1.4, two optimizations with probabilistic functions, which are evaluated with PCEs, are first performed on the 3D spine geometry without the use of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$. In the first optimization, the expected value of the sum of quadratic differences between the delivered dose and the prescribed dose is minimized in the full discretized domain. In the second optimization, the same objective function is minimized, but also a constraint is set on the expected mean dose in the tumor to be greater than or equal to 60 Gy, which is chosen to be the prescribed dose in the tumor. These two optimization problems coincide with cases 1 and 2 in Table 6.1 respectively. Furthermore, settings for the prescribed dose and voxel weight in each structure, the considered uncertain parameter and PCE characteristics need to be defined. These settings are shown in Table 7.1.

Table 7.1: Parameter settings for the two optimizations of cases 1 and 2.

Prescribed dose:	Structure:	Value:	
	Tumor	60 Gy	
	Spine	0 Gy	
	Tissue	0 Gy	
Voxel weight:	Structure:	Value:	
	Tumor	500	
	Spine	500	
	Tissue	1	
Uncertain parameter:	Type:	Distribution:	
	Set-up uncertainty in x-direction	$\mathcal{N}(0, 0.3^2)$ [cm, cm^2]	
PCE:	Object:	Characteristic:	Cut-off value:
	All PCEs for cases 1 and 2	$L5E0O1$	10^{-3}

For all PCEs that need to be constructed, the PCE characteristic $L5E0O1$ is used, as has also been used in [12]. A first order polynomial is reasonable as only the zeroth order PCE coefficients are required. Furthermore, the grid level $L = 5$ is assumed to be accurate enough for this purpose.

The probabilistic functions in the objective function and constraint are meant to irradiate the tumor with the prescribed dose in a larger range of error scenarios, than in robust and conventional treatment planning. It is therefore expected that the nominal and expected dose distributions not only show high dose levels inside of the tumor, but also outside of the tumor in the direction in which the uncertain parameter is introduced. This provides a more conformal dose coverage of the tumor in a large range of error scenarios. This 'automatic' expansion of the irradiated volume is similar to the idea of a PTV margin. However, the calculation of what this margin should be is completely different. Here, this margin is carefully calculated with probabilistic optimization, and not chosen from clinical experience.

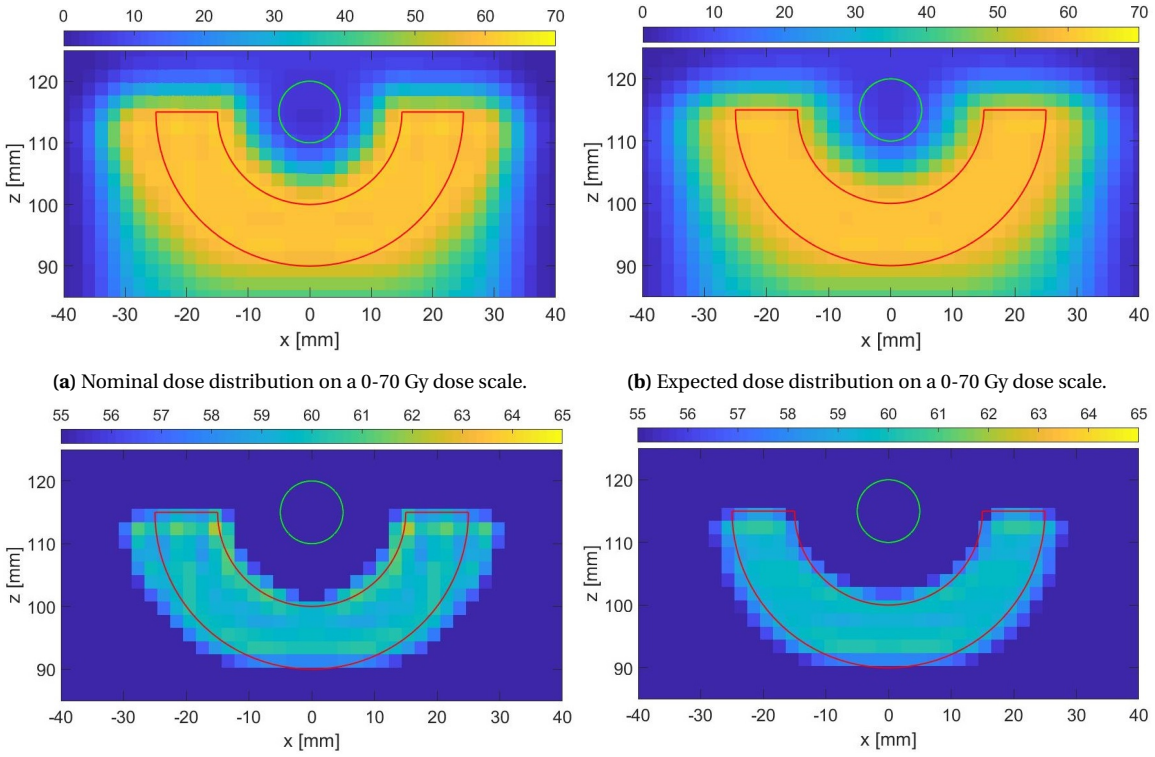
The PCEs are constructed with the OpenGPC package and the two optimizations are performed with the *fmincon* optimizer in Matlab, using the interior-point algorithm. The starting beam intensity vector is chosen as the zero vector in both optimizations. The optimizations are performed in the high performance computing (HPC) cluster of Delft University of Technology with 6 processors and 50 GB of memory reserved for each on one computer.

In case 1, the construction of the $\frac{1}{2}\left(N_b + N_b^2\right)$ and N_b PCEs for different objects in the objective function and the $N_v \cdot N_b$ PCEs for the elements of the dose deposition matrix, in total $\frac{3}{2}N_b + \frac{1}{2}N_b^2 + N_b \cdot N_v = 13,735,800$ PCEs, took 214.6 seconds. On the other hand, the optimization itself took only 11.8 seconds with 19 iterations. In case 2, there are N_b extra PCEs to be constructed of an object in the constraint. The construction of the $\frac{5}{2}N_b + \frac{1}{2}N_b^2 + N_b \cdot N_v = 13,737,672$ PCEs took 220.1 seconds and the optimization took 13.9 seconds with 20 iterations. The differences in computation time between the two optimizations are displayed in Table 7.2.

Table 7.2: Computation times of the construction of PCEs and the optimizations of the two probabilistic plans in cases 1 and 2.

Case:	PCE:	Optimization:
1	214.6 s	11.8 s
2	220.1 s	13.9 s

In Figure 7.1 axial cross-sections at the plane corresponding to $y = 5$ mm are shown with the optimized nominal and expected dose distributions of case 1. Figures 7.1a and 7.1b are colored according to a 0-70 Gy dose scale, whilst Figures 7.1c and 7.1d are colored according to a 55-65 Gy dose scale to have a better view of the dose fall-off around the edges of the tumor. All voxels with a dose less than 55 Gy are given the same color as the 55 Gy boundary for the 55-65 Gy scale.



(a) Nominal dose distribution on a 0-70 Gy dose scale.

(b) Expected dose distribution on a 0-70 Gy dose scale.

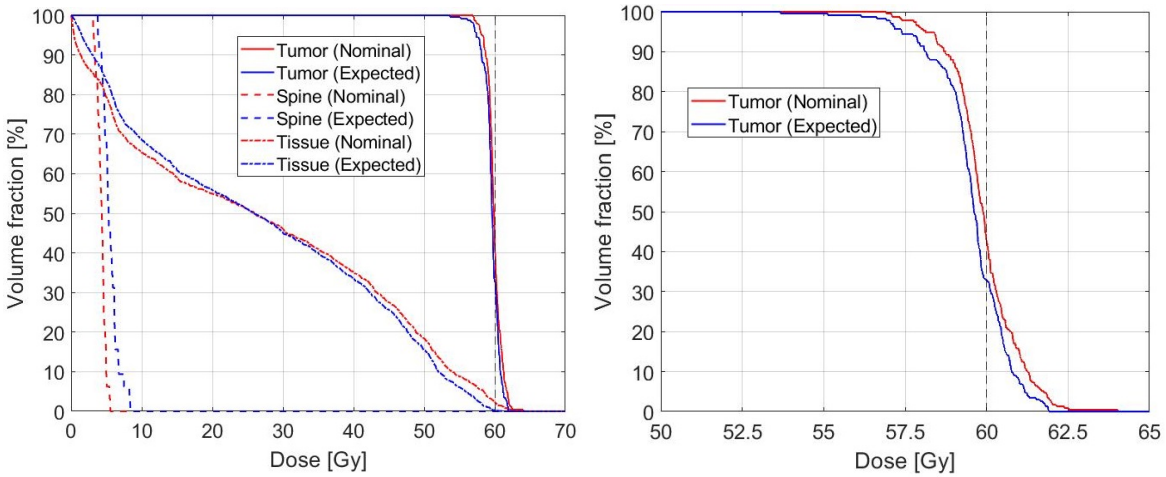
(c) Nominal dose distribution on a 55-65 Gy dose scale.

(d) Expected dose distribution on a 55-65 Gy dose scale.

Figure 7.1: Axial cross-sections of the 3D spine geometry at the plane corresponding to $y = 5$ mm with the nominal and expected dose distributions on a 0-70 Gy dose scale, **(a)** and **(b)** respectively, and on a 55-65 Gy dose scale, **(c)** and **(d)** respectively, for the optimized treatment plan of case 1. All voxels with a dose less than 55 Gy are given the same color as the 55 Gy boundary for the 55-65 Gy scale.

The dose distributions in Figure 7.1 show good results in terms of dose coverage in the tumor in the nominal scenario and also for the expected dose with an automatic expansion of the PTV in the x-direction as expected. The high weights that are assigned to the voxels in the spine and the tumor give rise to a larger expansion of the high dose levels near the outer-edge of the tumor, and less near the inner-edge of the tumor, where the spine is located. There is also a difference visible between the nominal dose and the expected dose. The expected dose distribution is more smooth than the nominal dose distribution, because the expected dose is the weighted average of the dose distributions in all possible error scenarios, following the underlying Gaussian probability space.

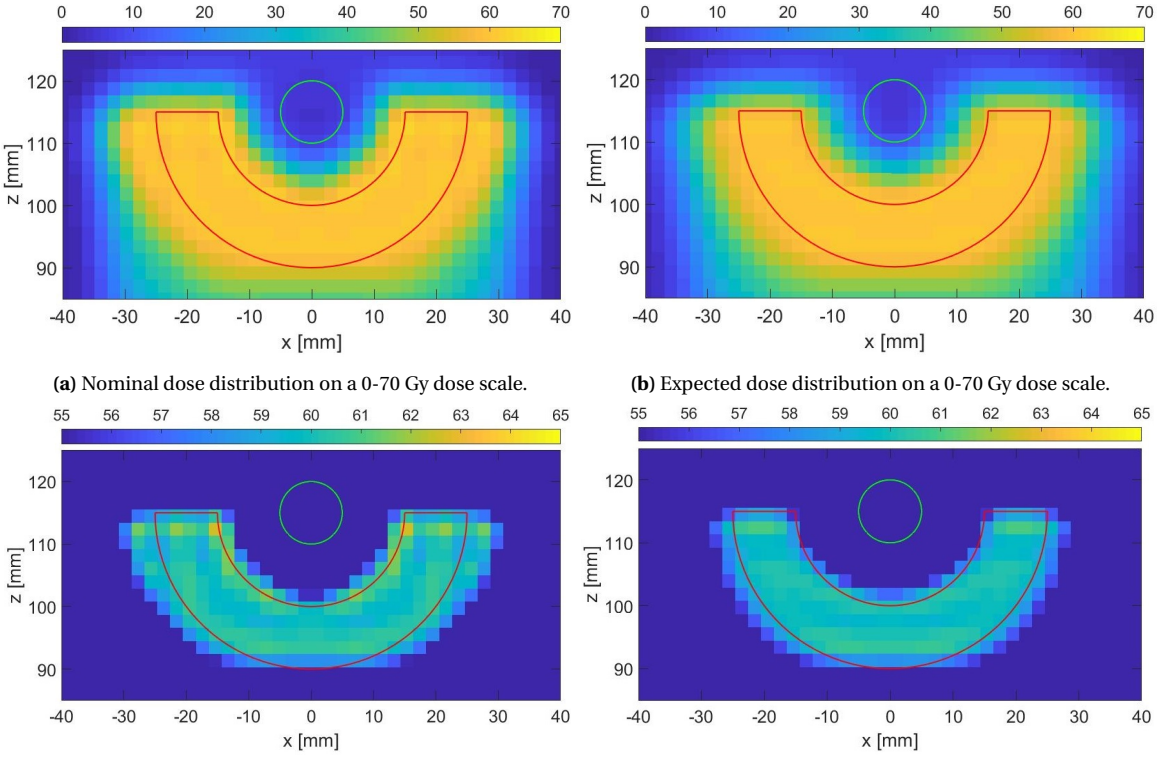
The results in Figure 7.1 only show axial cross-sections of the dose distributions. However, to fully compare the results of the nominal and expected dose distributions dose volume histograms are needed. Therefore, in Figure 7.2a the differences between the nominal and expected dose in the tumor, the spine and the healthy tissue are summarized in a dose volume histogram. In Figure 7.2b, the differences in nominal and expected dose in the tumor are highlighted.



(a) **(b)**

Figure 7.2: Dose volume histograms of the nominal and expected dose in the tumor, spine and healthy tissue **(a)** and only the tumor **(b)** for the optimized treatment plan of case 1.

The dose volume histograms in Figure 7.2 show that the expected dose distribution in the tumor reaches lower dose levels for each volume fraction than the nominal dose distribution. In the spine it is the other way around with the nominal dose reaching lower dose levels for each volume fraction than the expected dose. This is the result of optimizing probabilistically with the objective function of cases 1 and 2. Figures 7.3 and 7.4 show similar results as Figures 7.1 and 7.2, but then for the optimized treatment plan of case 2.



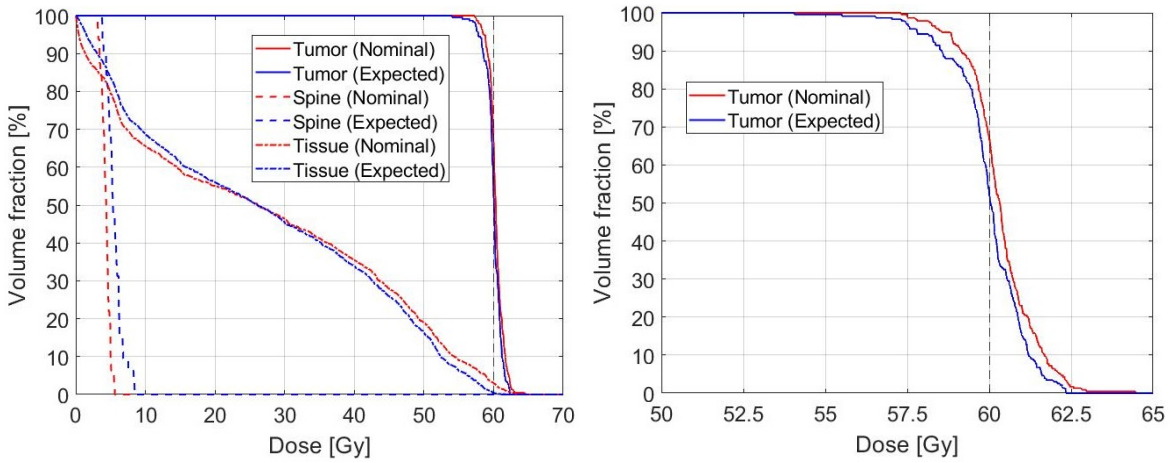
(a) Nominal dose distribution on a 0-70 Gy dose scale.

(b) Expected dose distribution on a 0-70 Gy dose scale.

(c) Nominal dose distribution on a 55-65 Gy dose scale.

(d) Expected dose distribution on a 55-65 Gy dose scale.

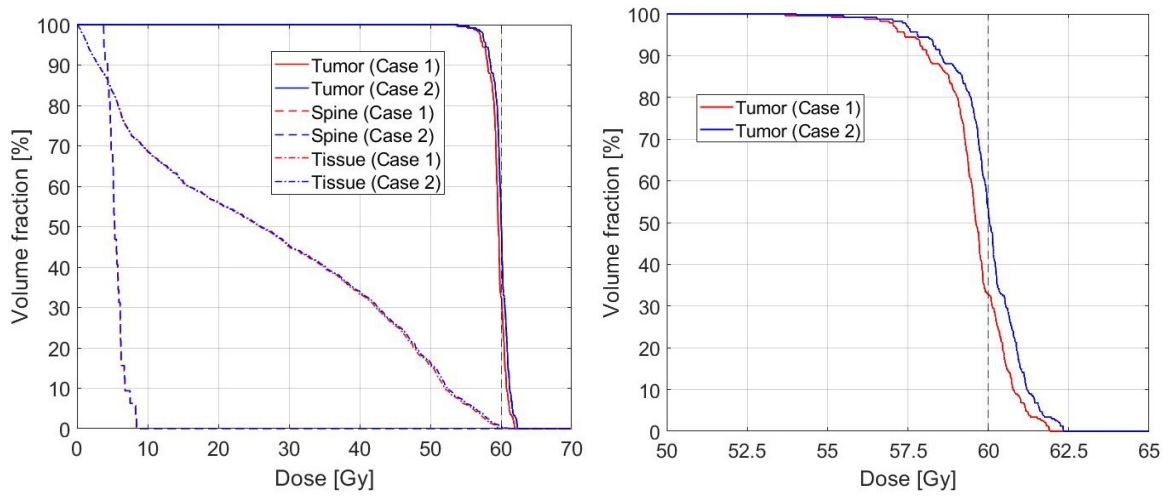
Figure 7.3: Axial cross-sections of the 3D spine geometry at the plane corresponding to $y = 5$ mm with the nominal and expected dose distributions on a 0-70 Gy dose scale, **(a)** and **(b)** respectively, and on a 55-65 Gy dose scale, **(c)** and **(d)** respectively, for the optimized treatment plan of case 2. All voxels with a dose less than 55 Gy are given the same color as the 55 Gy boundary for the 55-65 Gy scale.



(a) **(b)**

Figure 7.4: Dose volume histograms of the nominal and expected dose in the tumor, spine and healthy tissue **(a)** and only the tumor **(b)** for the optimized treatment plan of case 2.

The optimized treatment plan of case 2 is in the nominal and expected dose distributions similar to that of case 1, but the dose levels in the tumor are slightly higher due to the constraint on the expected mean dose in the tumor. To compare the results of cases 1 and 2, in Figure 7.5a the differences in the expected dose in the tumor, spine and healthy tissue between the two treatment plans are summarized in a dose volume histogram and in Figure 7.5b the differences in expected dose in the tumor are highlighted.



(a) **(b)**

Figure 7.5: Dose volume histograms of the expected dose to the tumor, spine and healthy tissue **(a)** and only the tumor **(b)** for the optimized treatment plans of cases 1 and 2.

The dose volume histograms in Figure 7.5 show that the expected dose in the tumor is consistently higher for all volume fractions in the treatment plan of case 2, resulting in almost equal parts of the tumor to be over- and underdosed. On the other hand, the expected dose in the tissue is slightly lower in the treatment plan of case 1 for the 0-40% volume fraction range. The expected dose in the spine seems to be the same for the two treatment plans. Through the assigned weights to the different voxels in the domain and possible constraints on the dose in the tumor or the spine, one could find a treatment plan that spares the spine more or irradiates the tumor more, depending on the desired outcome. In the end, the expected mean dose in the tumor is 59.6 Gy and 60.0 Gy in the treatment plans of cases 1 and 2 respectively. The solution of case 2 is found at the boundary of the constraint of 60 Gy for the expected mean dose in the tumor. A possible reason for this is that the optimization in case 2 with the constraint on the beam intensities to be non-negative and the constraint

on the expected mean dose in the tumor results in an optimization over a convex set, and that the objective function is convex, as claimed in [12].

The result of the optimized treatment plan of case 1 is in accordance with the work of [12]. The computation times for the construction of the PCEs are longer in this research, probably because of the difference in hardware that is used for the computations and a different number of jobs that was present in the cluster. The computation times of the optimization itself are also different, being faster in this research. A possible reason for this difference is that in [12] no Hessian matrix is supplied to the optimizer, and therefore resulting in longer calculations through the approximations of the Hessian matrix.

7.2. EXACT IMPROVEMENTS ON ALGORITHM

In this section, improvements on Algorithm 1 are discussed. In total three improvements are treated, which speed up the computation time, whilst yielding the exact same result. In Chapter 7.2.1, a method for parallel dose calculations in the PCE construction is discussed, and Chapters 7.2.2 and 7.2.3 focus on speeding up the gradient and Hessian matrix computation respectively. In Chapter 7.2.4, the overall improvement on the computation time of Algorithm 1 is summarized.

7.2.1. PARALLEL DOSE CALCULATIONS FOR PERTURBED BEAM INTENSITIES

The first improvement on Algorithm 1 is found in the construction of the matrices X_{grad} and X_{Hess} in Equations 6.2 and 6.3 in Chapter 6.2. In [12] the matrices X_{grad} and X_{Hess} are constructed through addition of the beam intensity vector to every column of a step matrix, S_{grad} and S_{Hess} respectively, containing the single (and double) step size increases h for all constructed PCEs. In Matlab this can be performed as:

$$\begin{cases} X_{grad} = \boldsymbol{x} + S_{grad}, \\ X_{Hess} = \boldsymbol{x} + S_{Hess}, \end{cases} \tag{7.1}$$

with $S_{grad} \in \mathbb{R}^{(N_b)\times(1+N_b)}$ and $S_{Hess} \in \mathbb{R}^{(N_b)\times\left(1+N_b+\frac{1}{2}\left(N_b+N_b^2\right)\right)}$. The matrices S_{grad} and S_{Hess} are constructed with one and two for-loops respectively to fill in the step matrix, which is quite costly. Afterwards, the matrix with the base and perturbed dose vectors is calculated through multiplication of the dose deposition matrix with the X_{grad} or X_{Hess} matrix. For the 3D spine geometry with the voxel and beamspot grid defined as in Chapter 6.1, the construction of $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right) = 439{,}453$ PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ for the base beam intensity vector and all single and double perturbed versions with PCE characteristic $L7E1O8$ in the HPC cluster with one processor and 50 GB of storage memory available takes approximately 1,300 seconds.

The improvement on this calculation is based on the fact that the dose in each voxel is a linear combination of the contribution of each beamspot to that voxel and the intensities of the beamspots, as shown in Equation 3.4. Therefore, for the single perturbed beam intensity cases the dose increases for the N_b cases can be computed all at once via $h \cdot D(\boldsymbol{\xi})$ for step size $h > 0$ and the resulting matrix with perturbed dose vectors $d_{grad}(\boldsymbol{x},\boldsymbol{\xi},h)$ can in Matlab be computed as:

$$d_{grad}(\boldsymbol{x},\boldsymbol{\xi},h) = D(\boldsymbol{\xi}) \cdot \boldsymbol{x} + h \cdot D(\boldsymbol{\xi}). \tag{7.2}$$

A similar, but slightly more complicated procedure can be used to calculate the matrix $d_{Hess}(\boldsymbol{x},\boldsymbol{\xi},h)$, containing all $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ perturbed dose vectors. In Matlab this can be coded as follows in Algorithm 2, in which $d_{Hess}[*,i]$ denotes the i-th column of matrix d_{Hess} and $||_{j=1}^{n} d_{Hess}[*,i]$ denotes the vertical concatenation of the i-th column of d_{Hess} n times.

Algorithm 2 Compute $d_{Hess}(\boldsymbol{x},\boldsymbol{\xi},h)$

$d_{Hess}[*,1] = D(\boldsymbol{\xi}) \cdot \boldsymbol{x}$
$d_{Hess}[*,2:N_b+1] = D(\boldsymbol{\xi}) \cdot \boldsymbol{x} + h \cdot D(\boldsymbol{\xi})$
$j = N_b + 2$
$k = 2N_b + 1$
for $i = 1 : N_b$ **do**
    $d_{Hess}[*,j:k] = d_{Hess}[*,1] + d_{Hess}[*,2+i:N_b+1] + ||_{n=1}^{N_b+1-i} d_{Hess}[*,1+i]$
    $j = j + N_b + 1 - i$
    $k = k + N_b - i$
end for

With the same geometry and PCE settings in the HPC cluster the construction of PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ for the base beam intensity vector and all single and double perturbed intensities takes approximately 200 seconds. This results in roughly a factor of six decrease in computation time for this part of Algorithm 1.

7.2.2. GRADIENT COMPUTATION

The improvement on the gradient computation of Algorithm 1 involves vectorization of the N_b calculations as in Equation 5.12. In previous research of [12] the elements of the gradient vector were calculated element-wise, whereas the computation time could easily be reduced with the use of vectorization. For the specified beamspot grid as in Chapter 6.1 with $N_b = 936$ the computation time has been reduced from 1.0 seconds to $5 \cdot 10^{-4}$ seconds. This is a small improvement compared to the other two improvements.

7.2.3. HESSIAN MATRIX COMPUTATION

The Hessian matrix computation of Algorithm 1 is based on Equation 5.13. In previous research of [12] the $\frac{1}{2}\left(N_b + N_b^2\right)$ elements of the upper, or lower, triangular part of the Hessian matrix, as the Hessian matrix is symmetric, were calculated with a double for-loop, one over the columns and one over the rows. This is a costly computation and takes approximately 1,500 seconds for the 3D geometry.

The improvement on the Hessian matrix computation involves the use of a sparse matrix and vector multiplication instead of the double for-loop. The $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ percentile values of the base case and all perturbed cases can be stored in a column vector. Then, the elements of the Hessian matrix can be calculated at once with a matrix-vector multiplication. Looking at Equation 5.13 each row of this multiplication matrix will consist of four nonzero elements; twice a factor $\frac{1}{h^2}$ and twice a factor $-\frac{1}{h^2}$ on the positions that correspond to the right elements of the vector with percentile values. As each row of this multiplication matrix has only four nonzero elements, this matrix can be made sparse to save storage memory for an $\frac{1}{2}\left(N_b + N_b^2\right)$ by $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ matrix. This large sparse matrix only uses 32 MB of memory in the case $N_b = 936$. The multiplication matrix only has to be constructed once for the full optimization as the multiplication coefficients are the same for each iteration. Only the column vector with percentile values changes per iteration. For the 3D geometry the construction of this sparse multiplication matrix takes approximately 29 seconds and the Hessian matrix calculation takes only 1.0 seconds. This shows a very large decrease in computation time.

7.2.4. OVERALL IMPROVEMENT ON ALGORITHM'S COMPUTATION TIME

After discussing the three different improvements on Algorithm 1, it is time to summarize the overall improvement on the computation time for a single iteration step in the optimization. These results are summarized in Table 7.3.

Table 7.3: Computation times of the different parts in Algorithm 1 for the approach in [12] and the three improvements.

Part of algorithm:	Old approach	New approaches
Construction of $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$	1,300 s	200 s
Computation of gradient of β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$	1 s	$5 \cdot 10^{-4}$ s
Construction of sparse multiplication matrix	-	29 s
Computation of Hessian of β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$	1,500 s	1 s
Total:	2,801 s	230 s

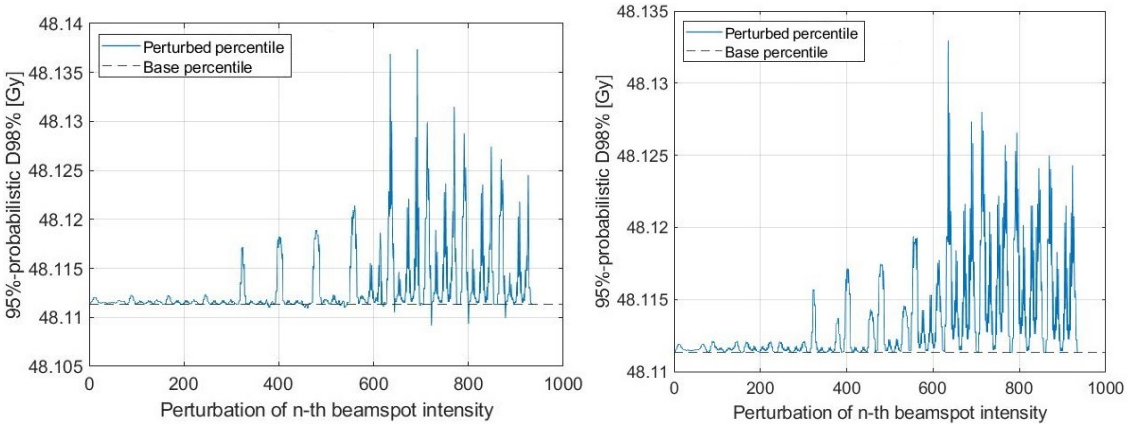
The first and third improvement show a large decrease in computation time compared to the second improvement. In total, the computation time is roughly decreased by a factor of 12. All three improvements result in the exact same outcome as the methodology used in [12]. The part of the algorithm that has not been covered yet is the actual percentile calculations for all PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$. This methodology and a proposition of an approximation method are discussed in Chapter 7.4.

7.3. MONOTONICITY CONSTRAINT ON PCE COEFFICIENTS

In Chapter 6.3, a constraint on the PCE coefficients is derived to ensure monotonicity of the PCE with respect to the beam intensity vector in the ideal case where the PCE is constructed with an infinite amount of

basis functions. As in practice always a finite polynomial expansion is constructed, still problems with monotonicity might occur. This section focuses on the differences in monotonicity of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ between the cases with and without constraint on the coefficients of the PCE of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$.

In Figure 7.6 the values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$, the probabilistic near-minimum dose, are shown for the base beam intensities and all the single perturbed versions without and with constraint on the PCE coefficients. Figure 7.7 shows similar results as Figure 7.6 for the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$, the probabilistic near-maximum dose. The results are retrieved from the 3D spine geometry with the base beam intensity vector obtained from the optimized treatment plan of case 1 in Chapter 7.1. The PCEs are constructed with the characteristics $L7E1O8$ and are sampled 500,000 times. The step size is chosen as $h = 0.005$, whereas the base beam intensities are in the range of 0-0.4 MU. To compare the cases with and without constraint on the PCE coefficients, the same set of samples is used for both cases.

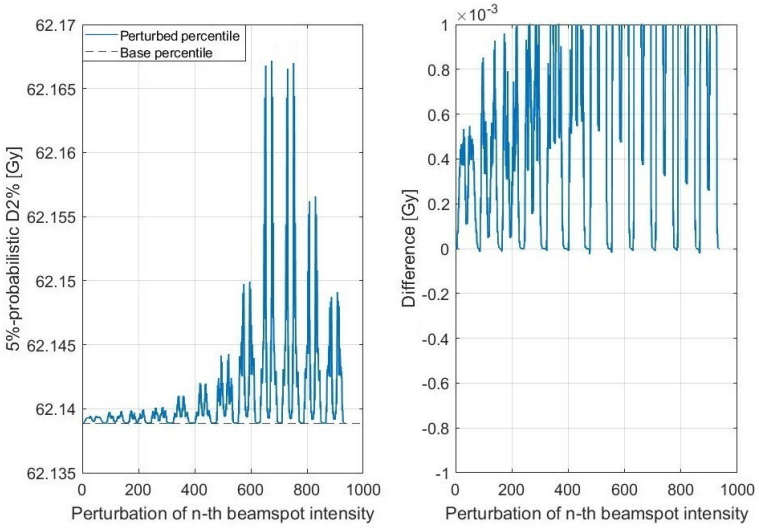


(a) Without constraint on PCE coefficients.

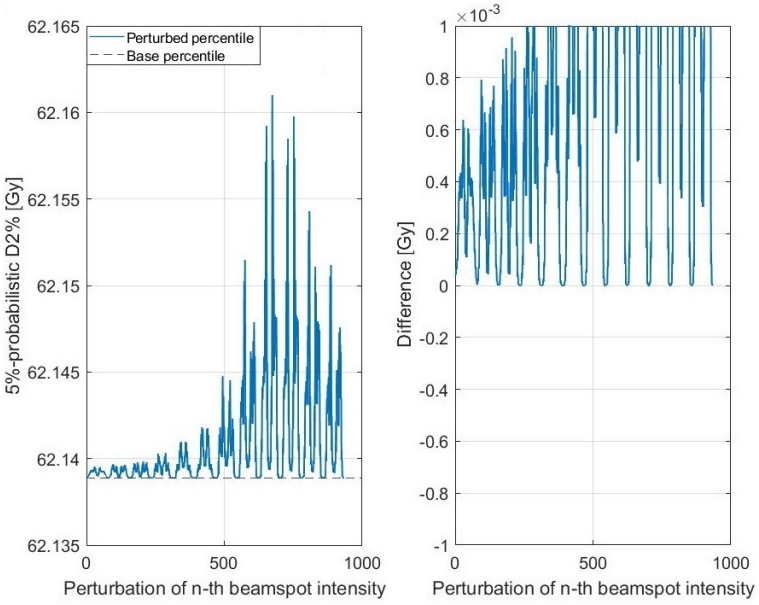
(b) With constraint on PCE coefficients.

Figure 7.6: Values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the base beam intensities and all the single step size increased versions with the base beam intensity vector from the optimized plan of case 1. The PCEs are constructed without **(a)** and with **(b)** constraint on the PCE coefficients respectively. The horizontal axis is based on an ordering of the N_b beamspots in the created grid, so $n \in \{1,\ldots,N_b\}$. The N_b percentile values for the perturbed beam intensities are connected for clarity.

The results in Figure 7.6 show good improvement in terms of monotonicity with respect to the beam intensity vector. The values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ become fully monotonic when the constraint on the PCE coefficients is applied. This constraint not only affects the value of non-monotonic perturbed percentiles, but improves the accuracy of the percentile value for all perturbed PCEs due to a more accurate construction of the PCEs. This results in quite some differences between Figure 7.6a and Figure 7.6b, when looking at the 'peaks'. Even though the non-monotonic behavior is small in Figure 7.6a, it effects the calculation of the gradient, becoming negative, which is physically incorrect and could drive the optimization in the wrong direction.



(a) Without constraint on PCE coefficients.



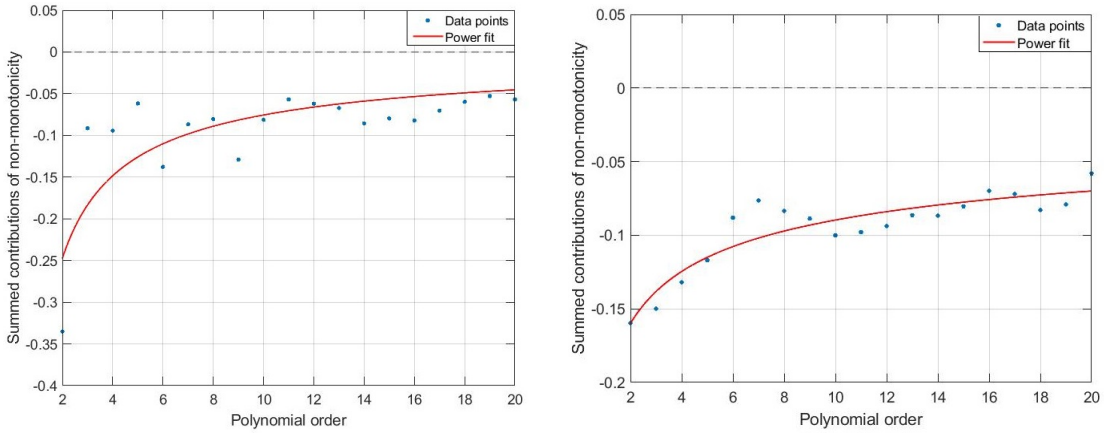
(b) With constraint on PCE coefficients.

Figure 7.7: Values of the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ for the base beam intensities and all the single step size increased versions with the base beam intensity vector from the optimized plan of case 1. The PCEs are constructed without **(a)** and with **(b)** constraint on the PCE coefficients respectively. In the left images the percentile values are shown and in the right images the differences between the perturbed percentile values and the base percentile value are shown. The horizontal axis is based on an ordering of the N_b beamspots in the created grid, so $n \in \{1,\ldots,N_b\}$. The N_b percentile values for the perturbed beam intensities are connected for clarity.

Similarly to Figure 7.6, the results in Figures 7.7 also show improvements in terms of monotonicity with respect to the beam intensity vector. In this case, the effects of non-monotonicity in Figure 7.7a are even smaller than in Figure 7.6a, but some of the visible 'peaks' are different when comparing Figure 7.7a and Figure 7.7b.

Even though the results of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ are fully monotonic, non-monotonic behavior could in theory still occur in the PCE of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ at certain arguments $\boldsymbol{\xi}$. This can occur at arguments for which some PC basis vectors in Equation 6.7 have a negative value, since in both cases with and without constraint the same set of PC basis vectors is used for the PCE construction. The PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ with constraint on the PCE coefficients are still an approximation. However, as derived in Chapter 6.3, any non-monotonic behavior should completely disappear in the case $P \to \infty$ in Equation 6.7. Nevertheless, the constraint on the PCE coefficients will mitigate the non-monotonic behavior and will thus improve the accuracy of the 'perturbed' probabilistic dose volume parameters.

To review the converging monotonic behavior of the perturbed PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ with constraint on the PCE coefficients, in Figure 7.8 the results of PCEs of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ and $d_{2\%}(\boldsymbol{x},\boldsymbol{\xi})$ are shown, comparing $N_s = 500{,}000$ sample values for the base PCE and all single step size perturbed PCEs for increasing polynomial orders and grid levels in the PCE construction. As a measure of non-monotonicity all negative differences between the perturbed PCE sample values and the base PCE sample values are summed. The step size is again chosen as $h = 0.005$ and the base beam intensities are retrieved from the optimized treatment plan of case 1 in Chapter 7.1.



(a) Results for PCEs of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$.

(b) Results for PCEs of $d_{2\%}(\boldsymbol{x},\boldsymbol{\xi})$.

Figure 7.8: Summed contributions of non-monotonicity of the PCEs of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ **(a)** and $d_{2\%}(\boldsymbol{x},\boldsymbol{\xi})$ **(b)** for single perturbed beam intensities, compared to the PCE for the base beam intensities for increasing polynomial orders and grid levels. The step size is chosen as $h = 0.005$ and each PCE is sampled on the same $N_s = 500{,}000$ arguments $\boldsymbol{\xi}$. The data points visualize the result for different PCE characteristics: $LaE1Ob$ for polynomial orders $b \in \{2,\dots,20\}$ and grid levels $a = b-1$. A power curve of the form $f(x) = a_1 * x^{a_2}$ is fitted through the data points.

In Figures 7.8a and 7.8b a decrease in non-monotonic behavior of the perturbed PCEs can be seen. In both figures, a curve of the form $f(x) = a_1 * x^{a_2}$ is fitted through the data points with the *Curve Fitting* toolbox in Matlab to review the trend of the data points for increasing polynomial orders and grid levels. As expected, a converging behavior is visible towards full monotonicity of the perturbed PCEs of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ and $d_{2\%}(\boldsymbol{x},\boldsymbol{\xi})$. However, it is difficult to conclude convergence towards monotonicity from these results alone, apart from the underlying theory.

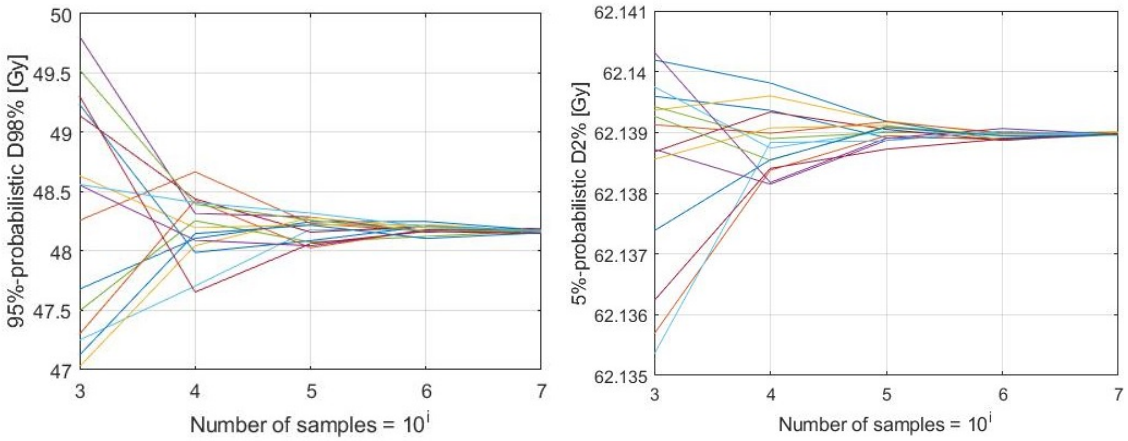
7.4. CALCULATION OF PROBABILISTIC DOSE VOLUME PARAMETERS

In previous research of [12], the $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ values of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for the base beam intensity vector $\boldsymbol{x}$ and all the single and double perturbed versions $\boldsymbol{x} + h\hat{\boldsymbol{e}}_n$ and $\boldsymbol{x} + h\hat{\boldsymbol{e}}_m + h\hat{\boldsymbol{e}}_n$ with $n, m \in \{1,\dots,N_b\}$ were calculated through separate constructions of PCEs of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ for each configuration of beam intensities, sampling each individual PCE N_s times and using the *Prctile* function in Matlab to obtain the percentile values. For every PCE this function sorts the PCE samples in increasing order, the index that corresponds to the desired $(100-\beta)$-th percentile is calculated via the following equation:

$$\text{Index} = \frac{100-\beta}{100} \cdot N_s + \frac{1}{2}, \tag{7.3}$$

and the right value out of the sorted list of PCE samples is obtained [45]. In the case that the percentile index is an integer, the percentile value is directly found from the sorted list, but otherwise, linear interpolation is used between the two surrounding values in the sorted list. This is a costly computation, and therefore, the choice of the sample size is a trade-off between the accuracy and the computation time of the percentile calculation.

In order to have an idea about the accuracy of the percentile calculation for different sample sizes, Figure 7.9 shows approximations of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ for 15 realizations of different sets of samples per specified sample size. The sample size is varied between $N_s = 10^3$ and $N_s = 10^7$, increasing in factors of 10. The beam intensity vector is obtained from the optimized treatment plan of case 1 in Chapter 7.1.



(a) Results of 95%-probabilistic $d_{98\%}(\boldsymbol{x})$.

(b) Results of 5%-probabilistic $d_{2\%}(\boldsymbol{x})$.

Figure 7.9: Approximations of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ **(a)** and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ **(a)** with varying sample sizes $N_s = 10^i$ for $i \in \{3,4,5,6,7\}$. For each value of N_s the results from 15 different realizations of sets of samples are shown.

The results in Figure 7.9 show a large difference in variance in the percentile value between the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$. The absolute differences between realizations are much larger in the $d_{98\%}$ case than in the $d_{2\%}$ case, which is expected as we are looking at the maximum dose in a much larger volume fraction in the first case, while considering a large set of error scenarios, and the dose distributions in the tumor are not completely homogeneous. Also, for this considered beam intensity vector from a probabilistic plan Figure 7.1 shows quite an expansion of the PTV in the direction of the set-up uncertainty. This makes the percentile values less sensitive to a difference in sample argument in the 95-th percentile case, the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$, where the less extreme error scenarios are more important compared to the 5-th percentile case, the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$, where the more extreme error scenarios are considered. In the extreme scenarios there is a steeper dose fall-off outside of the tumor, which makes the percentile value more sensitive to a difference in sample argument.

The computation times are exactly the same for the calculation of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for different values of α and β. The computation times for the calculation of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for the base beam intensities and all of the single and double perturbed beam intensity configurations are summarized in Table 7.4 for the same sample sizes as in Figure 7.9, and are performed in the HPC cluster with one processor and 50 GB of storage memory available. Since there is not enough storage capacity for all the $N_s \times \left(1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)\right)$ PCE samples to calculate all percentile values at once, this calculation has to be performed in batches with a specified memory size. This does lead to the same computation times as computing the percentile values at once.

Table 7.4: Computation times for the calculation of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for the base beam intensities and all single and double perturbed versions for the 3D spine geometry for different sample sizes.

Number of samples:	Computation time:
1,000	7.4 s
10,000	76.2 s
100,000	1,100 s ≈ 18 min
1,000,000	13,200 s ≈ 220 min ≈ 3.67 h
10,000,000	159,300 s ≈ 44.25 h

The computation times in Table 7.4 do not show a linear dependence on the sample size. This could be explained in two ways: the computation times are slowed down due to a larger memory usage, and the computation time of sorting the N_s PCE samples is not linearly dependent on the sample size. When combining the results in Figure 7.9 and Table 7.4, it seems the case that for smaller values of α and β the trade-off between the accuracy of the percentiles values and the computation time is less present. At least in the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ case, a small number of samples, around $N_s = 1,000 - 10,000$, already leads to a maximum difference

of 0.05 Gy between the different realizations. For larger values of α and β this trade-off needs to be taken into account. In the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ case at a large sample size of around $N_s = 100,000$ there is still a maximum difference of 0.3 Gy between the different realizations. Inaccuracies in the percentile values that are too large might drive the optimization of a treatment plan in the wrong direction through the values of the gradient and Hessian or might lead to oscillations when considering a small dose scale. However, the inaccuracies could partly be mitigated through the use of the same sample set for all PCEs in one iteration. On the other hand, a too large computation time is problematic for the application in medical clinics.

In [12] for all probabilistic dose volume parameters a sample size of $N_s = 500,000$ and the PCE characteristic $L5E1O6$ are used. Then, the percentile calculation takes approximately 5,600 seconds and the PCE construction 1,200 seconds in the HPC cluster with one processor and 50 GB of storage memory available. This computation time of the PCE construction corresponds to the old approach. With the new approach the computation time would be decreased to approximately 170 seconds. The results in Table 7.3 then show that in Chapter 7.2 the computation time of Algorithm 1 is decreased by $\left(1 - \frac{200s+5600s}{2700s+5600s}\right) \cdot 100\% \approx 30\%$ through the three improvements. This will also yield a similar speed-up for the full optimization of a treatment plan, depending on the dose parameters in the objectives and constraints, and their computation times. Unfortunately, the percentile calculations still consume $\frac{5600s}{2700s+5600s} \cdot 100\% \approx 67\%$ of the complete computation time of Algorithm 1 in the old approach. Therefore, other methods are needed to reduce the computation time of Algorithm 1 even further. For this purpose, an approximation method for the calculation of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for perturbed beam intensity vectors is devised.

7.4.1. APPROXIMATION METHOD FOR 'PERTURBED' PROBABILISTIC DOSE VOLUME PARAMETERS

The approximation method is based on the question whether in general there exists a realization of the uncertainty vector in the multidimensional sample space that corresponds to the percentile of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ for a certain beam intensity vector $\boldsymbol{x}$ and all its step size perturbed versions. If so, one could approximate the argument of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ through sampling for a relatively large sample size and use this argument to evaluate all PCEs for the perturbed beam intensity cases. This would reduce the computation time in the percentile calculation for the $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ PCEs significantly to the order of only 1 second. Also, there would not be a trade-off between the sample size and the computation time anymore as only the PCE for the base beam intensities is sampled and the percentile calculation is performed once. With this method, the computation time of Algorithm 1 would be reduced from roughly $5,600 + 2,700 = 8,300$ seconds to only $170 + 30 + 1 = 201$ seconds for the 3D spine geometry with the calculations on the HPC cluster with one processor and 50 GB of memory available when using a sample size of 500,000 and the PCE characteristic $L5E1O6$.

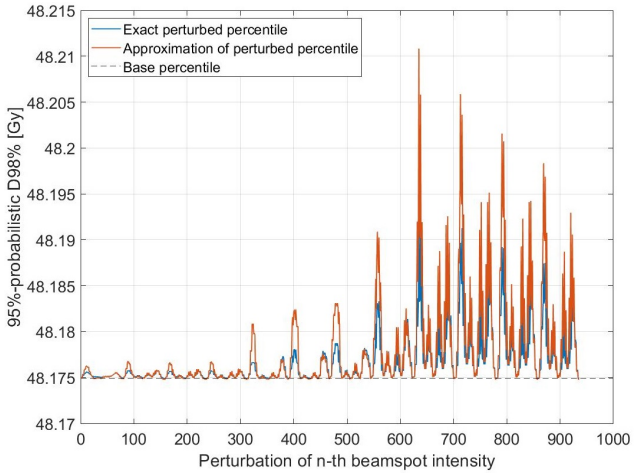
In the ideal case, the PCEs are constructed with an infinite amount of basis functions. Then, monotonicity of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ over the beam intensity vector is definitely satisfied. With this approach the PCE for the base beam intensity vector is sampled N_s times, with the samples chosen randomly from the Gaussian distribution of the uncertainty vector, to obtain the PCE sample set $\left\{d_{\alpha\%}\left(\boldsymbol{x}, \boldsymbol{\xi}^1\right), \ldots, d_{\alpha\%}\left(\boldsymbol{x}, \boldsymbol{\xi}^{N_s}\right)\right\}$ with the corresponding Gaussian sample set $\left\{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^{N_s}\right\}$. This set of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ values is then sorted in increasing order to obtain the set $\left\{d_{\alpha\%}^{(1)}(\boldsymbol{x}), \ldots, d_{\alpha\%}^{(N_s)}(\boldsymbol{x})\right\}$. For a certain percentile index γ, following from Equation 7.3, the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is found as $d_{\alpha\%}^{(\gamma)}(\boldsymbol{x})$, for which linear interpolation is used between the values $d_{\alpha\%}^{(\lfloor\gamma\rfloor)}(\boldsymbol{x})$ and $d_{\alpha\%}^{(\lceil\gamma\rceil)}(\boldsymbol{x})$ in the case that γ is not an integer. Here, $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling function respectively. Also, when γ is not an integer, as percentile argument either $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$ or $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$ should be chosen to evaluate the PCEs for the perturbed beam intensity vectors on. Linear interpolation between $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$ and $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$ will not work in general as there could be multiple arguments that yield the same percentile value.

When considering the same sample set from the uncertainty vector for all PCEs, the set of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ values are not the same for each PCE with perturbed beam intensities. For an arbitrary percentile index γ the same argument corresponds to the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for each PCE if the ordering $\left\{d_{\alpha\%}^{(1)}(\boldsymbol{x}), \ldots, d_{\alpha\%}^{(N_s)}(\boldsymbol{x})\right\}$ is unchanged. Otherwise, this will not hold. In general, this only holds if the value of $d_{\alpha\%}\left(\boldsymbol{x} + h \cdot \sum_{i=1}^{M} \hat{e}_{k_i}, \boldsymbol{\xi}\right)$ for an arbitrary, positively perturbed version of the beam intensity vector $\boldsymbol{x}$ with $M \in \mathbb{N}$, $h > 0$ and $k_i \in \{1, \ldots, N_b\}$ is the value of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ plus a certain shift H, which is dependent on the perturbation $h \cdot \sum_{i=1}^{M} \hat{e}_{k_i}$. Mathematically, this is written as:

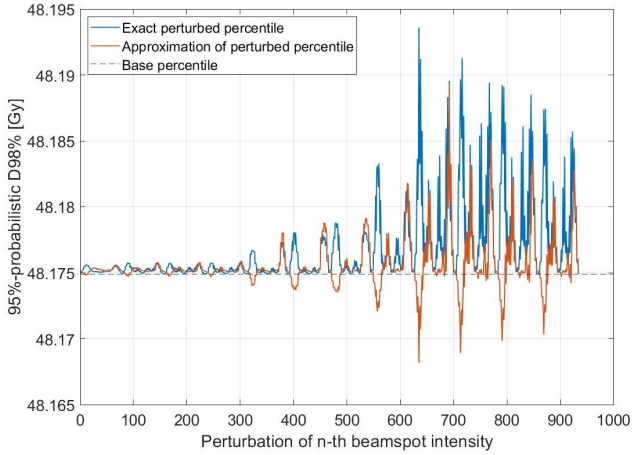
$$d_{\alpha\%}\left(\boldsymbol{x} + h \cdot \sum_{i=1}^{M} \hat{e}_{k_i}, \boldsymbol{\xi}\right) = d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi}) + H\left(h \cdot \sum_{i=1}^{M} \hat{e}_{k_i}\right), \quad \forall h > 0, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^N, \quad \forall M \in \mathbb{N}, \quad \forall k_i \in \{1, \ldots, N_b\}. \tag{7.4}$$

From Equation 3.4 it is clear that the dose vector is a linear function of the beam intensity vector, but it is not the case in general that a perturbation of beam intensities yields the same increase in the dose vector for each realization of the uncertainty vector. Therefore, the percentile argument $\boldsymbol{\xi}^{(\gamma)}$ becomes an approximation for the PCEs of perturbed beam intensity vectors in the case $h > 0$. For $h = 0$ this relation obviously holds with $H(0) = 0$, so the method should improve in accuracy for $h \to 0$. At the same time, in practice the PCEs are constructed with a finite number of basis functions, in which certain inaccuracies arise. The accuracy of this approximation method should therefore also improve for increasing number of basis functions in the PCE construction.

In the implementation of this approximation method on the 3D spine geometry, with the beam intensity vector from the optimized treatment plan of case 1 in Chapter 7.1, some inconsistency in results has been found in the approximation of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for the perturbed beam intensities as needed in the forward finite difference scheme. Figure 7.10 shows the approximation of the 95%-probabilistic $d_{98\%}(\boldsymbol{x} + h\hat{e}_n)$ for $n \in \{1, \ldots, N_b\}$ for a set of 500,000 Gaussian samples for this approximation method and the exact method with step size $h = 0.005$. As for $N_s = 500{,}000$ and $\beta = 95$ the percentile index γ is not an integer, a choice has to be made about which percentile argument to evaluate the PCEs for perturbed beam intensities on. In Figures 7.10a and 7.10b the results are shown when the percentile arguments $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$ and $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$ are used respectively.



(a) Evaluation of 'perturbed' PCEs with $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$.



(b) Evaluation of 'perturbed' PCEs with $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$.

Figure 7.10: Values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the base beam intensities (dotted line) and for the single step size perturbed beam intensities (solid lines) with step size $h = 0.005$. In blue the exact percentile values are shown, and in red the percentile values from the approximation method are shown for the two percentile arguments $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$ in (**a**) and $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$ in (**b**). The perturbed percentile values are connected for clarity.

The differences in the approximation of the percentile values in Figure 7.10 originates from the fact that the

PCE of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ with this specific beam intensity vector yields two distinct arguments in the sample space that correspond to the 5-th percentile and that for the used Gaussian sample set the results in Figures 7.10a and 7.10b follow from these two distinct arguments. The perturbed percentile values in Figure 7.10 are connected for clarity. The peaks that are visible seem to be the result from intensity increases at beamspots that are inside the tumor or near the tumor in the direction in which the set-up uncertainty is defined. The increased beam intensities have different influences on the $d_{98\%}(\boldsymbol{x}+h\cdot\hat{e}_n,\boldsymbol{\xi})$ for different arguments $\boldsymbol{\xi}$ due to the location of the beamspots. In the results of Figure 7.10a the percentile values at the peaks seem to be overestimated. Similarly, in Figure 7.10b the percentile values at the peaks seem to be underestimated. Figure 7.11 shows that using linear interpolation between $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$ and $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$ to evaluate the PCEs for perturbed beam intensities on will not work in general, and therefore, another method is needed for more accurate approximations. In Figure 7.12 the results are shown similarly to Figure 7.10, but now the percentile value of each 'perturbed' PCE is averaged over the outcomes at arguments $\boldsymbol{\xi}^{(\lfloor\gamma\rfloor)}$ and $\boldsymbol{\xi}^{(\lceil\gamma\rceil)}$. In Figures 7.10 and 7.12 the same set of Gaussian samples is used.

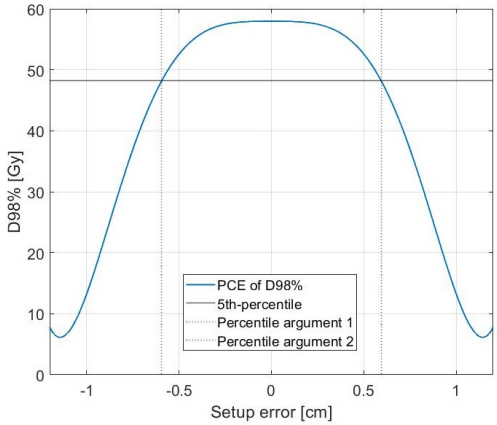


Figure 7.11: PCE of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ with characteristic $L7E1O8$ and the beam intensity vector obtained from the optimized treatment plan of case 1. Also, a horizontal line at the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$, calculated from 500,000 samples, and two vertical lines at the two distinct percentile arguments are drawn.

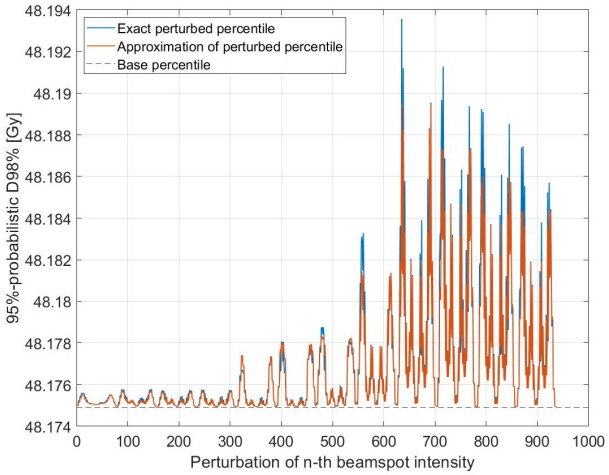
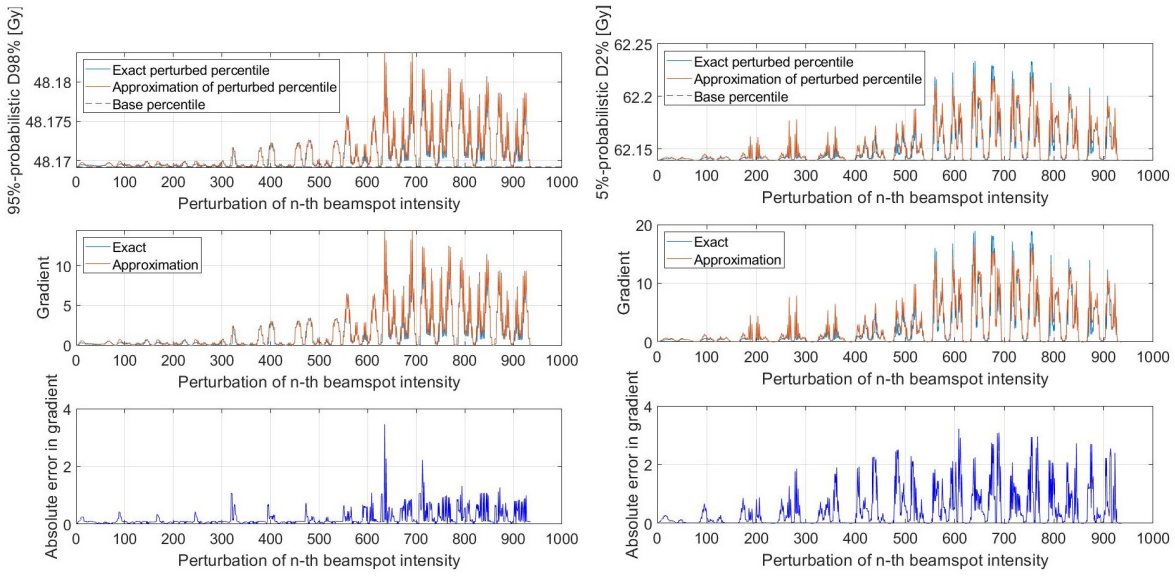


Figure 7.12: Values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the base beam intensities (dotted line) and for the single step size perturbed beam intensities (solid lines) with step size $h = 0.005$. In blue the exact percentile values are shown, and in red the percentile values from the approximation method are shown with the outcome of the 'perturbed' PCEs averaged over the two distinct percentile arguments. The perturbed percentile values are connected for clarity.

In Figure 7.12 the results with the approximation method are much better than in Figure 7.10. The problem is, however, that in general for an arbitrary patient geometry, beam intensity vector and set of uncertain parameters the number of distinct percentile arguments is unknown and that the arguments themselves are

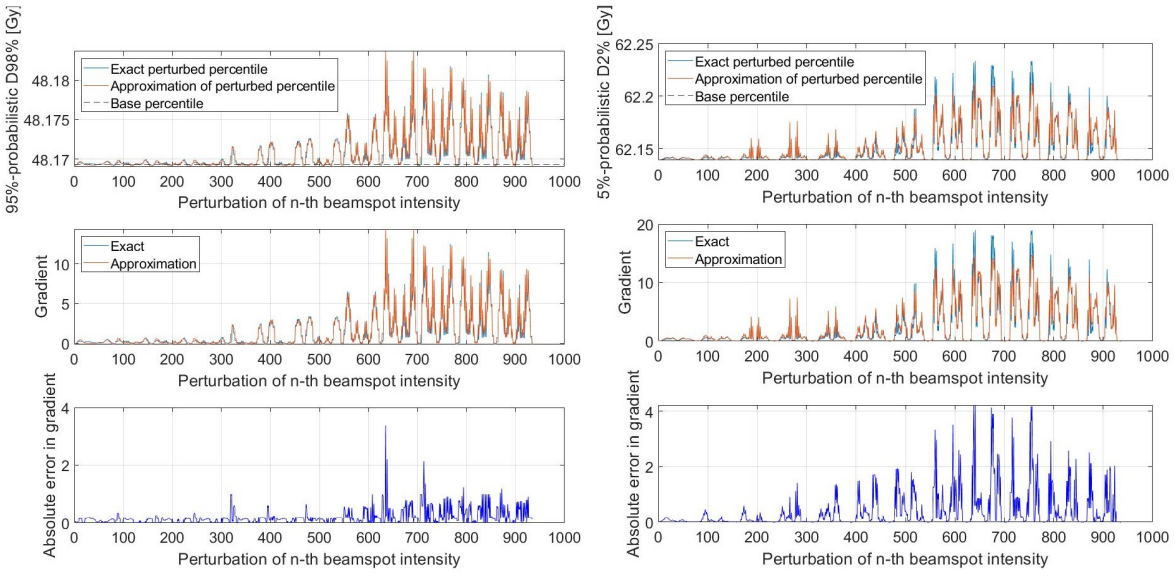
difficult to find in a robust manner. Therefore, instead of the two distinct percentile argument averaging an approximation method is tested in which the outcome of each PCE with a perturbed beam intensity vector is averaged over a number of arguments that yield values closest to $d^{(\gamma)}_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ for the PCE that is constructed with the base beam intensity vector. The idea is that then the inaccuracies are mitigated when averaging over multiple arguments. Figures 7.13, 7.14 and 7.15 show the results of this approximation method on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ for the 4, 6 and 8 nearest argument averaging methods respectively. Here, a different set of Gaussian samples is used compared to Figures 7.10 and 7.12.



(a) Results of 95%-probabilistic $d_{98\%}(\boldsymbol{x})$.

(b) Results of 5%-probabilistic $d_{2\%}(\boldsymbol{x})$.

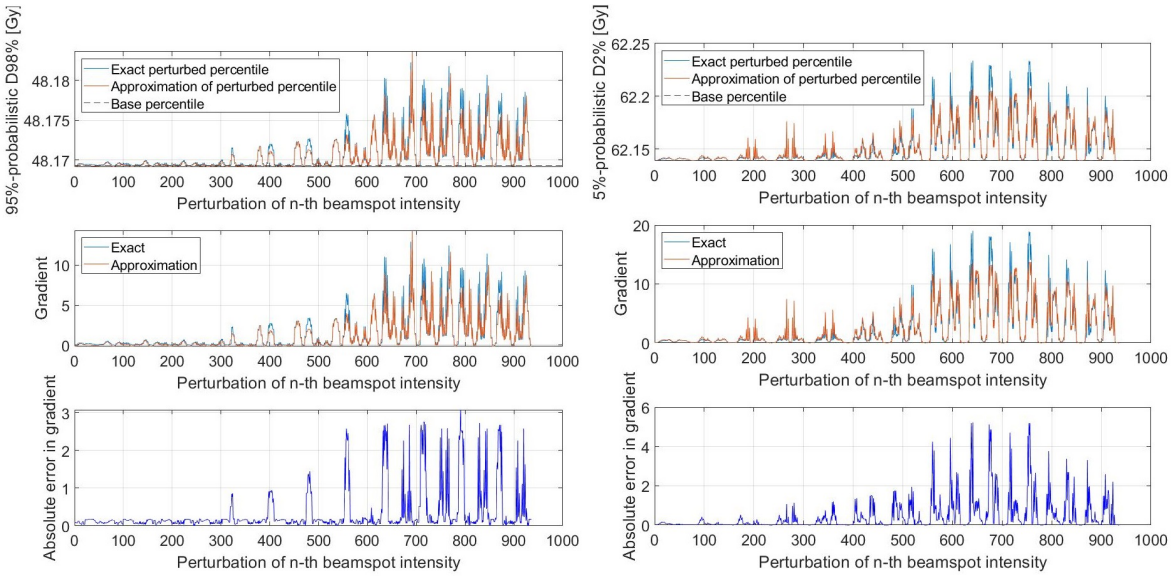
Figure 7.13: Results of the 4 arguments averaging approximation method on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ **(a)** and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ **(b)**. In the top images, the base percentile value is shown as the black dotted line and the exact and approximated perturbed percentile values are drawn as blue and red solid lines respectively. The middle images show the exact and approximated values of the gradient and the bottom images show the absolute errors in the gradient. The step size is chosen as $h = 0.005$ and the sample size as $N_s = 500,000$



(a) Results of 95%-probabilistic $d_{98\%}(\boldsymbol{x})$.

(b) Results of 5%-probabilistic $d_{2\%}(\boldsymbol{x})$.

Figure 7.14: Results of the 6 arguments averaging approximation method on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ **(a)** and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ **(b)**, similarly to Figure 7.13.

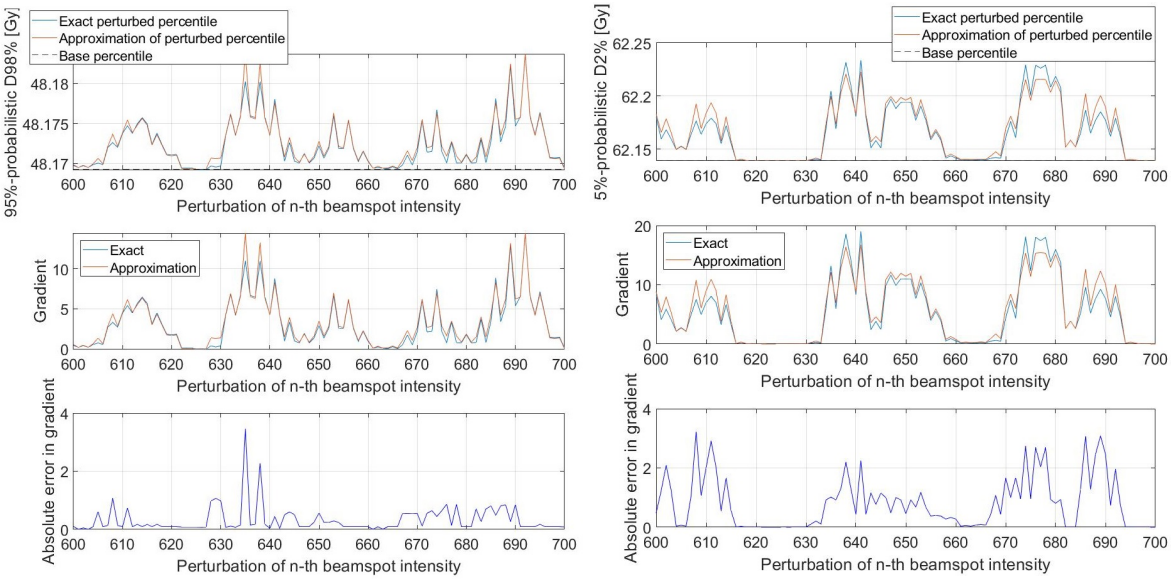


(a) Results of 95%-probabilistic $d_{98\%}(\boldsymbol{x})$.

(b) Results of 5%-probabilistic $d_{2\%}(\boldsymbol{x})$.

Figure 7.15: Results of the 8 arguments averaging approximation method on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ **(a)** and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ **(b)**, similarly to Figures 7.13 and 7.14.

In Figures 7.13, 7.14 and 7.15 the approximated percentile values are for most of the 'perturbed' PCEs close to the exact percentile values for both the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and the 5%-probabilistic $d_{2\%}(\boldsymbol{x})$, meaning that the absolute errors in the gradient are small compared to the actual gradient values. Also, from these three figures it is difficult to conclude for which number of arguments the averaging method leads to more accurate approximations. The accuracy will also differ for different sets of samples, and cannot be predicted. An important result in the accuracy of the approximated gradient is that the peaks are still present at the same locations as for the exact gradient, only having a slightly different height in some of the peaks. Figure 7.16 shows this result as a zoomed-in version of Figure 7.13 for a subset of 100 beamspots.



(a) Results of 95%-probabilistic $d_{98\%}(\boldsymbol{x})$.

(b) Results of 5%-probabilistic $d_{2\%}(\boldsymbol{x})$.

Figure 7.16: Results of Figure 7.13, zoomed-in at a subset of 100 beamspots.

Unfortunately, this approximation method is quite heuristic. It is not possible to predict the accuracy of the approximation of the perturbed percentile values, because the accuracy depends largely on the geometry,

the beamspot grid, the defined uncertain parameters, the beam intensities, the step size and the considered values of α and β in the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$. However, the results in Chapter 7.4.1 show quite a potential in the approximation method in terms of accuracy, and definitely in terms of computation time. Therefore, in Chapter 7.5 the approximation method is put into practice in optimizations on the 3D spine geometry. The purpose of the optimizations is to investigate whether a decrease in treatment plan optimization time can be achieved, and to investigate possible issues that might occur due to inaccuracies in the gradient and Hessian.

7.5. PROBABILISTIC PLANNING WITH PROBABILISTIC DOSE VOLUME PARAMETERS

In this last section of Chapter 7, optimizations on the 3D spine geometry are performed in which the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is used. As this has not been done before in Matlab, optimizations are first performed with the exact percentile calculations in Chapter 7.5.1. Here, the differences in performance and computation time for different sample sizes are investigated. Similarly, in Chapter 7.5.2 the results are shown for the same optimization problem, but with the approximation method used for the percentile calculations with different cases for the number of arguments for averaging.

One type of optimization problem is treated, namely case 3 of Table 6.1. In this optimization problem, just as in cases 1 and 2, the expected value of the quadratic differences between the prescribed dose and the delivered dose in the full discretized domain is minimized as objective function, and a constraint is set on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor to be greater than or equal to 50 Gy. As a comparison, in the optimized treatment plans of cases 1 and 2 the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor is 48.2 Gy and 48.3 Gy respectively. The goal of this constraint is to provide a better dose coverage in the tumor in the extreme error scenarios, resulting in a larger automatic PTV expansion in the direction of the setup uncertainty, than in cases 1 and 2.

The optimizations are again performed on the HPC cluster with 6 processors and 50 GB of storage memory available on one computer. The *fmincon* solver in Matlab is used with the optimization algorithm chosen as the interior-point algorithm. The settings for the prescribed dose and voxel weight in each structure, the considered uncertain parameter, the PCE characteristics, the step size for the calculation of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for perturbed beam intensities and the starting beam intensity vector are shown in Table 7.5.

Table 7.5: Parameter settings for the optimizations of case 3.

Prescribed dose:	Structure:	Value:	
	Tumor	60 Gy	
	Spine	0 Gy	
	Tissue	0 Gy	
Voxel weight:	Structure:	Value:	
	Tumor	500	
	Spine	500	
	Tissue	1	
Uncertain parameter:	Type:	Distribution:	
	Set-up uncertainty in x-direction	$\mathcal{N}(0, 0.3^2)$ [cm, cm^2]	
PCE:	Object:	Characteristic:	Cut-off value:
	All PCEs for objective function	$L5E0O1$	10^{-3}
	All PCEs for constraint	$L7E1O8$	10^{-3}
Step size:	Value:		
	0.005		
Starting vector:	Type:		
	Random values in range 0 - 0.4 MU		

Almost all parameters that coincide with cases 1 and 2 are kept the same in case 3. Only the starting vec-

tor is here chosen as a vector with random values between 0 and 0.4 MU, for which the value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ is in the feasible region of the constraint. Choosing the starting vector as the zero vector or the optimized beam intensity vector from either case 1 or 2 unfortunately results in starting at the infeasible region of the constraint. The range 0-0.4 MU is chosen, because this is the range of the beam intensities in cases 1 and 2. The same realization of this random vector is used for all optimizations of case 3, the starting value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ being 303 Gy.

In order to investigate the accuracy of the PCE of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ for different PCE characteristics, Figure 7.17 compares the exact values and the PCE approximations for various grid levels and polynomials orders. The beam intensity vector is obtained from the optimized treatment plan of case 1. The interval of setup error scenarios in Figure 7.17 is chosen as $[-1.2, 1.2]$ in centimeters, covering 99.9% of the underlying Gaussian sample space.

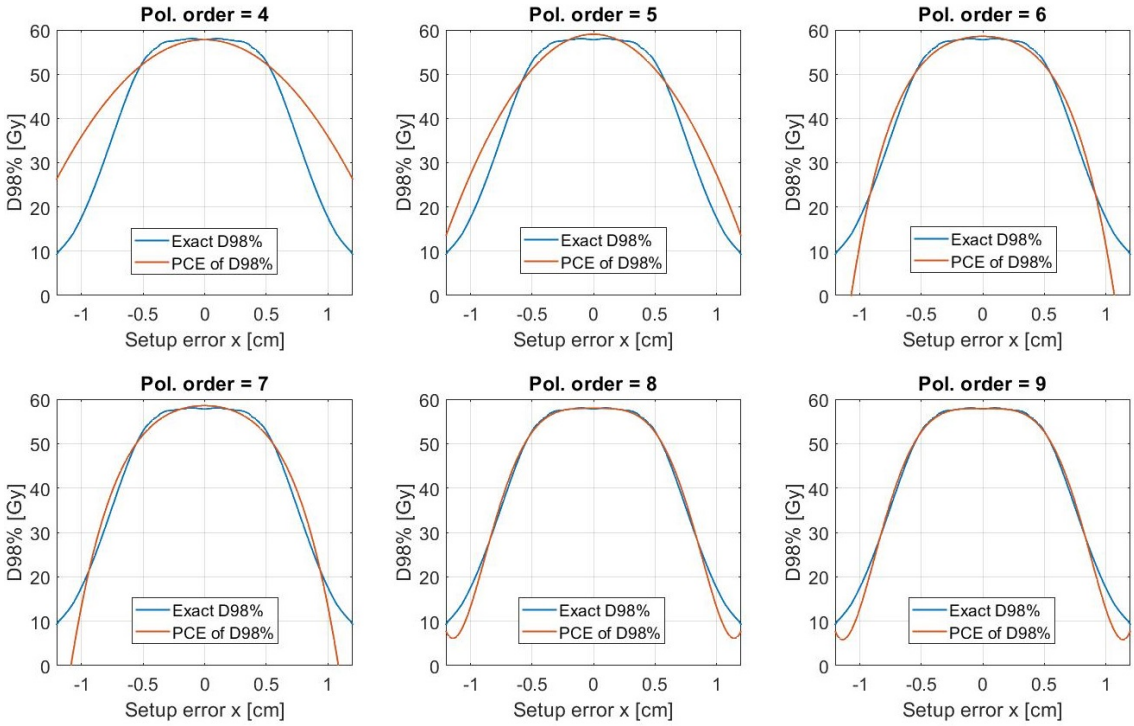


Figure 7.17: Exact values and PCE approximations of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ for increasing grid levels and polynomial orders. The PCE characteristics are $LaE1Ob$ for polynomial orders $b \in \{4,5,6,7,8,9\}$ and grid levels $a = b-1$.

From the results in Figure 7.17, the PCE characteristic $L7E1O8$ is chosen for all PCEs that need to be constructed for the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$. This PCE approximation is accurate up to almost 3 times the standard deviation of the setup uncertainty, covering 99.7 % of the underlying Gaussian sample space and thus also the 5-th percentile of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$, which we are interested in.

Furthermore, a step size needs to be chosen for the construction of PCEs of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the perturbed beam intensities. In order to compare the accuracy of the gradient calculation for different step sizes, Figure 7.18 shows for a subset of 100 beamspots the gradient values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for 10 different realizations of sets of 500,000 samples for various step sizes. The beam intensity vector is again obtained from the optimized treatment plan of case 1.

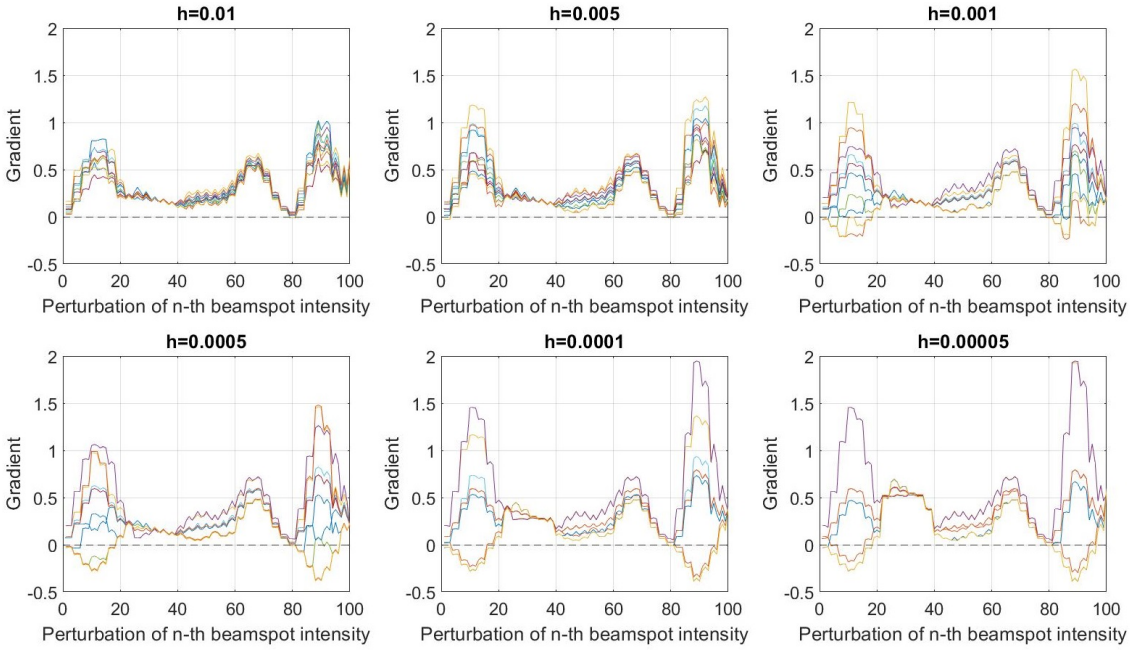


Figure 7.18: Values of the gradient of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for a subset of 100 beamspots for various step sizes. Per step size the results of 10 different realizations of sets of 500,000 samples are shown. The values of the gradient are connected for clarity.

The results in Figure 7.18 show decreasing accuracy in the gradient calculation for smaller step sizes. Also, the differences in the gradient between different realizations is larger for smaller step sizes. This inconsistency originates from the fact that the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ is a stochastic variable. The consistency deteriorates for smaller step sizes, because of two reasons. The first reason is that errors in the PCE construction are more prevalent when comparing the differences between the PCE for the base intensities and a PCE for perturbed intensities for smaller step sizes. The second reason is that through the finite difference scheme in Equation 5.12 the errors in the gradient blow up for smaller step sizes. From these results, the step size $h = 0.005$ is chosen for the optimizations, because the gradient values are relatively consistent across the realizations and because a better accuracy of the approximation method is expected for smaller step sizes.

7.5.1. EXACT METHOD

The exact method is used in three different optimizations, subcases 3.1, 3.2 and 3.3 of case 3. In these three subcases the number of samples for the calculation of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the base and perturbed beam intensity vectors is varied, as shown in Table 7.6.

Table 7.6: The three different subcases for optimizations of case 3 with the exact method.

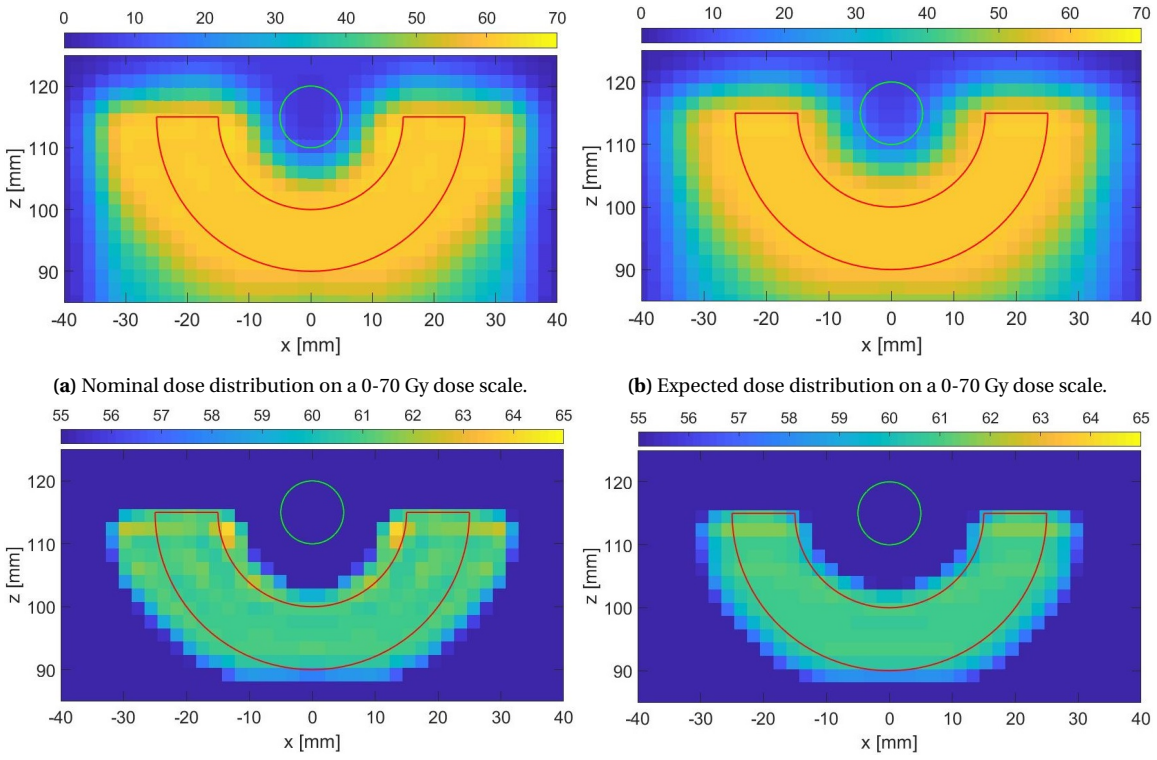
Case:	Number of samples:
3.1	50,000
3.2	250,000
3.3	500,000

The optimizations of cases 3.1, 3.2 and 3.3 took 49.3, 111.2 and 125.3 hours with 197, 139 and 73 iterations respectively. With the defined settings, one iteration of the optimization takes approximately 15, 48 and 103 minutes for cases 3.1, 3.2 and 3.3 respectively. The differences in computation time between the three subcases and the optimized constraint values are displayed in Table 7.7.

Table 7.7: Computation times of one iteration and the full optimization, and the number of iterations for the three probabilistic plans of the subcases of case 3 with the exact method.

Case:	Iteration:	Number of iterations:	Optimization:
3.1	15 min	197	49.3 h ≈ 2.1 days
3.2	48 min	139	111.2 h ≈ 4.6 days
3.3	103 min	73	125.3 h ≈ 5.3 days

The results in Table 7.7 show a decrease in the number of iterations for the optimizations when a larger sample size is used. This is due to a more accurate calculation of the value, gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$. However, due to the longer computation times for larger sample sizes, the optimization of case 3.1 results in the shortest computation time for the three subcases with the exact method. Apparently, with the inaccuracies in the value, gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for smaller step sizes the optimizations are still able to converge with the interior-point algorithm. The nominal and expected dose distributions of the optimized treatment plan of case 3.1 are shown in Figure 7.19.



(a) Nominal dose distribution on a 0-70 Gy dose scale.

(b) Expected dose distribution on a 0-70 Gy dose scale.

(c) Nominal dose distribution on a 55-65 Gy dose scale.

(d) Expected dose distribution on a 55-65 Gy dose scale.

Figure 7.19: Axial cross-sections of the 3D spine geometry at the plane corresponding to $y = 5$ mm with the nominal and expected dose distributions on a 0-70 Gy dose scale, **(a)** and **(b)** respectively, and on a 55-65 Gy dose scale, **(c)** and **(d)** respectively, for the optimized treatment plan of case 3.1. All voxels with a dose less than 55 Gy are given the same color as the 55 Gy boundary for the 55-65 Gy scale.

The dose distributions in Figure 7.19 show good results in terms of dose coverage in the tumor in the nominal scenario and also for the expected dose with an automatic expansion of the PTV in the direction of the set-up uncertainty, as expected. The resulting nominal and expected dose distributions are more uniform than in the optimizations of cases 1 and 2. Also, the PTV expansion is larger than in cases 1 and 2 due to the defined constraint on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor. At the same time, the spine is still nicely spared due to the defined voxel weights in the objective function, but receives more dose than in cases 1 and 2. In Figure 7.20, dose volume histograms are shown for the nominal and expected dose distributions in the tumor, spine and tissue for case 3.1 to fully review the optimized treatment plan.

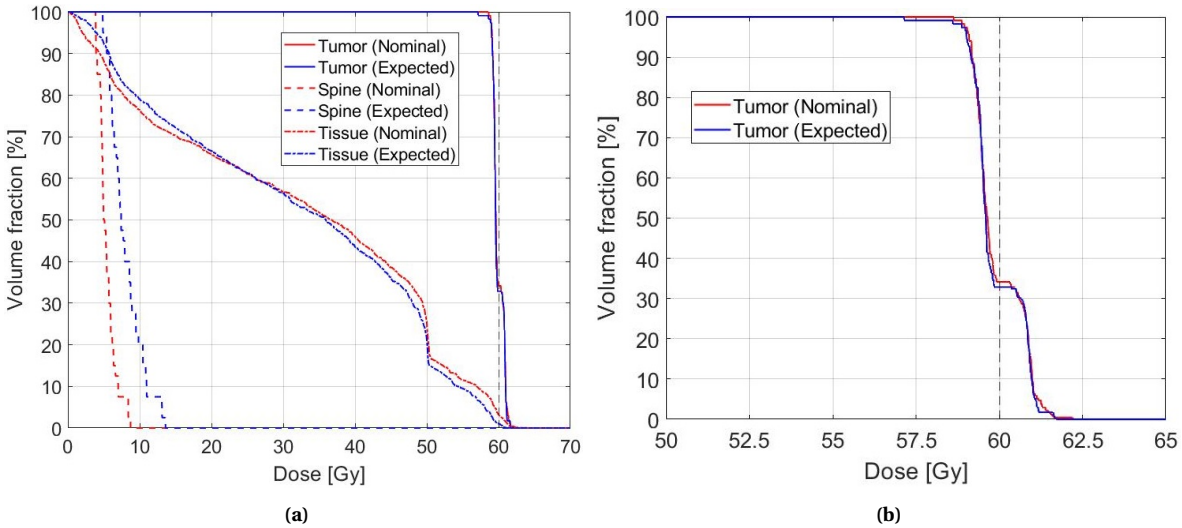


Figure 7.20: Dose volume histograms of the nominal and expected dose in the tumor, spine and tissue **(a)** and only the tumor **(b)** for the optimized treatment plan of case 3.1.

The dose volume histograms in Figure 7.20 show similar differences between the nominal and expected dose distribution as in cases 1 and 2. However, in case 3.1 the dose levels in the tumor are much closer to the prescribed dose of 60 Gy than in cases 1 and 2. On the other hand, the dose levels in the spine and tissue are higher in case 3.1 than in cases 1 and 2. This optimization shows nicely the possibility of improving the dose distribution in the tumor when using probabilistic dose volume parameters in treatment planning. The optimized treatment plan has clinically more relevance due to the use of probabilistic dose volume parameters compared to cases 1 and 2, however, the computation time for case 3.1 is much longer.

Possible ways to reduce the dose in the spine is to set a constraint on a probabilistic dose volume parameter in the spine or to increase the voxel weights of the spine in the objective function. To compare the optimized treatment plans of cases 3.1-3.3, Figure 7.21 shows dose volume histograms of the nominal and expected dose distributions in the tumor, spine and tissue for the three subcases.

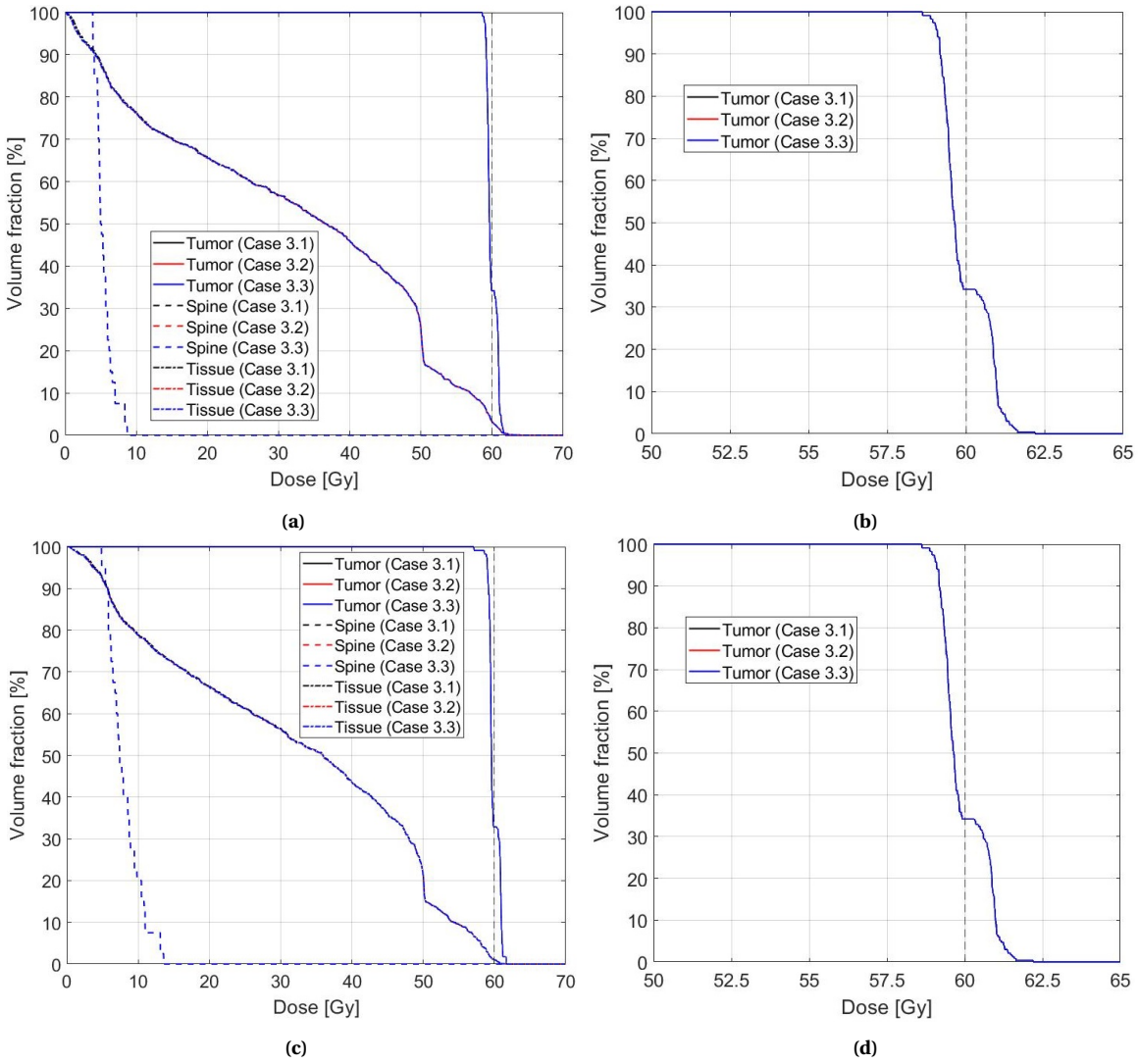
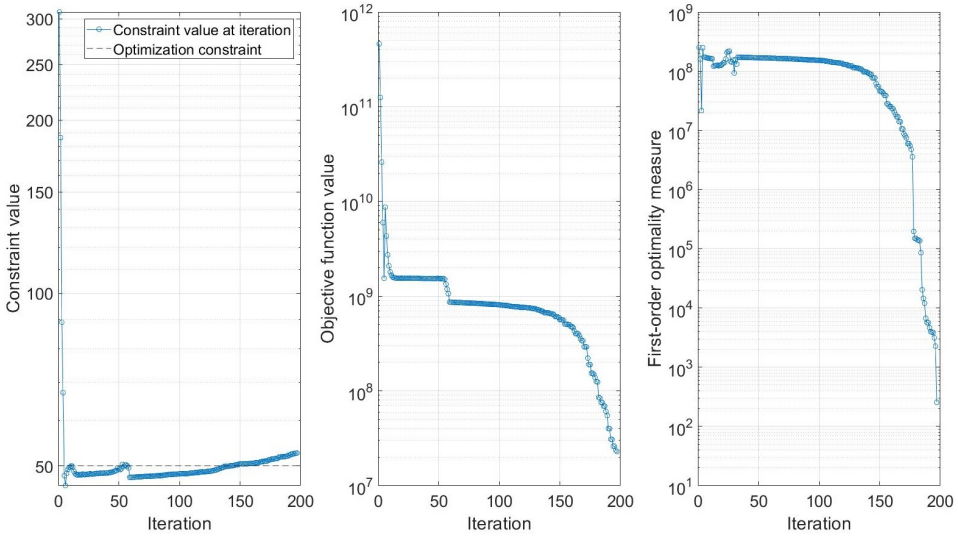


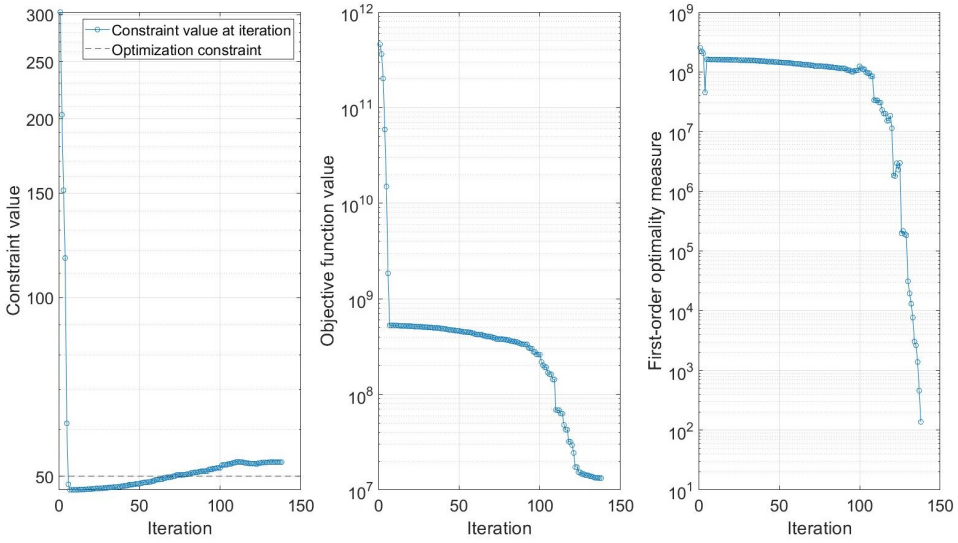
Figure 7.21: Dose volume histograms of the nominal and expected dose to the tumor, spine and tissue ((**a**) and (**c**)) and only the tumor ((**b**) and (**d**)) for the optimized treatment plans of cases 3.1-3.3.

The dose volume histograms in Figure 7.21 show that the optimized treatment plans in cases 3.1, 3.2 and 3.3 are almost identical. There is only a very small difference of approximately 0.2 Gy in the nominal dose in the tissue in Figure 7.21a between case 3.1 and the two other cases in the 0-5 Gy dose range. This is not a significant difference in the outcome of the optimizations and could be due to a less accurate calculation of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in case 3.1.

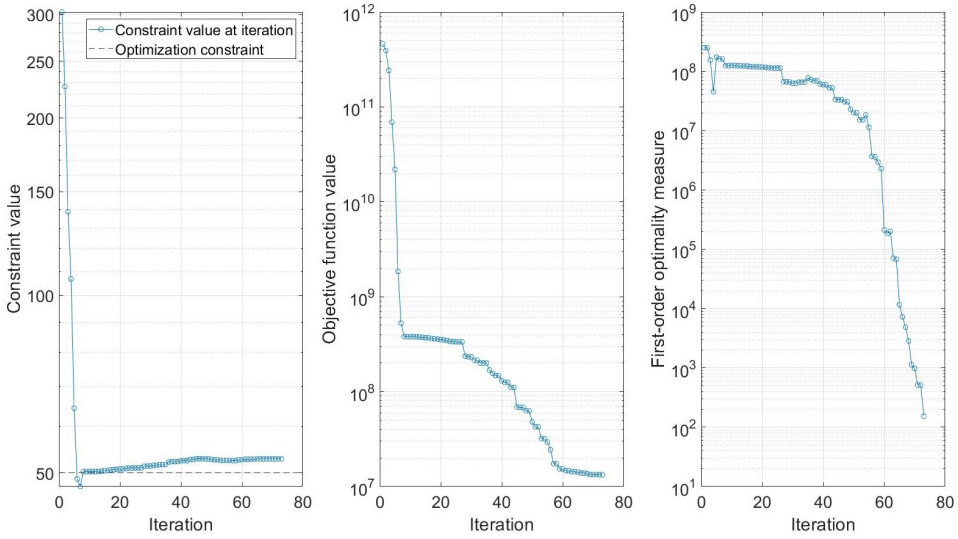
As a difference in the number of iterations is found in cases 3.1-3.3, Figure 7.22 provides for these three cases the value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor, the value of the objective function, and the value of the first-order optimality measure per iteration of the treatment plan optimization. The first-order optimality measure is a measure for the gradient of a constrained optimization problem and is calculated by the *fmincon* solver [46].



(a) Case 3.1.



(b) Case 3.2.



(c) Case 3.3.

Figure 7.22: Values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor (left), the objective function (middle) and the first-order optimality measure (right) per iteration during the optimization of the probabilistic treatments plans of cases 3.1 **(a)**, 3.2 **(b)** and 3.3 **(c)**.

The optimization data per iteration in Figure 7.22 show that for this starting vector with random values the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ is approximately 303 Gy. Then, the first 5-10 iterations the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ decreases rapidly towards the 50 Gy boundary, but also crosses this value. For the interior-point algorithm in the *fmincon* solver the constraint boundary is not an active constraint and the optimization is continued with an internal message that the constraint boundary has been violated. In cases 3.1 and 3.2, the solver takes many iterations for the value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ to get back to the feasible region, whereas in case 3.3 the feasible region is quickly found. This difference might be due to inaccuracies in the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ when smaller sample sizes are used. In cases 3.1-3.3, the optimizations stop at a point in the interior of the feasible constraint region, resulting in 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ values of 52.89, 52.78 and 52.72 Gy respectively. The solutions show a slight difference in the optimized 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ due to the inaccuracies for smaller sample sizes. At these solutions, the default value of the relative tolerance on the first-order optimality measure of 10^{-6} is violated. The value of the first-order optimality measure does not yet reach a value close to zero, as is expected for the gradient at an optimum, but this is due to the relatively large starting value of the first-order optimality measure for the starting beam intensity vector with random values and the chosen tolerance.

The solutions are not found at the boundary of the constraint of 50 Gy for the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor, but in the interior of the feasible region. A possible reason for this is that the optimizations in cases 3.1-3.6 with the constraint on the beam intensities to be non-negative and the constraint on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor result in an optimization over a non-convex set, even though the objective function is convex. To verify that the optimizations did converge, case 3.3 has been optimized again twice. The first optimization is performed with the same tolerance level on the first-order optimality measure, but starting at the solution of case 3.3. The second optimization is performed with a starting vector that is a slightly perturbed version of the solution of case 3.3. This perturbation is chosen as a decrease in the intensities of 0.5%. Both optimizations find the same solution as in case 3.3 within 5 iterations. From this, we can conclude that the optimizations did converge and that the found solution is a local minimum.

7.5.2. APPROXIMATION METHOD

The approximation method is also used in three different optimizations, namely subcases 3.4, 3.5 and 3.6 of case 3. In these three subcases, the number of samples for the calculation of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the base and perturbed beam intensity vectors is chosen as $N_s = 10^7$, but the number of arguments that is used for averaging is varied. Table 7.6 summarizes these three subcases.

Table 7.8: The three different subcases for optimizations of case 3 with the approximation method.

Case:	Number of samples:	Number of arguments for averaging:
3.4	10^7	4
3.5	10^7	6
3.6	10^7	8

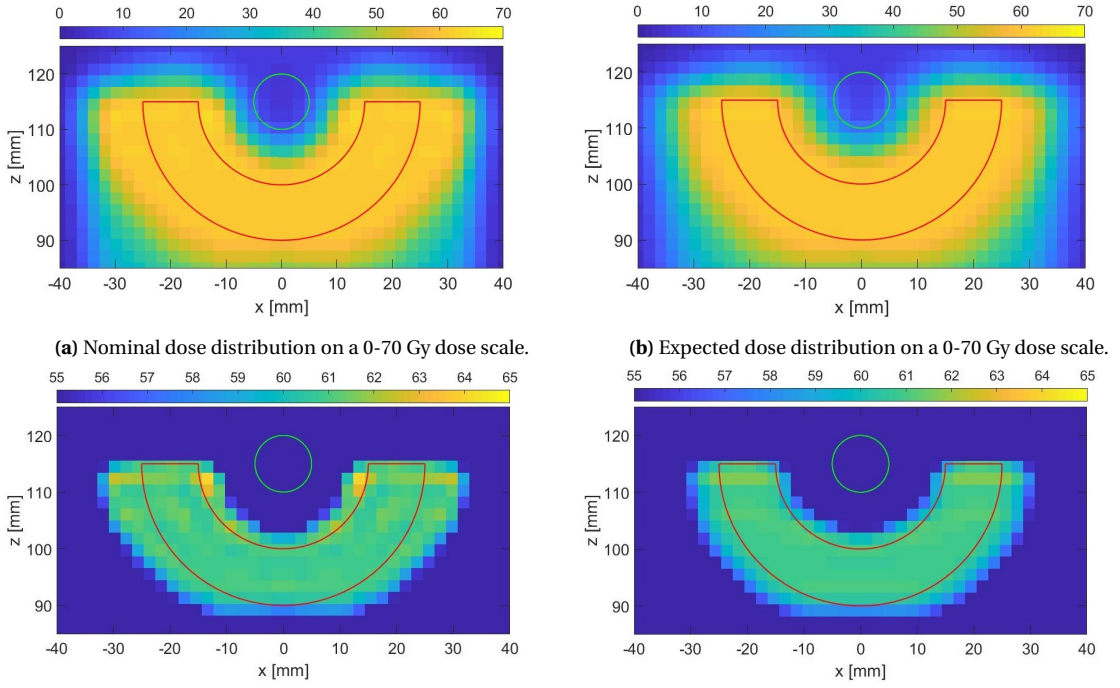
The optimizations of cases 3.4, 3.5 and 3.6 took 16.5, 16.6 and 16.3 hours with 190, 191 and 187 iterations respectively. For the defined settings one iteration of the optimization takes approximately 313 seconds. The differences in computation time between the three subcases are displayed in Table 7.9.

Table 7.9: Computation times of one iteration and the full optimization, and the number of iterations for the three probabilistic plans of the subcases of case 3 with the approximation method.

Case:	Iteration:	Number of iterations:	Optimization:
3.4	313 s	190	16.5 h
3.5	313 s	191	16.6 h
3.6	313 s	187	16.3 h

The results in Table 7.9 show that for the optimizations of cases 3.4-3.6 take almost the same number of iterations. The computation times per iteration are much shorter than in cases 3.1-3.3, and therefore, the total optimization time is shorter for the approximation method than for the cases with the exact method. Apparently, with the inaccuracies in the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ due to the

approximation method the optimizations are still able to converge with the interior-point algorithm. Figure 7.23 shows the nominal and expected dose distributions of the optimized treatment plan of case 3.4.



(a) Nominal dose distribution on a 0-70 Gy dose scale.

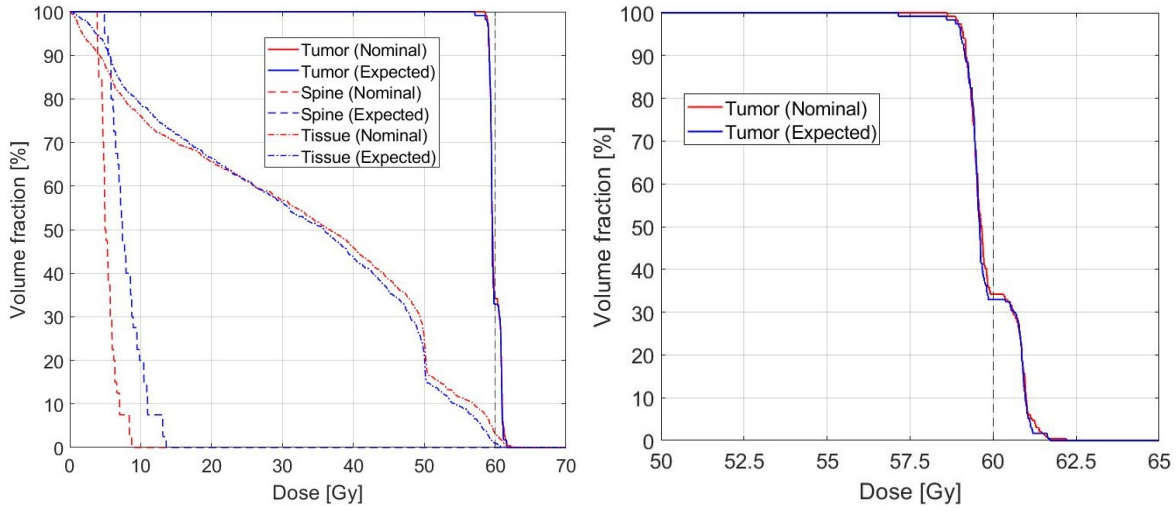
(b) Expected dose distribution on a 0-70 Gy dose scale.

(c) Nominal dose distribution on a 55-65 Gy dose scale.

(d) Expected dose distribution on a 55-65 Gy dose scale.

Figure 7.23: Axial cross-sections of the 3D spine geometry at the plane corresponding to $y = 5$ mm with the nominal and expected dose distributions on a 0-70 Gy dose scale, **(a)** and **(b)** respectively, and on a 55-65 Gy dose scale, **(c)** and **(d)** respectively, for the optimized treatment plan of case 3.4. All voxels with a dose less than 55 Gy are given the same color as the 55 Gy boundary for the 55-65 Gy scale.

The dose distributions in Figure 7.23 show good results in terms of dose coverage of the tumor in the nominal scenario and also for the expected dose with an automatic expansion of the PTV in the x-direction. The dose distributions are similar to the results in Figure 7.19. In Figure 7.24, dose volume histograms are shown for the nominal and expected dose distributions in the tumor, spine and tissue for case 3.4.



(a)

(b)

Figure 7.24: Dose volume histograms of the nominal and expected dose in the tumor, spine and tissue **(a)** and only the tumor **(b)** for the optimized treatment plan of case 3.4.

The dose volume histograms in Figure 7.24 show similar nominal and expected dose distributions in the tumor, spine and tissue as in cases 3.1-3.3. To compare the optimized treatment plans of cases 3.4-3.6, Figure 7.25 shows dose volume histograms of the nominal and expected dose distributions in the tumor, spine and tissue of the three cases.

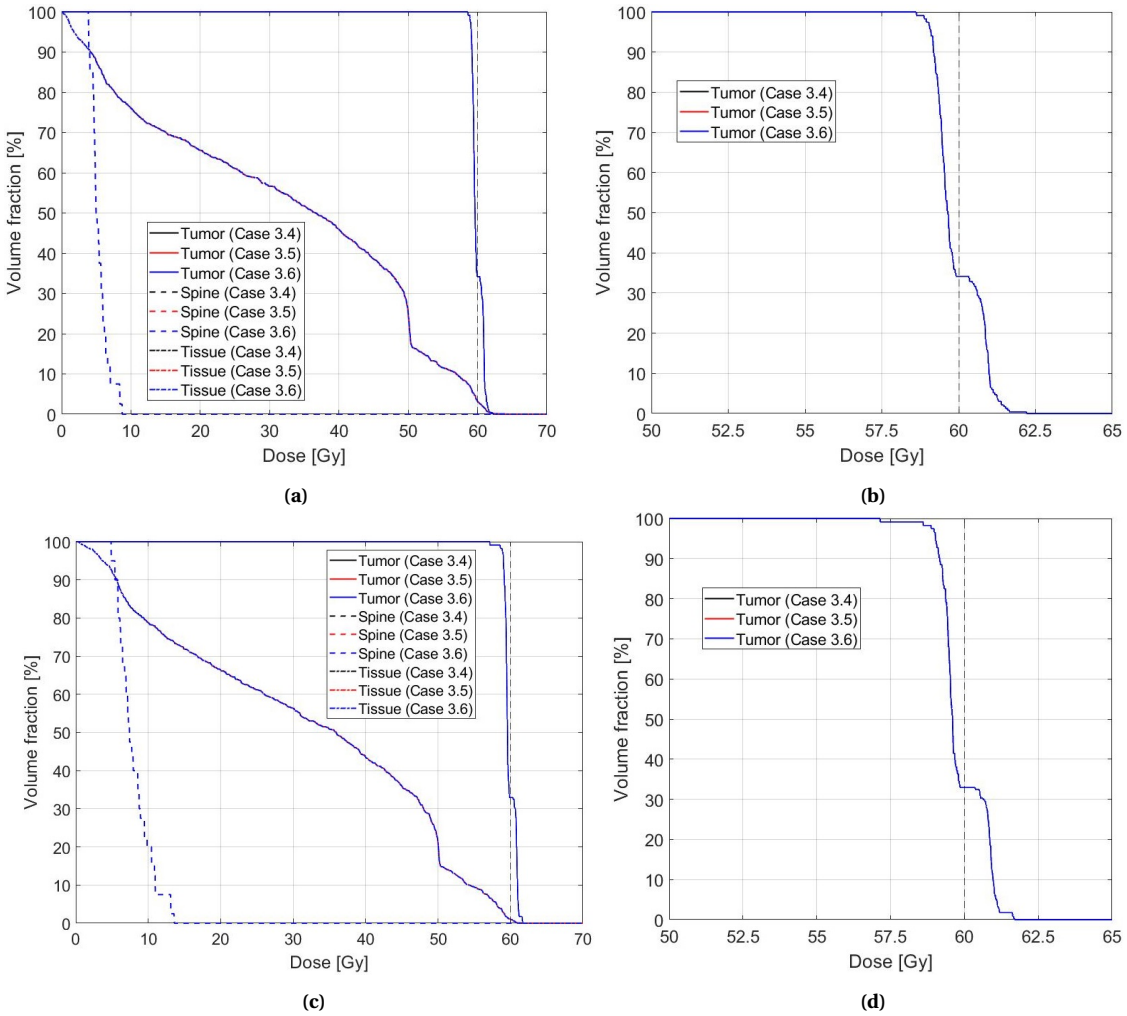
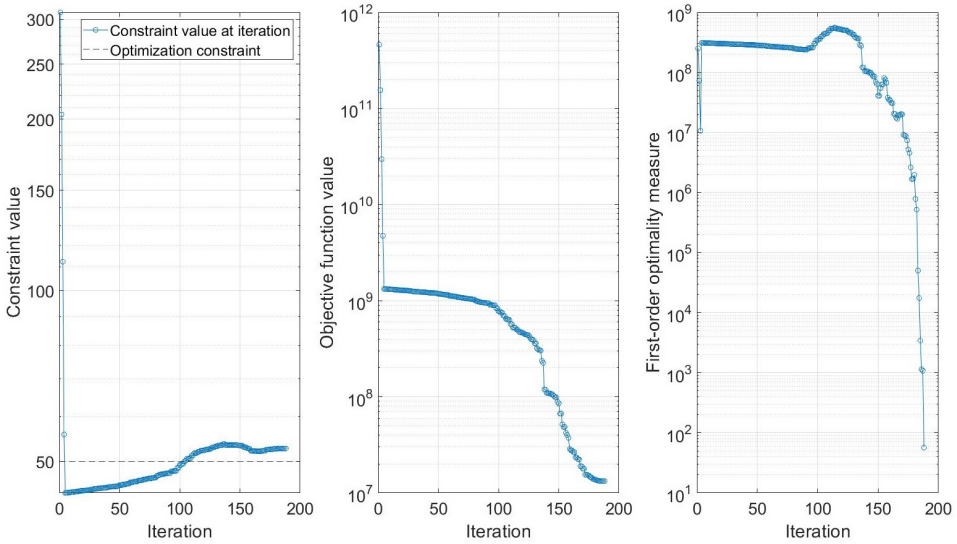
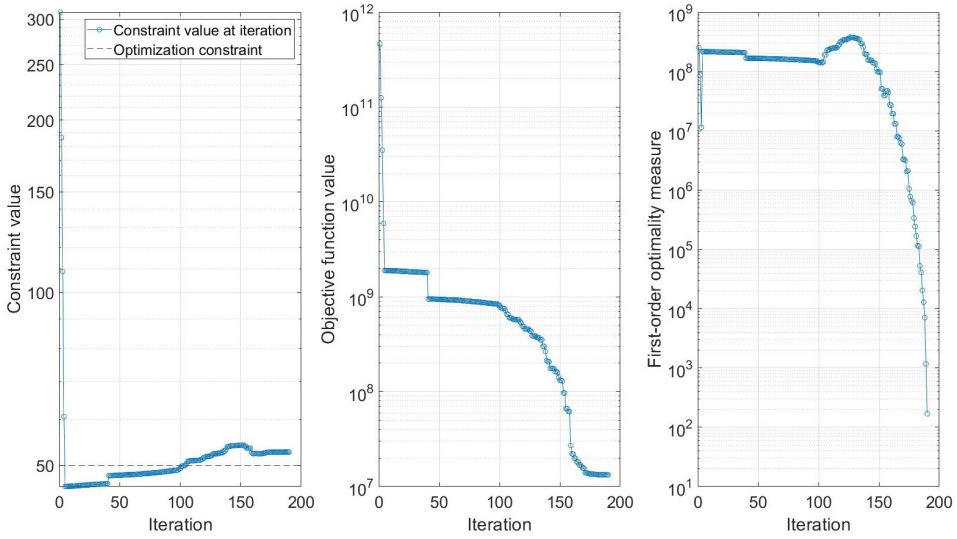


Figure 7.25: Dose volume histograms of the nominal and expected dose to the tumor, spine and tissue ((**a**) and (**c**)) and and only the tumor ((**b**) and (**d**)) for the optimized treatment plan of cases 3.4-3.6.

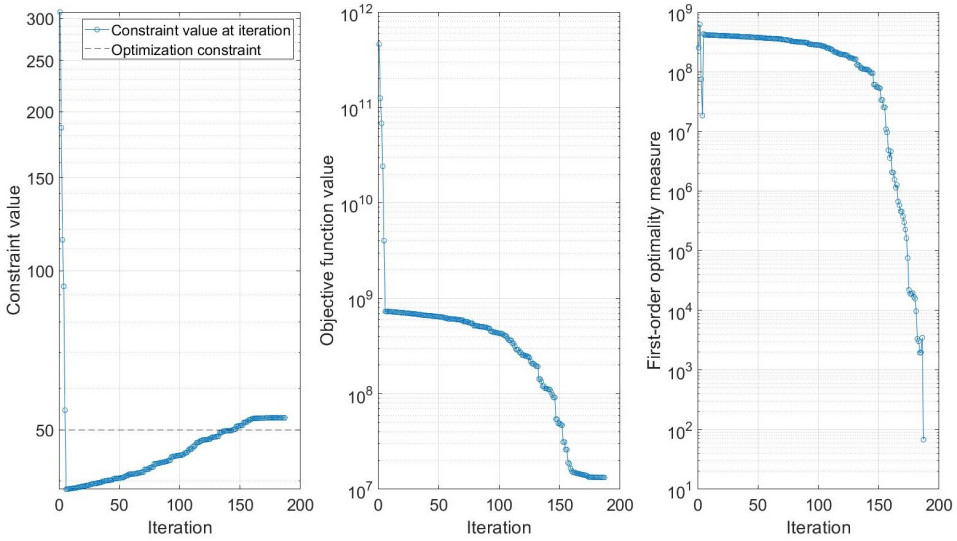
The dose volume histograms in Figure 7.25 show that the optimized treatment plans in cases 3.4-3.6 are identical. Looking at the optimization times and the number of iterations of the three cases, it seems that the approximation method does provide the same results in terms of performance for the 4, 6 and 8 point averaging cases. Figure 7.26 shows for cases 3.4, 3.5 and 3.6 the objective function value, value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and the first-order optimality measure to see what happens during the optimizations.



(a) Case 3.4.



(b) Case 3.5.



(c) Case 3.6.

Figure 7.26: Values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor (left), the objective function (middle) and the first-order optimality measure (right) per iteration during the optimization of the probabilistic treatments plans of cases 3.4 **(a)**, 3.5 **(b)** and 3.6 **(c)**.

The optimization data per iteration in Figure 7.22 show that in cases 3.4-3.6 the solver takes many iterations for the value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ to get back to the feasible region, just as in cases 3.1 and 3.2. The difference from case 3.3 is again the inaccuracies in the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ when the approximation method is used. The optimizations of cases 3.4-3.6 all converge to the same local minimum in the interior of the feasible region of the constraint at a value of 52.73 Gy for the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$. These optimized constraint values are different than in cases 3.1-3.3 due to the difference in sample size in the constraint calculations. Figure 7.27 shows dose volume histograms of the nominal and expected dose in the tumor, spine and tissue for cases 3.3 and 3.4 to compare the optimized treatment plans with the exact method and the approximation method.

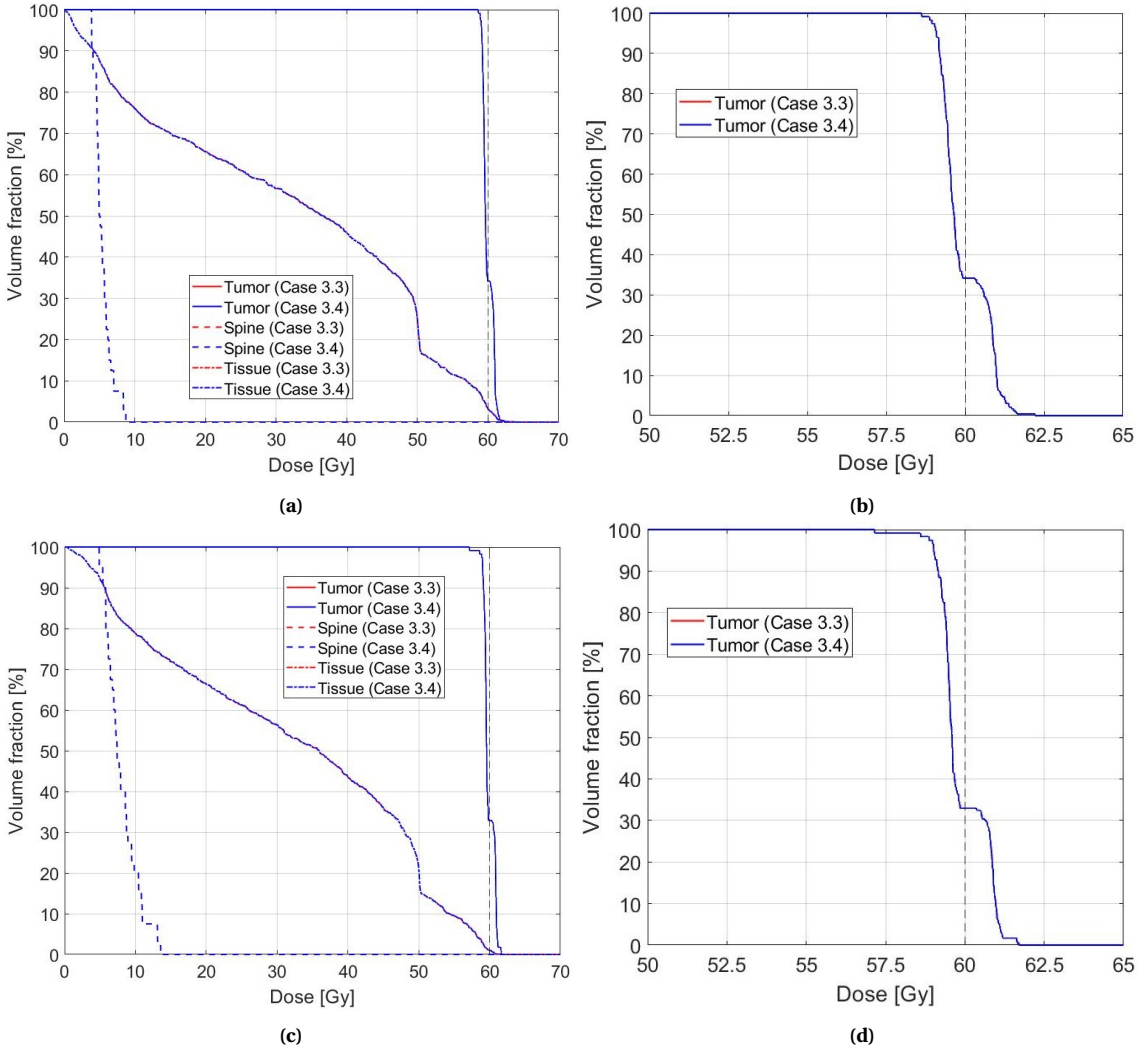


Figure 7.27: Dose volume histograms of the nominal and expected dose in the tumor, spine and tissue ((**a**) and (**c**)) and and only the tumor ((**b**) and (**d**)) for the optimized treatment plan of cases 3.3 and 3.4.

The dose volume histograms in Figure 7.27 show that case 3.3, with the largest sample size of 500,000 for the exact method, and cases 3.4-3.6 with the approximation method yield the same solution. The optimized values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ are most accurate in cases 3.4-3.6 due to the sample size of 10^7. In cases 3.4-3.6, the approximations of the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ lead to a higher number of iterations compared to case 3.3, but the computation times are much longer in case 3.3. The optimizations in cases 3.4-3.6 yield approximately a factor of 7.6 decrease in computation time compared to case 3.3. As a conclusion, good improvement in computation time is shown for probabilistic treatment planning on the 3D spine geometry with the use of probabilistic dose volume parameters.

8

DISCUSSIONS AND CONCLUSIONS

This chapter focuses on the discussion of the results in this research, the recommendations for future research and the main conclusions of this research.

8.1. DISCUSSIONS

In Chapter 8.1.1 the results are discussed per topic and in Chapter 8.1.2 recommendations for future research are mentioned, based on the results.

8.1.1. RESULTS

The results are divided into five parts. The first topic is the exact improvement on Algorithm 1. The second topic is the derived monotonicity constraint on the PCE coefficients. The third topic is the approximation method for probabilistic dose volume parameters. The fourth and fifth topic are probabilistic treatment planning without and with the use of probabilistic dose volume parameters respectively.

EXACT IMPROVEMENTS

In this research, three improvements on Algorithm 1 in terms of computation time are implemented that yield the same outcome as the methodology used in [12]. In total, a factor of 12 decrease in computation time is found for the 3D spine geometry from approximately 2,800 seconds to 230 seconds, when the percentile calculations are not considered. For a sample size of 500,000 in the percentile calculations and the use of the PCE characteristic $L5E1O6$, the total reduction in computation time for Algorithm 1 is found to be from 8,300 to 5,800 seconds. This shows a 30% speed-up in computation time. For case 3.3 in Chapter 7.5 with a sample size of 500,000, this would also approximately yield a 30% speed-up for the full optimization of the treatment plan. The actual speed-up in other treatment plan optimizations depends on the dose parameters in the objectives and constraints, and their computation times. Nevertheless, the goal in this research to reduce the computation time in probabilistic planning, in which probabilistic dose volume parameters are used, is definitely satisfied.

MONOTINICITY CONSTRAINT ON PCE COEFFICIENTS

In Chapter 6.3, a constraint on the PCE coefficients has been derived to ensure monotonicity of probabilistic dose volume parameters with respect to the beam intensity vector in the case that an infinite amount of basis vectors is used. In practice, a finite number of basis vectors is used, which results in possible errors in the approximation of the stochastic response. However, this constraint still improves the accuracy of probabilistic dose volume parameters for perturbed beam intensities in the gradient and Hessian calculations compared to the methodology used in [12].

The implementation of the constraint on the PCE coefficients shows good improvement in the monotonicity of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ for perturbed beam intensities. The corresponding gradient values become non-negative, which coincides with the physical constraint. Figure 7.8 shows that the non-monotonic behavior of PCEs of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ and $d_{2\%}(\boldsymbol{x},\boldsymbol{\xi})$ for perturbed beam intensities decreases when the number of basis functions and the polynomial orders of the basis functions in the PCE construction increase.

APPROXIMATION METHOD

An approximation method is devised for the calculation of probabilistic dose volume parameters for perturbed beam intensities. The purpose of this approximation method is to reduce the computation time of Algorithm 1 further as these calculations, with the *prctile* function in Matlab, consume around 67% of the total computation time for the 3D spine geometry and a sample size of 500,000. In this method, the percentile calculation only needs to be performed once per iteration instead of calculating percentiles for all the $1 + N_b + \frac{1}{2}\left(N_b + N_b^2\right)$ PCEs, when constructing PCEs of $d_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ as a whole. For the 3D spine geometry and a sample size of 500,000, this method reduces the computation time of the percentile calculations from 5,600 seconds to the order of only one second.

The approximation method turns out to be a heuristic method and the actual accuracy cannot be predicted in general. The accuracy depends on the geometry, the beamspot grid, the considered uncertain parameters, the beam intensity vector and the values of α and β in the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$. An important detail of the approximation method is that the actual value of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is calculated accurately with a large sample size, but the gradient and Hessian are approximated.

Results show relatively accurate approximations of the gradient of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and 5%-probabilistic $d_{2\%}(\boldsymbol{x})$ with small absolute errors. This method shows quite a potential in terms of accuracy, and definitely in terms of computation time, to alleviate the computational cost in probabilistic treatment planning for proton therapy. However, from these results alone it cannot be concluded what the effects of the gradient and Hessian approximations are on an actual treatment plan optimization.

If the approximation method turns out to be useful in practice, the accuracy of the PCEs could easily be improved through the chosen PCE characteristics without a significant increase in computation time. Also, a more accurate finite difference scheme could be used for the gradient and Hessian calculations. In that case, the number of PCEs that needs to be constructed per iteration does increase, however, the computation time of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ for perturbed beam intensities will stay considerably small. In the case that the approximation method fails to be robust enough for clinical use, still a gain could be obtained through investigation of the variance of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ as a stochastic variable for different sample sizes. Figure 7.9 shows that the variance at a certain sample size differs for different values of α and β. This could be used in the choice of the sample size for different probabilistic dose volume parameters in the objectives and constraints to reduce the computation time in the percentile calculations.

OPTIMIZATIONS WITHOUT PROBABILISTIC DOSE VOLUME PARAMETERS

In this research, eight optimizations have been performed on the 3D spine geometry, of which cases 1 and 2 in Chapter 7.1 are without and cases 3.1-3.6 in Chapter 7.5 are with the use of probabilistic dose volume parameters. Cases 1 and 2 are based on the use of the expected value as statistical measure in probabilistic treatment planning. Case 1 coincides with the optimized treatment plan in [12], and in case 2 a constraint is added on the expected mean dose in the tumor to be greater than or equal to 60 Gy. The expected value as statistical measure was in [12] found to be clinically less relevant in terms of interpretation, and therefore, as a comparison cases 3.1-3.6 are introduced.

In cases 1 and 2, the required PCEs only need to be constructed once for each optimization and the two optimizations converge fast in 11.8 and 13.9 seconds respectively. This is due to the fact that the objective function and constraint, and their gradient and Hessian, can be described analytically in terms of the zeroth order PCE coefficients and the defined geometry settings. The two treatment plans show good dose coverage in the tumor with slightly higher dose levels in case 2, as shown in Figure 7.5. Also, the results show an automatic PTV expansion in the direction of the 1D set-up uncertainty and show that the spine is nicely spared. However, there is still quite some over- and underdosage in the tumor present in the nominal and expected dose distributions. The corresponding values of the expected mean dose in the tumor are 59.6 and 60.0 Gy in cases 1 and 2 respectively. The solution of case 2 is found at the boundary of the constraint on the expected mean dose in the tumor. A possible reason for this is that the optimization in case 2 with the constraint on the beam intensities to be non-negative and the constraint on the expected mean dose in the tumor results in an optimization over a convex set, and that the objective function is convex, as claimed in [12]. The convexity of the objective function on this set still needs to be proven.

OPTIMIZATIONS WITH PROBABILISTIC DOSE VOLUME PARAMETERS

The optimization problems in cases 3.1-3.6 consist of the same objective function as in cases 1 and 2, but this time a constraint is set on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor to be greater than or equal to 50 Gy. The optimization times are much longer than in cases 1 and 2 due to the expensive computations of the

95%-probabilistic $d_{98\%}(\boldsymbol{x})$ for the perturbed beam intensities. Cases 3.4-3.6 yield the shortest optimization times of around 16.5 hours with approximately 190 iterations, whereas cases 3.1-3.3 yield longer computation times of 2.1, 4.6 and 5.3 days with 197, 139 and 73 iterations respectively. The faster percentile calculations with the approximation method in cases 3.4-3.6 outweigh the lower number of iterations in cases 3.2 and 3.3. The cases with the exact method also show faster optimization times for a sample size of 50,000 and 250,000 in cases 3.1 and 3.2 than for a sample size of 500,000 in case 3.3. The results of cases 3.1-3.6 are obtained with the implementation of the three exact improvements on Algorithm 1.

In cases 3.1-3.6, the nominal and expected dose distributions in the tumor are more uniform than in cases 1 and 2. Also, in cases 3.1-3.6 the dose levels in the tumor are closer to the prescribed dose and show a larger automatic PTV expansion in the direction of the set-up uncertainty than in cases 1 and 2. This shows better dose coverage in the tumor in the extreme error scenarios. On the other hand, the dose levels in the spine and tissue are in cases 3.1-3.6 higher than in cases 1 and 2. Possible ways to reduce the dose in the spine would be to also set a constraint on a probabilistic dose volume parameter for the spine or to increase the voxel weights of the spine in the objective function.

In cases 3.1-3.3, solutions are found at values of 52.89, 52.78 and 52.72 Gy for the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor respectively. The solutions show a slight difference in the optimized 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ due to the inaccuracies for smaller sample sizes. In cases 3.4-3.6, solutions are found at the same value of 52.73 Gy for the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$. The solutions are not found at the boundary of 50 Gy of the constraint, but in the interior of the feasible region. A possible reason for this is that the optimizations in cases 3.1-3.6 with the constraint on the beam intensities to be non-negative and the constraint on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor result in an optimization over a non-convex set, even though the objective function is convex.

The optimizations of cases 3.4-3.6 with the approximation method are still able to converge with the interior-point algorithm in the *fmincon* solver, despite the inaccuracies in the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$. Similarly, convergence is observed for cases 3.1-3.3 despite differences in accuracy in the value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$. This also introduces inaccuracies in the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the exact method, which are more prevalent for smaller sample sizes. The optimization data per iteration show that due to a more accurate calculation of the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in cases 3.2 and 3.3, the number of iterations is lower in cases 3.2 and 3.3 than in the other four cases.

Case 3.3, with the largest sample size of 500,000 for the exact method, and cases 3.4-3.6 with the approximation method yield an insignificant difference in the optimized treatment plans. The optimized values of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ are most accurate in cases 3.4-3.6 due to the sample size of 10^7. The optimizations in cases 3.4-3.6 yield approximately a factor of 7.6 decrease in computation time compared to case 3.3. This shows good improvement in computation time for probabilistic treatment planning on the 3D spine geometry with the use of probabilistic dose volume parameters.

Besides the random vector as starting vector for the optimizations of cases 3.1-3.6, also other vectors have been tested, all of which did not lead to a converged treatment plan. For example, the zero vector and the optimized beam intensity vectors from cases 1 and 2 lead to problems with convergence. This might be explained by the fact that the starting value of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ is not in the feasible region of the constraint. However, also problems were found with a starting vector in the feasible region. This starting vector was obtained from the optimization of case 1, but this time with a prescribed dose of 90 Gy in the tumor. This leads to an optimized beam intensity vector with a value of 63 Gy for the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ and a smooth dose distribution in the geometry. However, this starting vector also lead to problems with convergence. Therefore, the (default) interior-point algorithm in the *fmincon* solver does not seem robust for this optimization problem and other algorithms need to be investigated.

8.1.2. RECOMMENDATIONS FOR FUTURE RESEARCH

The discussion points about the results in this research lead to various recommendations for future research. The first recommendation is to test the exact improvements on Algorithm 1 on probabilistic treatment plan optimizations with patient data in iCycle, as performed in [12]. The exact improvements are guaranteed to alleviate the computational cost of probabilistic planning with the use of probabilistic dose volume parameters. The actual speed-up, however, should still be investigated.

The second recommendation is to test the accuracy of the approximation method further on the 3D spine geometry. Instead of considering one uncertain parameter, in future research a two or three-dimensional systematic set-up uncertainty and also range uncertainty can be introduced. Similarly, a third recommenda-

tion is to test the approximation method on a different simplified 3D geometry as well. This geometry could represent a tumor in an OAR, for example the liver or the prostate. These two recommendations will provide a better understanding of the accuracy of the approximation method in other, possibly more complex, situations before this method can be tested on treatment plan optimizations with patient data in iCycle.

A fourth recommendation is to perform the optimizations of case 3 with different optimization algorithms. In this research, only the interior-point algorithm of the *fmincon* solver in Matlab has been used and has been treated as a black box. However, it would be interesting to see which type of algorithm works best for this optimization problem and whether different results are found than in Chapter 7.5. As a fifth recommendation, in optimization problems of the same form as case 3, a constraint could also be set on a probabilistic dose volume parameter for the OAR.

Furthermore, in Appendix A the gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ are described mathematically to provide possible starting points for new approximation methods that could alleviate computational cost. A sixth recommendation is to invest time in such new approximation methods. The seventh, and final, recommendation is to investigate whether other computational programs written in C++, Python or R have faster percentile calculation algorithms and whether a fast coupling can be made to such programs. A starting point for different percentile calculation algorithms can be found in [47]. Furthermore, other research on probabilistic optimization with high-dimensional uncertainty sources that could inspire follow-up research are [48] and [49].

8.2. CONCLUSIONS

The goal of this research was to alleviate the computational cost of probabilistic treatment planning, in which percentiles of stochastic dose volume parameters are used, with the aim of a proof of principle for clinical use. As a first result, the three exact improvements on Algorithm 1 show an overall reduction in computation time of 30% in the calculation of the value, gradient and Hessian of probabilistic dose volume parameters. These improvements guarantee a similar speed-up in the total optimization time of a probabilistic plan, depending on the dose parameters in the objectives and constraints, and their computation times. However, the actual speed-up in probabilistic treatment plan optimizations with patient data in iCycle should still be investigated. Furthermore, the accuracy of the gradient and Hessian of probabilistic dose volume parameters has been improved through the introduction of a monotonicity constraint on the PCE coefficients.

To reduce the computation time of Algorithm 1 even further, an approximation method is devised for the calculation of the gradient and Hessian of probabilistic dose volume parameters. This approximation method and the exact improvements together have the potential to reduce the computation time of Algorithm 1 from 8,400 seconds to only 220 seconds, when using the 3D spine geometry and a sample size of 500,000. The problem is that the approximation method is quite heuristic and its accuracy cannot be predicted in general. However, relatively accurate percentile approximations were produced with small absolute errors when tested with the beam intensity vector of the optimized probabilistic treatment plan of case 1.

The exact calculations and the approximation method, both with the improvements on Algorithm 1, have been tested in treatment plan optimizations, in which a constraint is set on the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the tumor. These treatment plans clearly outperform the treatment plans of cases 1 and 2 in terms of the dose distributions in the tumor in the nominal scenario and in the (extreme) error scenarios. Overall, the three cases with the approximation method with four, six and eight argument averaging all show approximately a factor of 7.6 decrease in computation time from 5.3 days to 16.5 hours, compared to the case with the exact method and a sample size of 500,000. This is due to a rapid increase in the computation time of Algorithm 1 in the exact method for increasing sample size, which does not outweigh the increase in the number of iterations when using approximations of the gradient and Hessian of the 95%-probabilistic $d_{98\%}(\boldsymbol{x})$ in the approximation method. The cases with the exact method show to be 2.5 and 1.2 times faster for a sample size of 50,000 and 250,000 than for a sample size of 500,000 respectively.

A decrease in computation time is shown for probabilistic treatment planning, in which probabilistic dose volume parameters are used, in general, but also in implementations in the 3D spine geometry with and without the approximation method. Still, the approximation method should be tested on different geometries and for multiple uncertain parameters in the geometry to investigate whether this method is robust enough to be tested in optimizations with patient data in iCycle. Furthermore, different optimization algorithms should be tested to verify the decrease in optimization time. As a conclusion, the methods in this research do improve probabilistic treatment planning with polynomial chaos expansion for proton therapy by alleviating computational cost, but a proof of principle for clinical use is not yet achieved.

BIBLIOGRAPHY

[1] World Health Organization. "Cancer Mortality Rates Worldwide". URL: https://www.who.int/news-room/fact-sheets/detail/cancer. [Online. Posted: 12 September 2018. Accessed: February 2020].

[2] American Cancer Society. "Common Types of Cancer Treatment". URL: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types.html. [Online. Posted: March 2019. Accessed: February 2020].

[3] R. Baskar et al. "Biological response of cancer cells to radiation treatment". *Frontiers in molecular biosciences*, 1:24, 2014. ISSN: 2296-889X. DOI: 10.3389/fmolb.2014.00024.

[4] G. Delaney et al. "The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines". *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104(6):1129–1137, 2005. DOI: 10.1002/cncr.21324.

[5] American Cancer Society. "How Is Chemotherapy Used to Treat Cancer?" URL: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/chemotherapy/how-is-chemotherapy-used-to-treat-cancer.html. [Online. Posted: 22 November 2019;. Accessed: February 2020].

[6] A. Gardner et al. "Tumor localization, dosimetry, simulation and treatment procedures in radiotherapy: the isocentric technique". *American Journal of Roentgenology*, 114(1):163–171, 1972. DOI: 10.2214/ajr.114.1.163.

[7] T. Mitin and A.L. Zietman. "Promise and Pitfalls of Heavy-Particle Therapy". *Journal of Clinical Oncology*, 32(26):2855–2864, 2014. DOI: 10.1200/JCO.2014.55.1945.

[8] R.R. Wilson. "Radiological Use of Fast Protons". *Radiology*, 47(5):487–491, 1946. DOI: 10.1148/47.5.487.

[9] C.E. ter Haar. "Robustness Recipes for Proton Therapy". Master Thesis. Delft University of Technology, 2018.

[10] W. Liu et al. "Robust optimization of intensity modulated proton therapy". *Medical physics*, 39(2):1079–1091, 2012. DOI: 10.1118/1.3679340.

[11] S.R. van der Voort. "Application of polynomial chaos in proton therapy". Master Thesis. Delft University of Technology, 2015.

[12] J.H. Salverda. "Probabilistic Treatment Planning with Polynomial Chaos Expansion for Proton Therapy". Master Thesis. Delft University of Technology, 2019.

[13] National Cancer Institute. "Understanding Cancer Prognosis". URL: https://www.cancer.gov/about-cancer/diagnosis-staging/prognosis. [Online. Posted: 17 June 2019. Accessed: February 2020].

[14] W.D. Newhauser and R. Zhang. "The physics of proton therapy". *Physics in Medicine & Biology*, 60(8):R155–R209, 2015. DOI: 10.1088/0031-9155/60/8/r155.

[15] H. Paganetti. "Range uncertainties in proton therapy and the role of Monte Carlo simulations". *Physics in Medicine & Biology*, 57(11):R99–R117, 2012. DOI: 10.1088/0031-9155/60/8/r155.

[16] W. Ulmer. "Theoretical aspects of energy–range relations, stopping power and energy straggling of protons". *Radiation physics and chemistry*, 76(7):1089–1107, 2007. ISSN: 0969-806X. DOI: 10.1016/j.radphyschem.2007.02.083.

[17] U. Schneider et al. "The calibration of CT Hounsfield units for radiotherapy treatment planning". *Physics in Medicine & Biology*, 41(1):111, 1996. DOI: 10.1088/0031-9155/41/1/009.

[18] A. Elia. "Characterization of the GATE Monte Carlo platform for non-isocentric treatments and patient specific treatment plan verification at MedAustron-Vienna-Austria". PhD Dissertation. University of Lyon, 2019.

[19] H. Paganetti et al. "Proton Therapy Physics". *Medical Physics and Biomedical Engineering*, pages 103–690, 2012. ISSN: 0017-9078. DOI: 10.1097/HP.0b013e31824e7040.

[20] T. Bortfeld et al. "Proton Beam Radiotherapy—The State of the Art". *Medical Physics*, 32(6Part13):2048–2049, 2005. DOI: 10.1118/1.1999671.

[21] V. Rudat et al. "Combined error of patient positioning variability and prostate motion uncertainty in 3D conformal radiotherapy of localized prostate cancer". *International Journal of Radiation Oncology, Biology & Physics*, 35(5):1027–1034, 1996. ISSN: 0360-3016. DOI: 10.1016/0360-3016(96)00204-0.

[22] M. Yang et al. "Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration". *Physics in Medicine & Biology*, 57(13):4095–4115, 2012. DOI: 10.1088/0031-9155/57/13/4095.

[23] M. Engelsman et al. "Physics controversies in proton therapy". 23(2):88–96, 2013. DOI: 10.1016/j.semradonc.2012.11.003.

[24] C. Weltens et al. "Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging". *Radiotherapy and Oncology*, 60(1):49–59, 2001. ISSN: 0167-8140. DOI: 10.1016/S0167-8140(01)00371-1.

[25] Z. Perkó et al. "Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion". *Physics in Medicine & Biology*, 61(12):4646, 2016. DOI: 10.1088/0031-9155/61/12/4646.

[26] S. Breedveld et al. "iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans". *Medical physics*, 39(2):951–963, 2012. DOI: 10.1118/1.3676689.

[27] J. Unkelbach et al. "Robust radiotherapy planning". *Physics in Medicine & Biology*, 63(22):22TR02, 2018. DOI: 10.1088/1361-6560/aae659.

[28] S. van de Water et al. "Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams". *Physics in Medicine & Biology*, 58(19):6969, 2013. ISSN 0031-9155. DOI: 10.1088/0031-9155/58/19/6969.

[29] S. van de Water et al. "Shortening delivery times of intensity modulated proton therapy by reducing proton energy layers during treatment plan optimization". *International Journal of Radiation Oncology* Biology* Physics*, 92(2):460–468, 2015. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2015.01.031.

[30] S. Breedveld et al. "Multi-criteria optimization and decision-making in radiotherapy". *European Journal of Operational Research*, 277(1):1–19, 2019. DOI: 10.1016/j.ejor.2018.08.019.

[31] S.R. van der Voort et al. "Robustness recipes for minimax robust optimization in intensity modulated proton therapy for oropharyngeal cancer patients". *International Journal of Radiation Oncology, Biology & Physics*, 95(1):163–170, 2016. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2016.02.035.

[32] O.L. Le Maître and O.M. Knio. *"Spectral methods for uncertainty quantification: with applications to computational fluid dynamics"*. Springer Science & Business Media, 2010.

[33] Z. Perkó. "Sensitivity and Uncertainty Analysis of Coupled Reactor Physics Problems: Method Development for Multi-Physics in Reactors". PhD Dissertation. Delft University of Technology, 2015.

[34] D. Xiu. *"Numerical methods for stochastic computations: a spectral method approach"*. Princeton University Press, 2010.

[35] S.A. Smolyak. "Quadrature and interpolation formulas for tensor products of certain classes of functions". *Doklady Akademii Nauk*, 148(5):1042–1045, Russian Academy of Sciences. 1963.

[36] G. Blatman and B. Sudret. "Anisotropic parcimonious polynomial chaos expansions based on the sparsity-of-effects principle". *Proc ICOSSAR'09, International Conference in Structural Safety and Relability*, 2009.

[37] J. Stoer and R. Bulirsch. *"Introduction to numerical analysis"*, volume 12. Springer Science & Business Media, 2013.

[38] G. Blatman and B. Sudret. "An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis". *Probabilistic Engineering Mechanics*, 25(2):183–197, 2010. ISSN: 0266-8920. DOI: 10.1016/j.probengmech.2009.10.003.

[39] D. Xiu and G.E. Karniadakis. "The Wiener–Askey polynomial chaos for stochastic differential equations". *SIAM journal on scientific computing*, 24(2):619–644, 2002. ISSN: 1095-7197. DOI: 10.1137/S1064827501387826.

[40] G. Szegö. *"Orthogonal Polynomials"*. 2nd Edition, American Mathematical Society, New York. (1st Ed., 1939.), 1959.

[41] L. Gilli et al. "Uncertainty quantification for criticality problems using non-intrusive and adaptive polynomial chaos techniques". *Annals of Nuclear Energy*, 56:71–80, 2013. DOI: 10.1016/j.anucene.2013.01.009.

[42] T. Bortfeld. "An analytical approximation of the Bragg curve for therapeutic proton beams". *Medical physics*, 24(12):2024–2033, 1997. DOI: 10.1118/1.598116.

[43] R.H. Byrd et al. "A trust region method based on interior point techniques for nonlinear programming". *Mathematical programming*, 89(1):149–185, 2000. DOI: 10.1007/PL00011391.

[44] MathWorks. "Constrained Nonlinear Optimization Algorithms". URL: `https://nl.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html#brnpd5f`. [Online. Posted: unknown. Accessed: August 2020].

[45] MathWorks. "Prctile function". URL: `https://nl.mathworks.com/help/stats/prctile.html`. [Online. Posted: unknown. Accessed: September 2020].

[46] MathWorks. "First-Order Optimality Measure". URL: `https://nl.mathworks.com/help/optim/ug/first-order-optimality-measure.html`. [Online. Posted: unknown. Accessed: October 2020].

[47] E. Langford. "Quartiles in elementary statistics". *Journal of Statistics Education*, 14(3), 2006. DOI: 10.1080/10691898.2006.11910589.

[48] C. Sabater et al. "A Bayesian Approach for Quantile Optimization Problems with High-Dimensional Uncertainty Sources". 2020. DOI: 10.13140/RG.2.2.10177.79202.

[49] D. Tilly et al. "Probabilistic optimization of dose coverage in radiotherapy". *Physics and Imaging in Radiation Oncology*, 10:1–6, 2019. DOI: 10.1016/j.phro.2019.03.005.

[50] S. Wendt. "Uncertainty analysis in proton therapy using adjoint theory". Master Thesis. Delft University of Technology, 2020.

A

MATHEMATICS OF PROBABILISTIC DOSE VOLUME PARAMETERS

The methodology of the percentile approximation method for perturbed beam intensities in Chapter 7.4 is quite heuristic. It is not possible to describe the accuracy of the method in general, but the methodology could nevertheless provide approximations of the gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ with respect to the beam intensity vector in a fast manner. In this section of the Appendix, the gradient and Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ are described mathematically to provide possible starting points for new approximation methods.

The dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ has a certain probability density function $PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}(d_{\alpha\%},\boldsymbol{x})$ and a corresponding cumulative distribution function $CDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}(d_{\alpha\%},\boldsymbol{x})$. This PDF has no explicit dependence on the uncertainty vector $\boldsymbol{\xi}$, since this is the stochastic input. Mathematically, the CDF of a dose volume parameter $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ yields:

$$CDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}(d_{\alpha\%},\boldsymbol{x}) = \mathbb{P}(d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi}) \leq d_{\alpha\%},\boldsymbol{x}) = \int_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})\leq d_{\alpha\%}} p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi}, \tag{A.1}$$

with $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ the joint probability density of the uncertainty vector $\boldsymbol{\xi}$. Then, the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$, denoted as $d^{\beta\%}_{\alpha\%}(\boldsymbol{x})$ in the following equations, relates to Equation A.1 as:

$$CDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d^{\beta\%}_{\alpha\%}(\boldsymbol{x}),\boldsymbol{x}\right) = \frac{100-\beta}{100}. \tag{A.2}$$

From Equation A.2 it follows immediately that the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is defined as:

$$d^{\beta\%}_{\alpha\%}(\boldsymbol{x}) = CDF^{-1}_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(\frac{100-\beta}{100},\boldsymbol{x}\right). \tag{A.3}$$

Taking the derivative of both sides of Equation A.2 with respect to the beam intensity vector $\boldsymbol{x}$ yields:

$$\begin{aligned} 0 = \frac{d\left(\frac{100-\beta}{100}\right)}{d\boldsymbol{x}} &= \frac{d}{d\boldsymbol{x}} CDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d^{\beta\%}_{\alpha\%}(\boldsymbol{x}),\boldsymbol{x}\right) \\ &= \frac{d}{d\boldsymbol{x}} \int_{-\infty}^{d^{\beta\%}_{\alpha\%}(\boldsymbol{x})} PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%},\boldsymbol{x}\right) d(d_{\alpha\%}) \\ &= \frac{d\left(d^{\beta\%}_{\alpha\%}(\boldsymbol{x})\right)}{d\boldsymbol{x}} PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d^{\beta\%}_{\alpha\%}(\boldsymbol{x}),\boldsymbol{x}\right) + \int_{-\infty}^{d^{\beta\%}_{\alpha\%}(\boldsymbol{x})} \frac{dPDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%},\boldsymbol{x}\right)}{d\boldsymbol{x}} d(d_{\alpha\%}), \end{aligned} \tag{A.4}$$

in which Leibniz' rule of differentiation under the integral sign is used at the last equal-sign. By rewriting Equation A.4, an expression for the gradient of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ with respect to the beam intensity vector is obtained as:

$$\begin{aligned}\frac{d\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x})\right)}{d\boldsymbol{x}} &= -\frac{1}{PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)}\int_{-\infty}^{d_{\alpha\%}^{\beta\%}(\boldsymbol{x})}\frac{dPDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}(d_{\alpha\%},\boldsymbol{x})}{d\boldsymbol{x}}d(d_{\alpha\%}) \\ &= -\frac{1}{PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)}\int_{-\infty}^{d_{\alpha\%}^{\beta\%}(\boldsymbol{x})}\frac{d}{d\boldsymbol{x}}\frac{d}{d(d_{\alpha\%})}CDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}(d_{\alpha\%},\boldsymbol{x})d(d_{\alpha\%}) \\ &= -\frac{1}{PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)}\int_{-\infty}^{d_{\alpha\%}^{\beta\%}(\boldsymbol{x})}\frac{d}{d\boldsymbol{x}}\frac{d}{d(d_{\alpha\%})}\int_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})\leq d_{\alpha\%}}p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi}d(d_{\alpha\%})\end{aligned} \tag{A.5}$$

with substitution of Equation A.1 at the third equal-sign. A first note to Equation A.5 is that a dose parameter in general, and thus also a dose volume parameter, is physically constraint to be non-negative. Therefore, the lower integration limit can be set to zero, yielding:

$$\begin{aligned}\frac{d\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x})\right)}{d\boldsymbol{x}} &= -\frac{1}{PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)}\int_{0}^{d_{\alpha\%}^{\beta\%}(\boldsymbol{x})}\frac{dPDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}(d_{\alpha\%},\boldsymbol{x})}{d\boldsymbol{x}}d(d_{\alpha\%}) \\ &= -\frac{1}{PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)}\int_{0}^{d_{\alpha\%}^{\beta\%}(\boldsymbol{x})}\frac{d}{d\boldsymbol{x}}\frac{d}{d(d_{\alpha\%})}\int_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})\leq d_{\alpha\%}}p_{\boldsymbol{\xi}}(\boldsymbol{\xi})d\boldsymbol{\xi}d(d_{\alpha\%}).\end{aligned} \tag{A.6}$$

Equation A.6 gives rise to two different ways to express the gradient of the $\beta\%$-probabilistic $d_{\alpha\%}(\boldsymbol{x})$. Both expressions contain the value of the PDF of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ at the point $d_{\alpha\%}^{\beta\%}(\boldsymbol{x})$. One way to determine this value is to first construct a PCE of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ with the beam intensity vector from the current iteration step in the treatment plan optimization. Then, the PCE is sampled N_s times over different realizations of the uncertainty vector $\boldsymbol{\xi}$ and a normalized histogram is made according to these samples using a certain number of bins, for which there are multiple choices of rules of thumb to choose the number of bins. The value of $d_{\alpha\%}^{\beta\%}(\boldsymbol{x})$ can be obtained, for example with the *Prctile* function in Matlab from the set of PCE samples, and finally the value of $PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)$ is found by determining the value of the PDF at the bin in the normalized histogram in which the parameter $d_{\alpha\%}^{\beta\%}(\boldsymbol{x})$ is located. Figure A.1 shows an example of such a normalized histogram of the $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ with the beam intensity vector obtained from the optimized treatment plan of case 1 in Chapter 7.1.

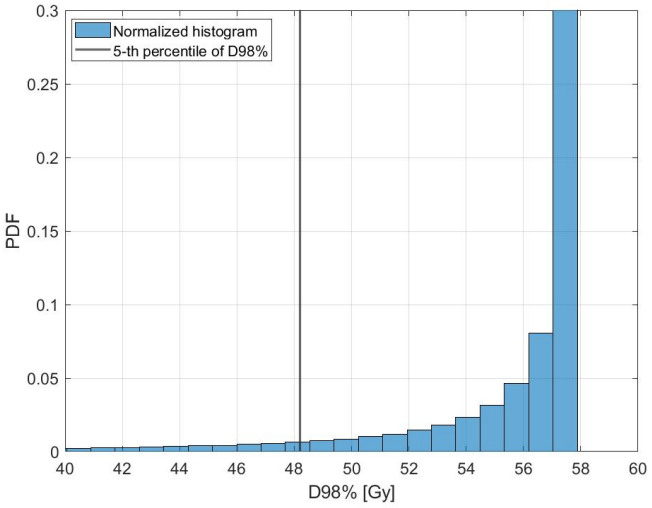


Figure A.1: Normalized histogram of 500.000 samples from a PCE of $d_{98\%}(\boldsymbol{x},\boldsymbol{\xi})$ with the beam intensity vector obtained from the optimized treatment plan of case 1 in Chapter 7.1. In the PCE construction, the characteristic $L7E1O8$ is used. The vertical black line shows the 5th-percentile of the samples. This histogram coincides with the constructed PCE in Figure 7.11.

Another way to determine the value of $PDF_{d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})}\left(d_{\alpha\%}^{\beta\%}(\boldsymbol{x}),\boldsymbol{x}\right)$ is to construct a continuous PDF of $d_{\alpha\%}(\boldsymbol{x},\boldsymbol{\xi})$ instead of a histogram, which is a non-continuous function. The set of PCE samples can be transformed into a PDF through parameterized or non-parameterized probability density fitting. Two possible types of distributions for parameterized fitting are the Beta and Kumaraswamy distribution, the latter being similar to the Beta distribution, but is characterized by closed form expressions of the PDF, CDF and quantile functions. With a change of variable the distribution could be defined on an interval $[0, d_{max}]$ instead of the default interval $[0,1]$. Also, Matlab contains a *Distribution fitter* toolbox, in which there are parametric and non-parametric distribution options. Besides this toolbox, non-parameterized probability density fitting could also be performed through kernel density estimation. However, no results with density fitting have been produced yet.

The integral in Equation A.6 is yet to be evaluated. A starting point here might be the evaluation of the inner integral over an inequality surface in the sample space of the uncertainty vector $\boldsymbol{\xi}$. Possibly, a relation might be found between this integral and the results that are derived in Appendix A of [50]. The Hessian of the β%-probabilistic $d_{\alpha\%}(\boldsymbol{x})$ is derived by taking the derivative of Equation A.6 with respect to the beam intensity vector.